# The exploitation of cell-free protein synthesis technologies for integrated structural biology

Vinesh Mervyn Jugnarain

Communauté UNIVERSITÉ Grenoble Alpes

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE LA COMMUNAUTE UNIVERSITÉ GRENOBLE ALPES

Spécialité : Chimie Physique Moléculaire et Structurale

Arrêté ministériel : 25 mai 2016

Présentée par


## Vinesh Mervyn JUGNARAIN

Thèse dirigée par **Edward MITCHELL**

**Préparée au sein du laboratoire de Synthelis, du laboratoire de deuteration du Life Science Group (D-lab, LS, Institut Laue-Langevin) et de l'European Synchrotron Radiation Facility (ESRF)**

dans **l'École Doctorale Chimie et Sciences du Vivant**


# Exploitation de la technologie acellulaire pour des applications en biologie structurale intégrée

# The exploitation of cell-free protein synthesis technologies for integrated structural biology

Thèse soutenue publiquement le **14 juin 2019**, devant le jury composé de :

**Monsieur Christophe Moreau**, Directeur de Recherche, CEA/CNRS/UGA
Président du Jury
**Madame Isabel Alves**, Directeur de Recherche, Université Bordeaux - CNRS
Rapporteur
**Monsieur Damien Baigl**, Professeur, Ecole Normale Supérieure (ENS)
Rapporteur
**Madame Pascale Crépieux**, Directeur de Recherche, Université de Tours - CNRS Examinateur
**Monsieur Pascal Demange**, Directeur de Recherche, Institut de Pharmacologie et de Biologie Structurale (IPBS) - CNRS
Examinateur

Aussi présents comme invités :
**Madame Magali Mathieu**, Sanofi Bio Structure and Biophysics/iDD
**Madame Montserrat Soler-Lopez**, Chargé de Recherche, ESRF
**Monsieur Alexey Rak** ( Autre iDD / Sanofi R&D )

# THE EXPLOITATION OF CELL-FREE PROTEIN SYNTHESIS TECHNOLOGIES FOR INTEGRATED STRUCTURAL BIOLOGY

**Vinesh Mervyn JUGNARAIN**

***EXPLOITATION DE LA TECHNOLOGIE ACELLULAIRE POUR DES APPLICATIONS EN BIOLOGIE STRUCTURALE INTÉGRÉE***

*A thesis submitted for the degree of Doctor of Philosophy*

École doctorale Chimie et Sciences du Vivant

Université Grenoble Alpes

Avril 2019

UNIVERSITÉ GRENOBLE ALPES

## Charte Anti-Plagiat

**Art 1 : Définition du plagiat**

Le plagiat consiste à reproduire un texte, une partie d'un texte, une illustration ou des idées originales d'un auteur, sans lui en reconnaître la paternité par un référencement bibliographique ou iconographique adéquat (Cf. art. 3).

**Art 2 : Objet des travaux universitaires**

Sont considérés comme travaux universitaires tous les documents réalisés par les étudiants et les enseignants, les chercheurs et les enseignants-chercheurs dans le cadre des activités de formation et de recherche. Ces travaux universitaires doivent toujours avoir pour ambition de produire un savoir inédit et d'offrir une lecture critique, nouvelle et personnelle du sujet.

**Art 3 : Méthodologie de référencement bibliographique**

La méthodologie d'un travail universitaire implique que les emprunts (par exemple par copier/coller) soient clairement identifiés et que le nom de l'auteur et la source de l'extrait soient mentionnés.

Les citations textuelles y compris dans une traduction personnelle, doivent obligatoirement être placées entre guillemets et être accompagnées d'une référence bibliographique à la suite de la citation, ou en note de bas de page.

Les emprunts non textuels *(*tableaux, graphiques, photos, formules scientifiques, etc.) doivent également être accompagnés d'une référence bibliographique à leur suite ou en note de bas de page.

En complément, toutes les références des documents cités, empruntés ou adaptés, doivent figurer en bibliographie.

**Art 4 : Détection du plagiat**

L'Université Grenoble Alpes est dotée d'un outil permettant de contrôler systématiquement les travaux universitaires et de détecter les similitudes, dans le but de rechercher le plagiat.

**Art 5 : Sanctions disciplinaires pour plagiat**

Les auteurs présumés de plagiat seront traduits devant la section disciplinaire compétente qui pourra prendre des sanctions pouvant aller jusqu'à l'exclusion définitive de tout établissement d'enseignement supérieur. La procédure disciplinaire n'exclut pas d'éventuelles poursuites judiciaires.

**Art. 6 : Engagement**

Les étudiants et les personnels s'engagent à ne pas commettre de plagiat dans leurs travaux universitaires. À cette fin, ils reconnaissent avoir pris connaissance des obligations décrites dans les articles 2 et 3 de la présente charte et s'engagent à s'y conformer.

Je certifie, JUGNARAIN Vinesh Mervyn *(indiquer nom prénom),* **avoir pris connaissance de la charte anti-plagiat et à la respecter.**

Date et signature : 28/06/2019

## ABSTRACT

Cell-free protein synthesis (CFPS) is an alternative method to cellular expression for recombinant protein production. Its lack of cellular boundaries enables direct modification of the CFPS reaction to allow optimised and rapid production of functional and labelled proteins. The aim of this industry-funded PhD was to demonstrate the applicability of the Synthelis CFPS technology for integrated structural biology by showing that CFPS proteins were functionally and structurally equivalent to those from *in vivo* expression systems.

These objectives were addressed with analytical approaches to the production and testing of different lysates used in the CFPS reactions and the use of model proteins for CFPS production and purification, and biophysical and structural analyses. The model proteins studied were the GPCR (G-protein coupled receptor) CXCR4 (C-X-C motif chemokine receptor 4) and its soluble ligand SDF1-α (stromal cell-derived factor alpha - 1), and the membrane protein CD4 (cluster of differentiation 4) and its truncated soluble variant 2dCD4. CXCR4 and SDF1-α are essential in the chemotaxis of cells, and CD4 plays a critical role in immune response. Pathologically, CXCR4/SDF1-α are involved in cancer progression, and CD4 is the primary receptor for HIV GP120 (glycoprotein 120) during viral infection.

For the purposes of integrated structural biology, a primary objective was to optimise CFPS to generate sufficient amounts of protein. From a screen of CFPS lysates derived from several *E. coli* strains, it was determined that Rosetta lysate enhanced yields of CXCR4. High-cell density cultures (HCDC) in fermenters were then established for lysate production. Subsequently, HCDC was used for the production of deuterated lysates for SANS (small-angle neutron scattering) studies.

Full-length CXCR4 and CD4 were produced as proteoliposomes or additionally, for CXCR4, as detergent solubilised protein. However, the expression and purity levels of solubilised CXCR4 and CD4 remained too low to permit advanced structural work. Nevertheless, soluble and functional SDF1-α and 2dCD4 were produced using lysate obtained from the SHuffle *E. coli* strain, which promotes disulphide bond pairing. CFPS is therefore advantageous over *in vivo* expression, where these occur as insoluble proteins and require an intensive refolding procedure. Subsequently, CFPS was established as a method for the rapid production of deuterated SDF1-α and 2dCD4.

The CFPS-generated proteins were assessed using biophysical techniques. ELISA (enzyme-linked immunosorbent assay) and SPR (surface plasmon resonance) analyses of CXCR4 in proteoliposomes

using conformational antibodies indicated binding and, therefore, a proper folding. Additionally, CD4 functionality in proteoliposomes was confirmed by using GP120 and conformational antibodies in ELISAs. Mass spectrometry of SDF1-α and its deuterated forms confirmed disulphide bonding and deuteration levels. The functionality of SDF1-α variants were confirmed using chemotaxis assays. With a view to long-term optimisation of CFPS, CXCR4 proteoliposomes were assessed using a NanoSight™ particle tracker, which provided insights into the changes in size and size distribution of proteoliposome particles during the CFPS reaction.

Advanced structural studies were performed on SDF1-α and 2dCD4. Structural determination of CFPS-generated SDF1-α by X-ray crystallography showed that it was identical to that of the published *E. coli* refolded form. The preliminary SANS assessment of deuterated 2dCD4 in complex with GP120 provided initial confirmation of the success of CFPS produced protein to provide viable SANS data. Finally, a SANS feasibility study demonstrated the possibility to obtain structural information about model proteins despite in the presence of unfractionated lysates. This novel technique could lead to the characterisation of CFPS membrane proteins in proteoliposomes, following a simple partial purification step.

**KEYWORDS:** Cell-free protein synthesis (CFPS), CXCR4, CD4, GP120, SDF1-α, membrane proteins, proteoliposome, X-ray crystallography, small-angle neutron scattering (SANS)

## RESUME

La synthèse des protéines acellulaire/ « cell-free » (cell-free protein synthesis, CFPS) est une méthode alternative à l'expression *in vivo* pour la production de protéines recombinantes. L'absence de barrières cellulaires permet de modifier aisément la réaction CFPS afin d'optimiser le rendement d'expression de la protéine et de favoriser son marquage. Le but de cette thèse était de démontrer que la technologie CFPS de Synthelis permet de produire des protéines de qualité, fonctionnellement et structurellement équivalentes à celles produites dans des systèmes *in vivo*.

L'optimisation du CFPS et de la purification ont été abordées avec des protéines modèles : le RCPG (récepteur couplé à la protéine G) CXCR4 (récepteur de chimiokine de type C-X-C), son ligand naturel soluble SDF1-α (facteur 1 de cellule stromale), la protéine membranaire CD4 (cluster of differentiation 4), et la forme soluble et tronquée de CD4, 2dCD4. CXCR4 et SDF1-α sont essentiels dans la chimiotaxie des cellules, et CD4 joue un rôle essentiel dans la réponse immunitaire. CXCR4 et SDF1-α sont impliquées dans la progression du cancer, tandis que le CD4 sert de récepteur primaire pour GP120 (glycoprotéine 120) du VIH pendant la phase d'infection.

Grace à un criblage de lysats dérivés de plusieurs souches d'*E. Coli*, il a été montré que le lysat de Rosetta était le plus adapté pour produire CXCR4. Par la suite, des cultures à haute densité cellulaire en fermenteur ont été réalisées pour la production du lysat. Ainsi, les protocoles de culture en fermenteur ont été adaptés pour la production de lysats deutérés. Ces lysats ont été utilisés pour des études de diffusion de neutrons aux petits angles (small-angle neutron scattering, SANS).

Ensuite, les protéines CXCR4 et CD4 ont été produites dans des protéoliposomes. De plus, CXCR4 a été produite sous forme solubilisée avec un détergent. Cependant, les niveaux d'expression et de pureté de ces deux protéines sont restés trop faibles pour permettre une caractérisation structurale. Néanmoins, en utilisant du lysat obtenu à partir de la souche SHuffle d'*E. coli* qui favorise les ponts disulfures, les molécules SDF1-α et 2dCD4 ont été produites sous forme solubles et fonctionnelles. Le CFPS est donc avantageux par rapport à l'expression *in vivo* qui ne permet pas une expression de ces molécules sous forme soluble. De plus, le CFPS a permis la production deutérée de ces deux protéines.

Les protéines CFPS ont ensuite été évaluées à l'aide de techniques biophysiques. Grâce à des analyses par ELISA (Enzyme-linked immunosorbent assay) and SPR (surface plasmon resonance) la liaison spécifique entre des protéoliposomes CXCR4 et des anticorps conformationnels ont été validée. Des

analyses similaires ont montré la liaison entre GP120 et des protéoliposomes CD4. La spectrométrie de masse sur la molécule SDF1-α et sur ses formes deutérées a confirmé les niveaux de deutération, et la formation des ponts disulfures. La fonctionnalité de SDF1-α a été confirmée par des tests de chimiotaxie. Enfin, une étude avec le traqueur de particules NanoSight$^{TM}$ sur les protéoliposomes CXCR4, a permis de mieux comprendre les changements de tailles et de la distribution des protéoliposomes au cours de la réaction CFPS.

Finalement, des études structurales ont été réalisées avec SDF1-α et 2dCD4. SDF1-α, a été cristallisée et caractérisée avec de la diffraction aux rayons X, et a permis de démontrer que la protéine produite par CFPS était identique à celle obtenu *in vivo*. Ensuite, l'étude du complexe 2dCD4 deutéré et GP120 au SANS a fourni des résultats préliminaires sur l'application du CFPS pour le SANS. Enfin, une étude de faisabilité SANS a indiqué la possibilité d'obtenir des informations structurales sur les protéines modèles malgré la présence de lysats non fractionnés. Cette nouvelle technique pourrait conduire à la caractérisation des protéines membranaires produites par CFPS dans des protéoliposomes.

**MOT-CLES :** Synthèse des protéines par « cell-free » (CFPS), CXCR4, CD4, GP120, SDF1-α, protéine membranaire, protéoliposome, cristallographie, diffraction aux rayons X, diffusion de neutrons aux petits angles (small-angle neutron scattering, SANS).

## ACKNOWLEDGEMENT

As a CIFRE funded PhD, where the worlds of academia and industry collide and combine, the making of this thesis was a great journey full of discrovery and learning. As with any journeys however, the road was not always smooth, and new paths had to be carved. Surmounting these challenges required a lot of strength and perseverance, but fortunately, I was lucky to find myself with a number of very understanding people who would encourage, and push me when needed, to finally complete the project.

I would like to thank my academic supervisors Prof. Edward Mitchell (European Sychrotron Radiation Facility/ESRF, Grenoble), Prof. Trevor Forsyth (Institut Laue-Langevin/ILL, Grenoble) and Dr. Michael Haertlein (ILL, Grenoble) for their guidance, help and support throughout the all of these years.

I would like to thank my supervisor from the industrial side, Dr. Sandra Cortès for help and support, and for teaching all of the basic principles and advanced techniques about the cell-free protein expression system. I would like to thank Mr. Bruno Tillier for welcoming me to the Synthelis team and offering me this incredible opportunity for carrying out my PhD as part of the ANRT/CIFRE scheme.

I would like to thank Dr. Martine Moulin (ILL, Grenoble) for all of her help and support, and for showing me all of the very handy tips and tricks in the lab. I would also like to thank her for great assistance in the setting up, monitoring and production of high-cell density cultures in fermenters, in both hydrogenated and deuterated media. I am also thankful to her for help with purification protocols.

I would like to thank Prof. Maria Papathanosopoulos at the HIV Pathogenesis Research Unit (HIVPRU) from the University of Witwatersrand (South Africa) for collaborating with us, and providing practical perspectives on the project. I would also like to thank Prof. Theresa Coetzer for financing our trip and work in South Africa. From the HIVPRU, I would like to thank Dr. Nichole Cerutti from the HIVRPL for her help and expertise for the ELISA, SPR setup and SPR analysis of CXCR4. I would also like to thank Dr. Gavin Owen and Nancy Tumba for their help with ELISAs of CD4 and GP120.

I would like Dr. Carlo Petosa and Dr. Montserrat Soler-Lopez for kindly being present as jury, and for their advices and encouragement, at each and every of my Comité de Suive de These (CST).

This work used the platforms of the Grenoble Instruct-ERIC Centre (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). From the Institut de Biologie Structurale (IBS, Grenoble), I would like to thank Dr. Winnie Ling, Dr. Fenel Daphna, and Dr. Schoen Guy, for their help and expertise with electron microscopy and data interpretation. The electron microscope facility is supported by the Rhône-Alpes Region, the Fondation Recherche Medicale (FRM), the fonds FEDER, the Centre National de la Recherche Scientifique (CNRS), the Centre Energie Atomique (CEA), the University of Grenoble Alpes, EMBL, and the GIS-Infrastrutures en Biologie Sante et Agronomie (IBISA).I would also like to thank thank Dr. Lucas Signor (IBS, Grenoble) for mass spectrometry and the analysis of the subsequent data.

I would like to thank Dr. Rabia Sadir (IBS, Grenoble) from the Groupe Structure et Activité des Glycosaminoglycanes (SAGAG) for providing the chemotactic cells, and her help and expertise in setting the chemotaxis assay for verifying SDF1-α function. I would also like to thank Dr. Corinne Deniaud from the Membrane and Pathogen group (IBS, Grenoble) for providing us with their plasmids of SDF1-α.

I would like to thank Mr. Romain Grondin and Dr. Aymeric Audfray from Malvern Instruments (Parc Club, Moulin à Vent, Vénissieux, France) for their help, expertise and demonstration using the Malvern Panalytical *NanoSight*$^{TM}$ NS300 Instrument.

I would also like to thank Dr. Anne Martell for her help and expertise with SANS experiments on D22 instrument (ILL, Grenoble), and with the subsequent analysis and data reduction. I would like to thank Dr. Judith Houston for her help and expertise with SANS experiments on the KWS-2 instrument (Forschungs-Neutronenquelle Heinz Maier-Leibnitz/FRMII, Munich), and with the consequent data analysis and data reduction. I woul also like to thank Dr. Sylvain Prevost for kindly taking his time to help us in interpretation of the SANS data.

I would like to thank all of the people from the entire Life Science Group, Synthelis team and HIVRU for their help and support throughout the entire PhD journey. A special thanks goes to Jennifer Channell, with whom we collaborated with for the production and structural characterisation of CFPS 2dCD4.

Finally, very special thanks to Dr. Alexio Capovilla, my MSc supervisor, who really incited the scientific spirit in me. As it turns out, Alex was the main person responsible for my three months internship at the ILL/ESRF, and incidentally, my first ever trip to Grenoble and Europe. From there, I was then recruited to begin the PhD project… “May the wheel of Science keep on turning” – A.C.

# CONTENTS

## GLOSSARY

2dCD4 : Soluble truncated variant of CD4 consisting of first two distal domains

3PG : 3-phosphoglycerate

AMP : Adenosine monophosphate

APC : Antigen presenting cells

ATP : Adenosine triphosphate

BSA : Bovine serum albumin

CD4 : Cluster of differentiation 4

CFPS : Cell-free protein synthesis

CIP : Complete protease inhibitor cocktail (Roche)

CMC : Critical micelle concentration

CMP : Cytosine monophosphate

CXCR4 : CXC – Chemokine receptor 4

DDM : n-Dodecyl β-D-maltoside

$D_2O$ : Deuterium oxide (Heavy water)

*E. coli* : *Escherichia coli*

ELISA : Enzyme-linked immunosorbent assay

ECL : Electro chemi-luminescent

EDTA : Ethylenediaminetetraacetic acid

Glob : Bovine γ-globulin sample (Sigma-Aldrich, G5009)

GMP : Guandine monophosphate

GP120 : Glycoprotein 120

HCDC : High-cell density culture (fermenter)

HEPES : 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid

HIV : Human immunodeficiency virus

HL : Hydrogenated lysate

IPTG : Isopropyl β-D-1-thiogalactopyranoside

K/Glu : Potassium glutamate

LB : Loading buffer

MBP : Maltose binding protein

Mg/Glu : Magnesium glutamate

MHC : Major histocompatibility complex

ML : Matchout cell-free lysate

MW : Molecular weight

NAD : Nucleoside adenosine

PBS : Phosphate buffered saline

PEP : Phosphoenol phosphate

PL : Perdeuterated cell-free lysate

Rg : Radius of gyration

Rpm : Revolutions per minute

SANS : Small-angle neutron scattering

SAXS : Small-angle X-ray scattering

SDF1-α : Stromal cell-derived factor 1-alpha

SDS-PAGE : Sodium dodecyl sulphate-polyacrylamide gel electrophoeresis

SEC : Size exclusion chromatography

SLD : Scattering length density

SOC : Super optimal broth with catabolite repression

Tris : Tris (hydroxymethyl) aminomethane

UMP : Uracil monophosphate

YPTG : Yeast peptone tryptone glucose

# RESUME DE LA THESE

## EXPLOITATION DE LA TECHNOLOGIE CELL-FREE POUR DES APPLICATIONS EN BIOLOGIE STRUCTURALE INTÉGRÉE

## OBJECTIF DU PROJET DE THESE

En tant que projet financé par le dispositif CIFRE (**C**onventions **I**ndustrielles de **F**ormation par la **Re**cherche), l'objectif général de cette thèse était de favoriser la collaboration entre partenaires industriels et académiques. Synthelis (France), une entreprise qui émane de l'Université de Grenoble Alpes (UGA, France), est une société de biotechnologie spécialisée dans la production, la purification et la caractérisation de protéines. Grace à sa technologie brevetée exploitant l'utilisation du système d'expression « Cell-Free » (Cell-Free Protein Synthesis, CFPS), Synthelis se démarque par son expertise à produire des protéines membranaires, et protéines difficiles à exprimer. L'objectif principal du projet était de démontrer, grâce à des techniques structurales de pointe, que l'intégrité structurale des protéines produites par la technologie de Synthelis était équivalente voire supérieure à celles produites par des systèmes d'expression classique *in vivo*. L'accès aux grands instruments permettant des études structurales a été possible grâce à une collaboration fructueuse avec les partenaires académiques comme l'Institut Laue-Langevin (ILL, France) et l'Installation Européenne de Rayonnement de Synchrotron (ESRF, France) (Figure1). Une collaboration internationale a également été initiée avec le laboratoire de recherche sur la pathogenèse du VIH (HIV Pathogenesis Research Unit, HIVPRU) de l'Université de Witwatersrand (Johannesbourg, Afrique du Sud) afin de caractériser la fonctionnalité de CD4, protéine produite dans cette étude et impliquée dans l'interaction avec le virus du SIDA.

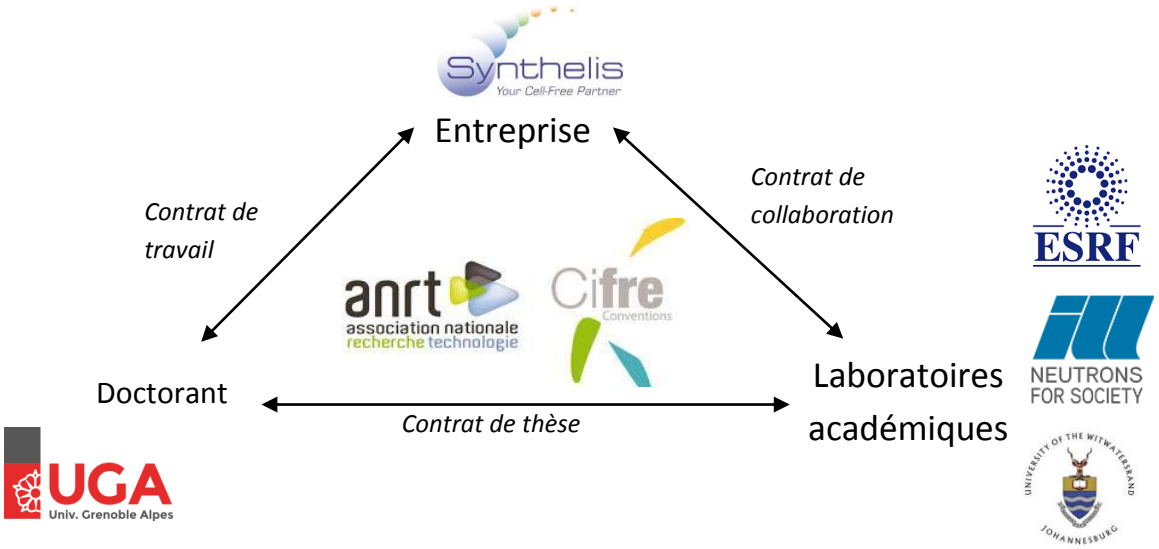



***Figure 1:*** *Déroulement de la thèse avec les partenaires industrielle et académique.*

Afin de démontrer le potentiel grandissant du CFPS comme méthode d'expression des protéines, 3 protéines modèles ont été étudiées lors de cette thèse. Ces 3 protéines sont les suivantes : le RCPG (récepteur couplé à la protéine G) CXCR4 (récepteur de chimiokine de type C-X-C), son ligand naturel soluble SDF1-α (facteur 1 de cellule stromale), la protéine membranaire CD4 (cluster of differentiation 4), et la forme soluble et tronquée de CD4, 2dCD4. Les critères d'évaluation ont été le rendement d'expression et de purification, ainsi que l'obtention d'un bon repliement de ces protéines. Grâce à la collaboration des partenaires académiques mentionnés ci-dessus, la diffusion des neutrons aux petits angles (small-angle neutron scattering, SANS) et la cristallographie aux rayons-X, ont été les deux méthodes structurales privilégiées lors de cette étude. Dans le premier chapitre, la technologie cell-free sera présentée avec ses avantages pour l'expression des protéines et pour des études structurales ainsi que son potentiel pour une exploitation industrielle. La caractérisation structurale nécessite des quantités de l'ordre du milligramme de protéines pures ce qui a constitué un réel premier challenge. Le second chapitre sera dédié à l'optimisation de la production de l'extrait bactérien (lysat), élément crucial dans le système CFPS. Des extraits issus de différentes souches bactériennes ainsi que l'utilisation de culture à haute densité pour la production à plus grande échelle seront présentés. Dans le troisième chapitre, en utilisant les lysats les plus performants, les résultats de la production et de la purification des 3 protéines modèles marquées ou non marquées isotopiquement, seront présentés. Le quatrième chapitre portera sur la validation de l'intégrité biophysique et fonctionnelle des protéines modèles à l'aide de méthodes utilisées dans le cadre du contrôle de la qualité au niveau industriel. Dans le chapitre suivant, une étude structurale aux rayons-X de SDF1-α sera décrite et montrera l'intérêt de la technique CPFS en comparaison avec un système d'expression classique *in vivo*. Le chapitre six se focalisera sur la caractérisation structurale en solution du complexe formé par CD4 et GP120 (protéine virale du virus du SIDA) par la méthode SANS. Enfin, le dernier chapitre décrira une étude de faisabilité en SANS, qui pourrait permettre de caractériser structurellement directement les protéines au sein d'un lysat (non purifiées) produites par CFPS.

## INTRODUCTION A LA SYNTHESE DE PROTEINES EN « CELL-FREE »

Utilisée initialement comme une technique réductionniste pour comprendre les fondamentaux de la transcription et de la traduction, la synthèse des protéines par « cell-free » (cell-free protein synthesis, CFPS) a connu un développement important au cours des soixante dernières années, pour en faire une alternative viable pour l'expression des protéines recombinantes. Les étapes essentielles dans le développement du CFPS sont la maitrise de l'élimination des acides nucléiques endogènes qui entravent la synthèse de protéines, l'utilisation de la machinerie transcriptionnelle et traductionnelle, et l'amélioration dans le système de régénération d'énergie et d'élimination des déchets.

Le CFPS est une technique *in vitro* utilisée pour l'expression de protéines recombinantes (Figure 2). Il implique l'extraction et l'utilisation d'une partie cytosolique d'une cellule, appelée extrait ou lysat, qui contient les composants nécessaires à la transcription et à la traduction. Ainsi, en fonction des besoins de l'utilisateur, diverses cellules (*Escherichia coli*, mammifères, protozoaires, levure, insecte, germe de blé ….) peuvent être utilisées pour la production de ce lysat. Le CFPS basé sur des souches bactériennes d'*E. Coli (Escherichia coli)*, est le système le mieux établi et le plus rentable pour la production d'un rendement élevé en protéines. La société Synthélis utilise principalement cet organisme source. En sus, le CFPS d'*E. Coli* est une technique relativement simple à mettre en place car il permet la production d'une grande quantité de lysat. La réaction CFPS est initiée par l'ajout au lysat de l'acide nucléique codant pour la protéine d'intérêt (ADN plasmide ou linéaire) etles réactifs critiques de la réaction (bases nucléotides, acides aminés, molécules énergétiques, cofacteurs et sels). La réaction se poursuit jusqu'à épuisement de l'un des substrats, ou inhibée par accumulation de sous-produits.

Contrairement aux systèmes *in vivo* (à condition que le lysat soit déjà disponible), le CFPS est une méthode beaucoup plus simple et plus rapide pour la production de protéines recombinantes. De plus, un autre avantage de CFPS réside dans l'absence des barrières cellulaires. Ceci permet à l'utilisateur d'avoir un accès direct à la réaction pour le modifier selon ces besoins. Par exemple, le CFPS peut être adapté pour l'expression optimisée de protéines difficiles à exprimer – notamment des protéines membranaires par l'ajout de liposomes dans la réaction.

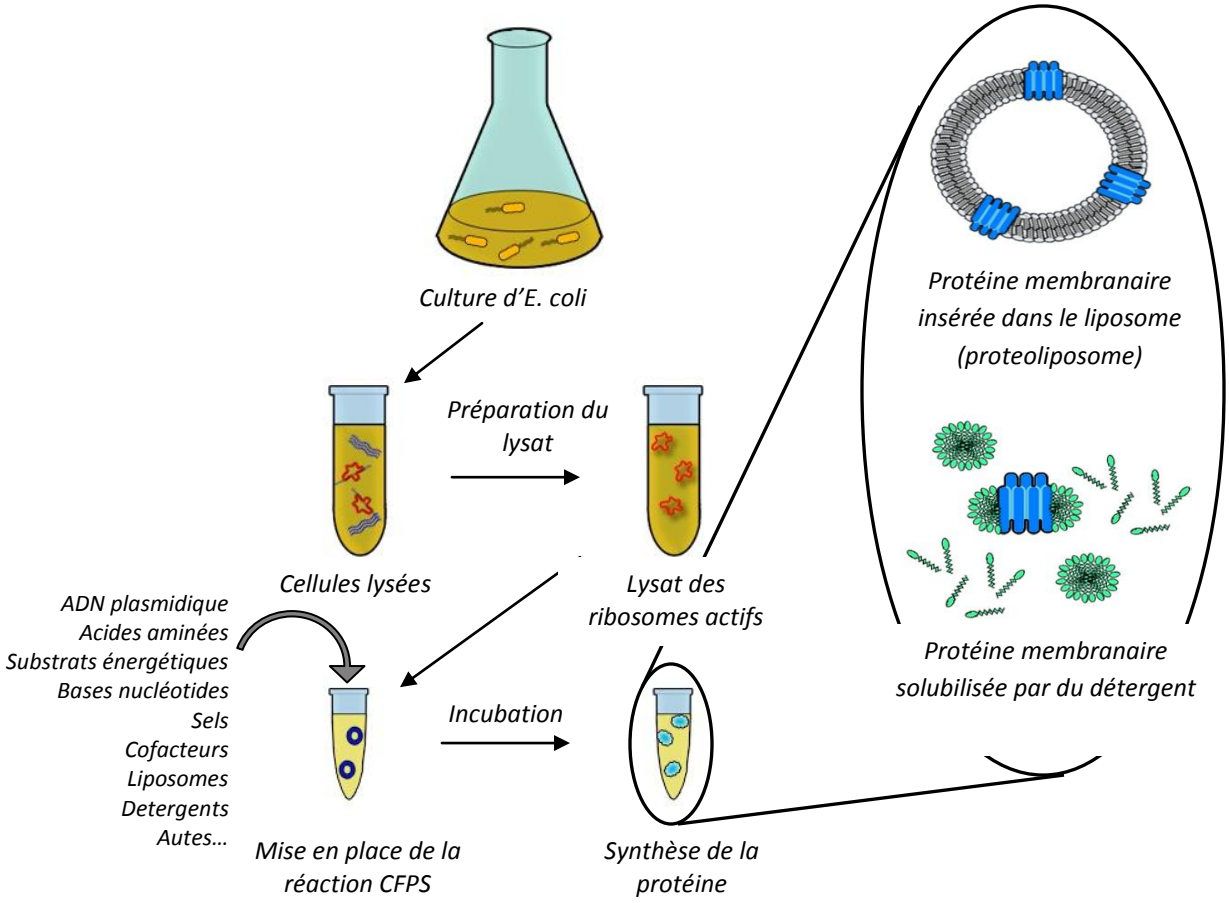



***Figure 2:*** *La synthèse des protéines membranaires en « cell-free » (cell-free protein synthesis, CFPS). Les cellules d'E. coli sont tout d'abord cultivées. La culture est récoltée, lysée et traitée pour éliminer l'ADN génomique, les membranes cellulaires et d'autres débris insolubles afin de produire le lysat actif, qui peut être conservé à -80 ° C jusqu'à utilisation ultérieure. La réaction CFPS est préparée en ajoutant de l'ADN plasmidique codant pour la protéine cible, des bases nucléotides, des acides aminés, des substrats énergétiques, des cofacteurs, des sels, des liposomes ou des détergents, des agents stabilisants ou d'autres additifs. La réaction est incubée à une température adéquate pour la production de protéines, qui peuvent ensuite être purifiées. Notamment, l'ajout de liposomes et des détergents comme support hydrophobes, permettent respectivement l'expression de protéines membranaires insérées dans une bicouche lipidique (proteoliposome) ou solubilisées.*

Les protéines membranaires, qui représentent 20 à 30 % du génome procaryote et eucaryote, jouent un rôle crucial dans la physiologie en assurant une diversité des fonctions biologiques.

Un défi majeur à leurs études revient à la nature hydrophobe des protéines membranaires, qui engendre des dépliements et agrégations lorsqu'elles sont exprimées *in vivo* par *E. coli*. Grâce à l'absence des barrières cellulaires, des matrices hydrophobes peuvent être ajoutées à la réaction CFPS afin de permettre l'expression de protéines membranaires structurellement intègres et fonctionnellement actives. Pour se faire, des liposomes (mimétiques de bicouche lipidique artificielle) permettant la formation de protéoliposomes, et des détergents, sont utilisés comme matrices hydrophobes. De plus, l'absence de barrières cellulaires permet au CFPS d'être un moyen relativement facile pour produire des protéines marquées en substituant les acides aminés non-marqués par des acides aminés marqués. L'utilisation du CFPS pour la production des protéines deutérées (enrichies en $^{2}H$) pour la diffusion des neutrons aux petits angles (SANS; small-angle neutron scattering), en particulier, a été de grand intérêt dans ce projet.

Dans cette étude, trois protéines modèles ont été exprimées par CFPS. Présente en particulier sur les globules blancs, la protéine membranaire CXCR4, de la famille des récepteurs couplés aux protéines G (RCPG), est impliquée dans la chimiotaxie des cellules vers le ligand SDF1-α. Dans des conditions pathologiques, CXCR4 et SDF1-α sont impliquées dans la croissance et métastase des cancers. CXCR4 est difficile à exprimer *in vivo* par la souche *E. coli* notamment car elle se produit généralement sous forme de corps d'inclusion insolubles, qui peuvent être toxiques pour la cellule. Cette toxicité limite donc la production de la protéine CXCR4 par la bactérie. De même, SDF1-α peut s'exprimer sous forme de corps d'inclusion en système i*n vivo E. coli* et ce, malgré sa nature soluble. La protéine CD4 est une protéine transmembranaire qui joue un rôle essentiel dans le système immunitaire. Cependant, elle sert aussi de récepteur primaire pour l'infection par le VIH dans sa cellule hôte. Cette protéine, comme toute protéine membranaire, est très difficile à produire et il a été reporté dans la littérature l'expression et l'étude d'une forme variante tronquée de CD4, appelée 2dCD4, comprenant seulement les deux premiers domaines. Néanmoins, l'expression de 2dCD4 en *E. coli* conduit également à des corps d'inclusion insolubles. Le rendement de 2dCD4 active après dénaturation et renaturation reste très faible. L'ensemble de ces protéines modèles, présentant des difficultés d'expression *in vivo*, sont donc des candidats idéals pour l'utilisation du CFPS, et démontrer ainsi son efficacité.

Grâce à l'amélioration continue de la technologie au cours des dernières décennies, notamment en terme d'augmentation des rendements d'expression et de réduction des coûts d'exploitation, le CFPS

commence a trouvé sa place en tant que système d'expression au sein des industries. Le CFPS est aujourd'hui utilisé dans des bioréacteurs de plus de 200 litres afin de produire des protéines solubles et fonctionnelles avec de haut rendement de production (SutroBiopharma Inc). La montée en échelle du système CFPS, dans le contexte de l'industrie, n'est donc plus un obstacle à son utilisation et peut constituer une alternative très avantageuse et rentable à l'expression *in vivo*, en particulier, pour la production des protéines membranaires actives qui représentent la cible de plus de 50 % des médicaments modernes. Ces perspectives envisagent donc un intérêt commercial très considérable.

## LA CULTURE BACTERIENNE ET PRODUCTION DES LYSATS

La production du lysat est une étape cruciale dans la procédure du CFPS. Dans le but d'améliorer la qualité et les rendements de production des protéines produites par CFPS, le lysat a été produit en utilisant diverses souches bactériennes: Nico21, C43 et Rosetta, Rosetta-Gami et SHuffle. Dans un contexte industriel, la standardisation et la reproductibilité des protocoles sont des éléments cruciaux permettant d'harmoniser et de valider un procédé. Dans cette étude, la croissance des cultures bactériennes et les rendements en biomasse de ces nouvelles souches ont été enregistrés et comparés les uns par rapport aux autres. Pour les protéines modèles étudiées, les lysats issus de souches bactériennes Rosetta et SHuffle ont été trouvés plutôt efficaces en termes de rendement d'expression. Au niveau industriel, nous observons une demande accrue de protéines produites par CFPS pour notamment des applications en biologie structurale. La mise en place de cultures bactériennes à haute densité cellulaire (HCDC; high cell density culture) dans des fermenteurs, en condition hydrogénée ou deutérée, a donc été réalisée. Cette expérience qui permet des analyses par SANS représente donc une étude de faisabilité d'une croissance bactérienne automatisée et contrôlée à grande échelle pour la production du lysat.

## L'OPTIMISATION DE L'EXPRESSION PROTEINE ET PURIFICATION

Après la mise au point des lysats bactériens, les protéines recombinantes CXCR4, CD4 et SDF1-α ont été exprimées par CFPS. La protéine CXCR4 a été exprimée sous deux formes: sous format protéoliposomes

enchâssées dans des liposomes, ou sous format solubilisé avec des détergents. Après optimisation de la réaction CFPS, les protéines SDF1-α et de 2dCD4 ont été produites sous forme soluble (sans matrice hydrophobe). Compte tenu des grandes quantités de protéines généralement requises pour les études structurales, l'expression et la purification de chacune de ces protéines produites par CPFS a dû être optimisées: la nature du lysat utilisé, le temps et la température d'expression, ainsi que la teneur en magnésium dans les réactions CFPS ont été étudiés. Un criblage de différents détergents à différentes concentrations a également été réalisé afin d'exprimer la protéine CXCR4 solubilisée. Malgré la panoplie de conditions testées, l'expression et la pureté de CXCR4 sont malheureusement restées trop faibles pour mener des études structurales. Par contre, des rendements élevés ont été obtenus pour les protéines solubles, SDF1-α et 2dCD4, ce qui a permis leurs caractérisations structurales. En effet, avec la mise enplace de méthodes de production par CFPS à grande échelle (jusqu'à 100 ml) et des protocoles de purification optimisés, des quantités adéquates de ces protéines ont été obtenues. Pour les études en SANS, des protocoles optimisés pour les formes deutérées des protéines ont également été développés. À la suite de ces travaux, des procédures opérationnelles standard (SOP) ont été établies afin d'aider à la standardisation et à la reproductibilité dans le cadre industriel.

## CHARACTERISATION BIOPHYSIQUE ET FONCTIONELLE DES PROTEINES CFPS

Après l'expression et la purification des protéines par CFPS, elles ont ensuite été caractérisées à l'aide d'un ensemble de méthodes biophysiques et de tests de fonctionnalité. Il s'agit non seulement d'une étape importante dans la validation des protéines, mais cela permet également d'avoir un aperçu sur les outils qui peuvent être utilisés en routine pour le contrôle de la qualité dans le contexte industriel. En confirmant la liaison à leur ligand, ou à un anticorps conformationnel via des mesures par ELISA et par SPR, des indices sur l'intégrité structurale de CD4 et CXCR4 ont été démontrés. Des tests de chimiotaxie utilisant la molécule de SDF1-α, marqués et non-marqués, ont permis de comparer et de confirmer la fonctionnalité des protéines CXCR4 en démontrant que leurs propriétés chimiotactiques étaient maintenues. Enfin, une analyse de SDF1-α par spectrométrie de masse a démontré que le CFPS permettait de reconstituer les ponts disulfures, et que d'autre part SDF1-α a bien été deuterée à hauteur de 77 %.

La caractérisation des protéoliposomes reste cependant difficile en raison de leur hétérogénéité et de leur complexité. Par conséquent et en vue d'améliorer l'homogénéité des produits, la taille et la distribution des liposomes/protéoliposomes ont été étudiées. Le traqueur de particules NanoSight™ (Malvern) a été utilisé avec succès pour observer les changements dans la distribution des tailles des liposomes/protéosoliposomes de CXCR4 avant, pendant et après la réaction CFPS. Il a été démontré que les petits liposomes de taille uniforme (<100 nm) introduits en début de réaction CFPS subissaient une augmentation de taille drastique due à l'insertion de CXCR4, formant ainsi des protéoliposomes. De plus, pour obtenir une représentation visuelle des protéoliposomes, la microscopie électronique à transmission (TEM) a été évaluée. Cependant, en raison de la fragilité des protéoliposomes, ceux-ci ont été détruits pendant la TEM rendant l'interprétation difficile. Néanmoins, la TEM a constitué un premier pas vers des techniques plus avancées de microscopie cryo-électronique, qui sont moins destructives pour les échantillons.

## LA STRUCTURE DE SDF1-α PRODUITE PAR CFPS

SDF1-α, générée par CFPS, a été cristallisée avec succès et soumis à la diffraction par rayons-X. La structure cristalline de SDF1-α s'est révélée être identique à celle précédemment publiée qui a été générée à partir de SDF1-α produit *in vivo* par *E. coli*. Cette expérience suggère donc fortement que les protéines produites par CFPS sont structurellement équivalentes à celles produites par des méthodes *in vivo* classiques.

## DIFFUSION A PETITS ANGLES DE 2DCD4 EN COMPLEXE AVEC GP120

La diffusion des neutrons aux petits angles (SANS; small-angle neutron scattering) est une technique qui utilise les propriétés de diffusion différentielle que possèdent les neutrons sur les protéines et le solvant dans l'échantillon afin d'obtenir des informations structurales sur les protéines d'intérêt. Le SANS est une technique de contraste (Figure 3A): lorsqu'il n'y a pas de différence entre les propriétés de diffusion neutron entre la protéine et le solvant, il n'y a pas de signal visible de la protéine. Ainsi, la protéine rendue « invisible », est dite être « matched-out » par rapport au solvant. En modifiant la teneur

$H_2O/D_2O$ dans le solvant, le contraste du solvant par rapport aux protéines peut être modifié pour que les protéines soit « matched-out » (Figure 3B). Grâce à la technique de « matching-out », les protéines deutérées ayant des propriétés de diffusion très différentes de celles des protéines non-marquées (dite hydrogénée), peuvent être distinguées individuellement lorsqu'elles sont présentes dans le même échantillon. Cette technique est particulièrement utile pour l'étude des complexes de protéines en solution, et pour l'observation des changements conformationnels résultant de l'interaction (Figure 4).

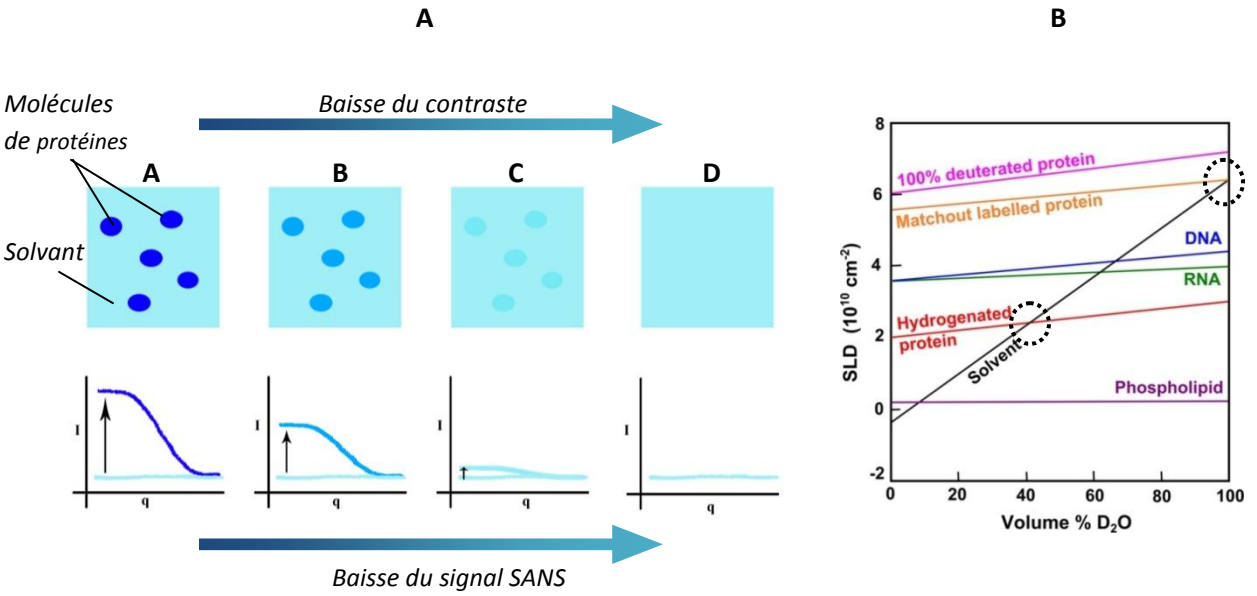



***Figure 3**: Schéma illustrant une baisse du contraste entre le solvant et la protéine d'intérêt, qui amène à la condition de « matching-out » où la protéine est rendue « invisible ». Cela s'observe par une baisse du signal SANS (A). L'altération du contraste se fait grâce à l'utilisation de protéines hydrogénées ou deutérées, qui ont chacune des propriétés de diffusion de neutrons très différentes, et en changeant la teneur $H_2O/D_2O$ dans le solvant pour arriver à leurs conditions de « matchout » (B, encerclé).*

Une étude préliminaire par SANS du complexe formé par 2dCD4 deuterée produit par CFPS, et le dimère de GP120, a été réalisée. L'étude par SANS de ce complexe est intéressante car elle permet l'utilisation

de la forme native de GP120, qui est une molécule flexible et glycolysée (37 % de son poids). Le principe de variation du contraste est de « matching-out » pour analyser individuellement (i) le complexe en entier, (ii) le 2dCD4 dans le complexe où GP120 est rendue « invisible » et (iii) le dimère GP120 où 2dCD4 est rendu « invisible » (Figure 4). Bien que les analyses aient montrées que 2dCD4 deutéré était effectivement « matched-out », les signaux restaient très faibles et difficile à interpréter. Toutefois, nous sommes parvenus à déterminer quelques critères sur la taille du complexe et ses composants, et suggérons un modèle du complexe. Néanmoins, l'importance de cette étude était de définir les conditions requises pour le SANS du complexe 2dCD4-GP120, ainsi que ses limites. Particulièrement, l'amélioration des signaux de SANS se ferait par l'utilisation de complexe qui comprend la molécule de 2dCD4 deuterée avec le monomère GP120. Ce complexe, avec une proportion de 2dCD4 :GP120 de 1 :1, serait effectivement vraiment homogène et concentré pour favoriser les signaux de SANS.


***Figure 4:*** *Schéma des conditions expérimentales pour la visualisation du complexe 2dCD4-GP120 par SANS. En ajustant la teneur de $D_2O$ du solvant, on peut ainsi visualiser le complexe entier, uniquement 2dCD4, ou uniquement GP120.*

## ETUDE DE FAISABILITE SUR L'ANALYSE STRUCTURALE D'UNE PROTEINE NON-PURIFIE DANS LE LYSAT CFPS

La caractérisation structurale des protéines membranaires produites par CFPS dans un format protéoliposome a été difficile en raison des faibles niveaux d'expression. De plus, les protéines contaminantes issues du lysat bactérien sont présentes au sein des protéoliposomes limitant ainsi

fortement le degré de pureté de ce format. Néanmoins, ce format reste le format le plus pertinent et le plus adapté pour produire des protéines membranaires avec un repliement proche de la conformation native. Par conséquent, il a fallu développer d'autres méthodes permettant la caractérisation structurale des protéines membranaires dans les protéoliposomes. Comme première étape vers cet objectif, une étude de faisabilité en utilisant le SANS, avec les techniques de variation du contraste et de « matching-out », a été examinée afin de visualiser des protéines modèles *en situ*, et en présence du lysat bactérien. Un tel concept, s'il est possible, permettrait donc de valider la possibilité d'obtenir des informations structurelles d'une protéine d'intérêt malgré la présence de protéines contaminantes. En outre, le système présenterait de nombreux avantages en termes de simplicité et de pertinence pour les études de protéines dans leur environnent natif non-purifié.

Deux strategies ont été envisageés (Figure 5): l'une dans laquelle une protéine deutérée a été produite avec un lysat hydrogéné, et l'autre dans laquelle cette protéine hydrogénée a été produite avec un lysat deutéré. La question clé dans les deux cas est de déterminer si le signal du SANS attribué à la protéine d'intérêt est suffisamment fort et cohérent pour pouvoir être interprété, par rapport à la diffusion incohérente des signaux du lysat. Le travail de faisabilité que nous décrivons, a été réalisé pour tester cette idée dans son ensemble, et permettre d'évaluer les besoins et les limites de la technique. Différentes protéines modèles ont donc été étudiées par SANS en présence de lysats hydrogénés ou deutérés, Il a été montré que les signaux obtenus des protéines modèles étaient fortement maintenus malgré la présence croissante de la concentration du lysat. A partir de ces signaux, les paramètres de taille des protéines se sont révélés conforme aux échantillons de contrôle pur, où le lysat n'était pas présent. Enfin, les modèles *ab-initio* se sont avérés conformes les uns par rapport aux autres malgré la présence croissante du lysat dans l'échantillon. La perte de résolution de la surface des modèles était cependant une limitation de la technique. En conclusion, il est possible d'obtenir des informations structurales sur les protéines modèles malgré la présence du lysat, qui était présent à des quantités équivalentes à celles utilisées lors d'une réaction CFPS. Ces résultats définissent donc les exigences pour les prochaines études en SANS sur des protéines non purifiées et, à plus long terme, pour l'étude des protéines membranaires sous format protéoliposome produite par CFPS (Figure 5).

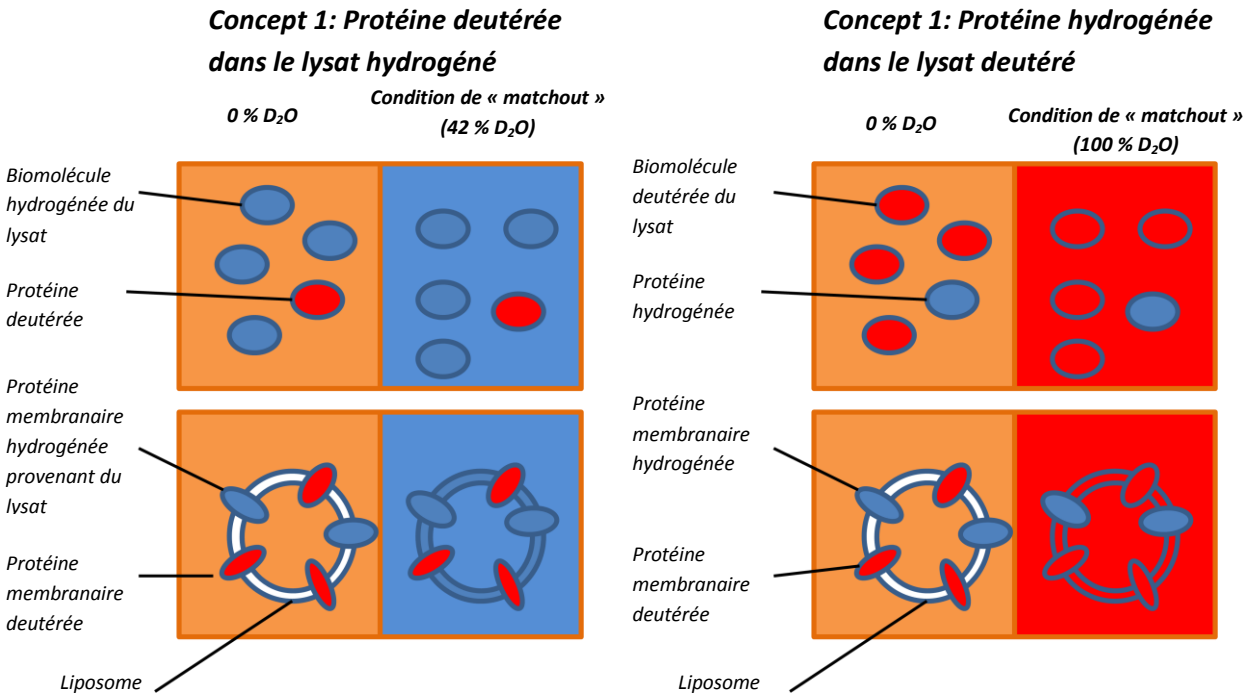



***Figure 5:*** *Schéma des échantillons pour l'analyse SANS des protéines en présence de lysat. Dans le premier cas, lorsqu'on est dans les conditions de « matching-out » avec 42 % de $D_2O$ dans le solvant, les biomolécules hydrogénées du lysat sont rendues « invisibles », tandis que la protéine d'intérêt deutérée est «visible». À l'inverse, dans le deuxième cas, lorsqu'on est dans les conditions de « matching-out » avec 100 % de $D_2O$ dans le solvant, les biomolécules deutérée du lysat sont rendues « invisibles », tandis que la protéine d'intérêt hydrogénée est rendue « visible ». La faisabilité de cette technique ouvrirait la voie vers la caractérisation structurale par SANS des protéines membranaires CFPS dans les protéoliposomes. Dans ce dernier cas, le liposome doit aussi être deutérée pour qu'il soit « matched-out ».*

## DISCUSSIONS, CONCLUSIONS ET PERSPECTIVES

L'objectif principal du projet était de démontrer, grâce à des techniques structurales de pointe, que l'intégrité structurale des protéines produites par la technologie de Synthelis était équivalente voire supérieure à celles produites *in vivo*. Pour cela, les protéines modèles CXCR4, SDF1-α, CD4 et 2dCD4 ont été sélectionnées :

• CXCR4 et CD4: En tant que protéines membranaires, CXCR4 et CD4 doivent être insérés dans la bicouche lipidique de la membrane cellulaire pour avoir leurs conformations natives. Les rendements de ces deux protéines recombinantes dans les cellules de mammifère sont généralement trop faibles pour envisager une caractérisation structurale. L'expression recombinante dans des cellules d'*E. coli*, en revanche, ne permet pas l'insertion de ces protéines membranaires mammaliennes dans la membrane cellulaire. Par conséquent, CXCR4 et CD4 sont exprimées sous formes insolubles et agrégées. Pour obtenir ces protéines CXCR4 et CD4 fonctionnelles, il est donc nécessaire de procéder à des méthodes drastiques de dénaturation et de repliement.

• SDF1-α et 2dCD4: Ces protéines peuvent être produites sous forme soluble avec de cellules mammaliennes. Cependant, l'utilisation du système d'expression mammalienne pour la production des protéines à des fins de caractérisation structurale, peut être coûteuse et incompatible avec le marquage isotopique des protéines (par exemple, la deutération). De plus, les rendements d'expression sont vraiment très faibles. L'utilisation de cellules d'*E. coli* permet une expression mais sous formes insolubles et agrégées. Ne pas pouvoir former des protéines solubles et correctement repliées en *E. coli*, peut être attribué au microenvironnement réducteur dans les cellules qui limite la formation de ponts disulfure. SDF1-α et 2dCD4 possèdent deux paires de ponts disulfures qui doivent être formées afin de favoriser un bon repliement. Pour obtenir des protéines solubles et fonctionnelles, SDF1-α et 2dCD4 exprimées en E. coli doivent subir une procédure de dénaturation et repliement hautement adaptée. Cependant, les rendements en protéines solubles et fonctionnelles à la fin du processus de repliement sont très faibles et la possibilité d'un mauvais repliement est élevée.

Afin d'optimiser ces expressions, différentes souches bactériennes *E. coli* (Nico21, C43, Rosetta, Rosetta-Gami and SHuffle) ont été cultivées et des lysats ont été faits pour la production de protéines par CFPS. Avec l'introduction de ces nouveaux lysats dans le répertoire de Synthelis, les courbes de croissance de ces nouvelles souches ont été enregistrées. Conforme à l'engagement industriel pour le contrôle de la

qualité, ces courbes de croissance ont été reproduites et le protocole standardisé écrit afin de servir comme référence pour les prochaines cultures.

Au fur et à mesure que nous augmentons le volume de culture d'*E. coli*, la procédure de fabrication du lysat devient rapidement difficile et laborieuse lorsqu'on utilise de simples erlens bafflés. Dans le but d'augmenter les volumes de production du lysat, nous avons adapté les protocoles établis pour la culture des bactéries à haute densité dans des fermenteurs (High-Cell Density Cultures, HCDC). Cette étape était importante pour permettre l'utilisation du CFPS à grande échelle, qui est nécessaire pour la production de grandes quantités de protéines nécessaire à des études structurales. De plus, les HCDC sont bien adaptés aux besoins de Synthelis pour le développement et l'automatisation de la chaine de production. L'utilisation de fermenteur, qui permet un contrôle automatisé sur la croissance à haute densité de bactéries, constitue donc un élément qui ouvrira la voie à d'importants volumes de culture.

Suite au criblage des lysats, il a été constaté que les lysats provenant des souches *E. coli* Rosetta et de SHuffle, étaient les plus adaptés aux protéines modèles présentées dans ce travail. Par conséquent, les perspectives d'améliorer l'efficacité du système cell-free en utilisant différentes souches d'*E. Coli,* sont certainement plausibles. Avec une grande sélection de souches d'*E. Coli,* ayant été spécialement conçue pour améliorer l'expression des protéines *in vivo* et qui sont disponibles commercialement, nous considérons leur potentiel pour le CFPS. En outre, compte tenu de l'absence de barrières dans le système CFPS, l'idée de mélanger les différents lysats, dérivés de souches d'*E. Coli*, ou voire même de cellules d'eucaryotes, est une perspective interessante à explorer. Nous bénéficierions ainsi des avantages des systèmes procaryotiques (rendement important) et des systèmes eucaryotiques (modifications post-traductionnelles). Néanmoins, sachant que les modifications post-traductionnelles comme la glycosylation compliquent la caractérisation structurale des protéines, la production de protéines en *E. coli* et donc sans glycolsylation est parfois mieux adaptée à la situation.

Nous avons par la suite exprimé par CFPS, deux protéines, CXCR4 et CD4 dans des proteoliposomes. De plus, nous sommes parvenus à exprimer la protéine CXCR4 solubilisée par des détergents. Malgré nos efforts, les rendements et la pureté de CXCR4 et de CD4 sont trop faibles pour être caractérisées structurellement. Dans le cas d'expression en liposomes/protéoliposomes, l'adhésion non-spécifique de protéines du lysat au protéoliposome était un problème persistant qui entravait la pureté des échantillons de CXCR4 et CD4. Dans le cas d'expression CFPS en présence de détergents, même s'il était possible d'obtenir la forme solubilisée et purifiée de CXCR4, le rendement final de celle-ci était extrêmement faible. Néanmoins, l'incapacité de produire du CXCR4 en grande quantité et pure pour

permettre la caractérisation structurale par des méthodes classiques, nous a encouragé à développer de nouvelles méthodes afin de caractériser des protéines en format protéoliposomes malgré la présence de protéines contaminantes.

Avec l'expression CFPS en protéoliposome de CXCR4/CD4, l'accrochage des protéines contaminantes issues du lysat bactérien sur les protéoliposomes, nuisait à la caractérisation structurale. Pour éviter ce problème, il est possible de produire des protéines membranaires pures enchâssées dans les liposomes en réinsérant ces protéines membranaires purifiées solubilisées par des détergents dans un liposome vide (Klammt *et al.*, 2004, Costello *et al.*, 1984). Bien que cela puisse être faisable, la question des faibles niveaux d'expression de CXCR4 en particulier peut toujours poser problème.

Le problème des faibles niveaux d'expression dépend toutefois des protéines. Plusieurs protéines membranaires produites à Synthelis ont un niveau d'expression élevé. L'expression de la forme entière de CXCR4/CD4 en CFPS a été relativement faible, et on observait des formes tronquées. Pour augmenter l'expression de CXCR4/CD4, l'optimisation de la réaction pour favoriser la formation de CXCR4/CD4 en forme entière via l'élimination des espèces tronquées, est une voie à explorer. La séquence de l'ADN codante étant cruciale pour l'expression de la forme entière ou incomplète des protéines, une étude dans laquelle les séquences codantes de CXCR4/CD4 seraient modifiées pour éviter les arrêts de transcription ou traduction serait à envisager.

Nous avons optimisé l'expression par CFPS et la purification de SDF1-α et de 2dCD4. Contrairement au lysat de référence, en utilisant le lysat SHuffle pour l'expression de ces deux protéines, nous avons démontré que ce dernier lysat était essentiel pour la formation de ponts disulfures. Ainsi, cela a permis l'obtention de protéines solubles et correctement repliées. CFPS constitue donc un avantage majeur car lorsqu'elles sont exprimées *in vivo* en *E. coli*, ces deux protéines se présentent sous formes insolubles et inactives (Picciocchi *et al.*, 2014, Murphy *et al.*, 2007, Cerutti *et al.*, 2014, Cerutti *et al.*, 2010). L'étape de renaturation nécessaire à un repliement correct de la protéine agrégée est ardue et potentiellement peu rentable, car elle ne garantit pas que les protéines soient solubles, fonctionnelles et correctement repliées en fin de procédé. Le système CFPS optimisé peut être utilisé avantageusement pour la production presque immédiate de protéines solubles et fonctionnelles, qui peuvent ensuite être purifiées à l'aide de méthodes standards.

De plus, nous avons démontré que le CFPS permettait la production aisée de SDF1-α et de 2dCD4 deutérées. CFPS s'est avéré avantageux par rapport au système classique d'*E. coli in vivo* pour

l'expression de protéines deutérées de part sa simplicité et sa vitesse de production. La rapidité et la facilité de production de protéines, non-marquées ou marquées, sont des aspects critiques dans le contexte industriel et académique où des échéances sont à respecter. Le CFPS a été présenté comme un système efficace pour l'expression des protéines deutérées. En particulier, étant une protéine cruciale dans la recherche VIH/SIDA, la facilité d'expression de 2dCD4 en CFPS, deutérée ou pas, constitue un progrès technique important. Ainsi, le travail présenté dans cette thèse sert à démontrer et à démocratiser le CFPS comme un outil accessible pour l'expression quotidienne, de non-seulement des protéines membranaires, mais également des protéines solubles.

Après l'optimisation initiale à petite échelle, nous sommes passés à une montée en échelle du CFPS pour permettre la production en grande quantité de SDF1-α et de 2dCD4 (lots de 100 mL). Malgré des niveaux de production faibles pour des demandes industrielles, la montée en échelle en bioréacteurs nous a permis d'utiliser SDF1-α et de 2dCD4 pour la caractérisation structurale. De plus, cela est une étape essentielle pour Synthelis dans le cadre d'une éventuelle montée à échelle du processus cell-free. Grâce à ces études d'optimisation et de montée à l'échelle, des protocoles d'exploitation standard (Standard operating protocoles, SOPs) ont été rédigés. Dans ce contexte, ces protocoles peuvent servir à normaliser, et à assurer la reproductibilité de la production.

Nous avons ensuite confirmé l'intégrité biophysique et fonctionnelle des protéines produites par CFPS en utilisant plusieurs méthodes. En plus de fournir des informations scientifiques importantes sur leurs propriétés fonctionnelles, ces techniques peuvent également être suggérées comme des méthodes pour caractériser les protéines CFPS au quotidien. L'ELISA, est une méthode peu coûteuse, mais très polyvalente, qui pourrait être mise en place dans le contexte des petites et moyennes entreprises. Conforme au besoin de la qualité au sein de l'industrie, l'utilisation de ces méthodes garantirait en outre la standardisation du CFPS en permettant d'identifier des anomalies entre les lots de production. Grâce à la caractérisation fonctionnelle, ces outils apporteraient une valeur ajoutée aux protéines CFPS sur le marché, et renforceraient la confiance des clients dans les produits CFPS.

La difficulté à caractériser les protéines membranaires CFPS dans les proteoliposomes à cause de l'abondance des protéines contaminantes, nous a orientés vers l'étude de deux outils : le NanoSight™ Particle Tracker et le TEM. Le NanoSight™ Particle Tracker s'est révélé être un outil simple d'utilisation, qui pouvait donner un résultat visuel sur la distribution des tailles et l'homogénéité des liposomes/protéoliposomes de CXCR4 avant, pendant et après la réaction CFPS. Ces résultats peuvent être utilisés pour évaluer, améliorer et standardiser la qualité des protéines, ainsi que la purification des

protéoliposomes. Dans un contexte industriel, cet outil peut être utile comme contrôle qualité des lots protéiques. Par exemple, le NanoSight™ Particle Tracker donne des indications sur l'état des protéoliposomes (taille, agrégat…). De plus, en ajoutant des marqueurs fluorescents ou en attachant des anticorps spécifiques fluorescent sur CXCR4, le dispositif pourrait permette la détermination du nombre réel des molécules de CXCR4 sur les protéoliposomes qui sont correctement insérées et orientées. La réalisation de cet objectif répondrait à une préoccupation des clients : à savoir la quantité de molécules CXCR4 fonctionnelles présentes sur les protéoliposomes. Repoussant davantage les limites du NanoSight™ Particle Tracker, le dispositif pourrait ensuite être utilisé pour détecter simultanément la liaison de SDF1-α à CXCR4, après que ces derniers aient été spécifiquement marqués avec des étiquettes fluorescentes.

Quant aux protéoliposomes difficile à être caractériser, ils ont été visualisés par TEM dans un premier temps, comme une étape préliminaire pour accéder à la cryo-EM. Contrairement au TEM où les proteoliposomes ont étaient détruites suite à la coloration pré-requise, le cryo-EM, peut se dispenser de cette étape : il a la capacité d'analyser les échantillons sans préparation majeure et dans leur état d'origine. Le Cryo-EM, ayant été utilisé pour l'étude de protéines membranaires isolées, des protéines membranaires intégrées dans la membrane, et de la dynamique de la membrane (Goldie *et al.*, 2014, Abe et Fujiyoshi, 2016, Mitsuoka et Gerle, 2016, Rawson *et al.*, 2016), offre donc une perspective très intéressante pour l'étude de la protéine membranaire produite par CFPS intégrée dans les protéoliposomes.

Par la suite, la structure de SDF1-α a été déterminée par cristallographie à rayons X. La structure de CFPS SDF1-α étant identique à la celle produite dans E. *coli in vivo,* ceci démontre qu'elles sont structurellement équivalentes. Ce résultat nous ramène donc vers le but premier du projet, qui s'agissait de démontrer l'intégrité structurale des protéines produites par le système CFPS de Synthelis.

Dans le but de démontrer davantage l'exploitation du CFPS pour la production de protéines marquées à des fins de caractérisation structurale, 2dCD4 deutérée générée par CFPS, complexée au GP120 virale, a été analysée par diffusion de neutrons aux petits angles (SANS). Malgré les signaux SANS médiocres qui contraignaient l'interprétation des données et des modèles, l'expérience nous a néanmoins aidé à définir les conditions requises pour les prochaines analyses SANS sur le complexe. Par conséquent, les travaux futurs doivent porter sur l'amélioration du signal de 2dCD4. De plus, l'utilisation de GP120 monomère plutôt que de GP120 dimère pourrais aussi aider à améliorer les signaux SANS. Contrairement au GP120 dimère où l'échantillon se constitue d'un mélange hétérogène, l'utilisation du

GP120 monomère permettra la production d'une solution hautement homogène constituée seulement de GP120 monomère lié à une seule 2dCD4. Toute aussi importante, l'étude de SANS a été essentielle pour comprendre les difficultés qui limiteront l'interprétation structurale. Notamment, les chaines flexibles et la surface glycosylée de GP120 (≈37 % de son poids) poseront un défi majeur pour l'interprétation précise des modèles. Des techniques permettant un marquage sélectif des chaines flexibles ou de la surface glycosylée pourraient être envisagées afin de distinguer ces éléments pendant le SANS.

Finalement, une étude de faisabilité pour le développement d'une nouvelle méthode de SANS pour caractériser structurellement une protéine CFPS, malgré sa présence des protéines contaminantes issue du lysat, a été lancée. En utilisant des protéines modèles, nous avons constaté que même avec une concentration élevée de lysat, ce dernier pouvait être efficacement « matched-out ». Ainsi, les informations structurales des protéines modèles étaient bien conservées et pouvaient être interprétées de manière à permettre la détermination des paramètres structuraux, et la génération des modèles à basse résolution. Cette étude de faisabilité, sert donc de tremplin pour établir les méthodes qui permettront l'interprétation structurale de la protéine membranaire intégrée dans les protéoliposomes, malgré les protéines contaminantes.

Néanmoins, la voie pour atteindre ce dernier objectif est difficile à plusieurs niveaux. Premièrement, dans le cas notamment de la protéine CXCR4, le système CFPS ne permet pas de générer des quantités suffisantes pour l'étude. Néanmoins, avec l'utilisation de protéoliposomes partiellement purifiés sur gradient de sucrose, les concentrations de protéines membranaires sont nettement plus élevées, tout en réduisant la quantité globale de lysat contaminant. Ensuite, le prochain défi concernerait la production des liposomes avec des lipides deutérée, de sorte que le lysat et le liposome soient simultanément « matched-out » au cours du SANS, ne laissant ainsi, que le signal de la protéine membranaire. Une autre difficulté s'agit de l'incapacité de produire des protéines CFPS hydrogénées lorsqu'on utilise le lysat deutéré. Nous pouvons suggérer donc de compléter la réaction CFPS avec les composants fonctionnels (par exemple, la polymérase) ce qui pourrait aider à rétablir son activité. Pour le moment, seule la production de protéines CFPS deutérées lorsqu'on utilise du lysat hydrogéné est possible. Finalement, nous devons admettre que les protéines non purifiées produites par CFPS sont généralement un mélange hétérogène comprenant des monomères, des oligomères et des espèces tronquées. Il s'agit en effet d'un échantillon SANS non idéal, qui limiterait donc la fiabilité les données.

Néanmoins, cette étude de faisabilité peut voir une grande application dans le domaine du SANS et de la biologie structurale. En effet, la capacité d'obtenir des informations structurales sur une protéine malgré la surabondance de biomolécules hétérogènes, rendrait possible l'utilisation de cette technique pour l'étude de proteines dans leurs états natifs. De plus, la technique peut permettre l'analyse structurale de protéines qui sont instables et susceptibles à la détérioration lorsqu'on les enlève de leur environnent natif après la purification. Enfin, nous avons aussi des idées qui impliqueraient l'analyse *in situ* de la réaction CFPS où nous observerions les protéines en cours de traduction.

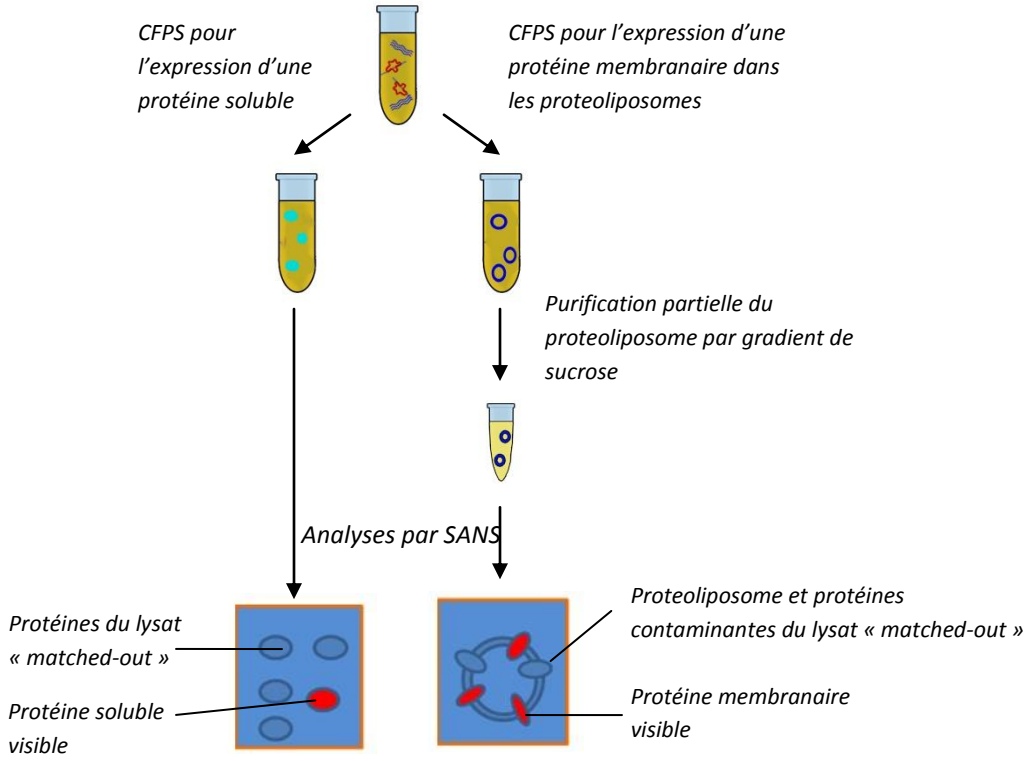



***Figure 6**: Schéma illustrant l'exploitation du SANS pour l'analyse de la protéine CFPS malgré la présence de protéines du lysat. Les protéines CFPS solubles se trouvent dans une très grande quantité de lysat ; ce qui contribue a beaucoup de bruit dans le signal SANS. Dans le cas des protéines membranaires produit dans protéoliposomes par contre, il serait possible d'avoir moins de bruit dans les signaux car il y a la possibilité de les purifier partiellement avant les analyses de SANS.*

L’objectif principal de ce projet était de démontrer que l’intégrité structurale des protéines produites par système CFPS de Synthelis était équivalente à celles produites *in vivo* et ceci via l’utilisation d’outils structuraux poussés. Cet objectif était indispensable pour Synthelis afin d’améliorer la valeur de ses produits, et de renforcer la confiance des clients dans son système CFPS. De plus, le projet a permis à l’entreprise un débouché sur le marché de la biologie structurale. Pour pouvoir caractériser structurellement les protéines produites par CFPS, il était donc nécessaire d'optimiser les méthodes et procédures du système pour permettre la production de protéines à grande échelle, en quantité et en qualité appropriées. Grace à ces travaux, il a été démontré que le CFPS est un système d’expression alternatif efficace, qui peut être utilisé avec beaucoup d’aise au quotidien pour la production rapide de protéines fonctionnelles.

# CHAPTER ONE

## INTRODUCTION TO CELL-FREE PROTEIN SYNTHESIS

## ABSTRACT

Cell-free protein synthesis (CFPS), which has existed for nearly 60 years, was used initially for understanding the fundamentals of protein transcription and translation. After decades of technological development however, CFPS is proving to be a powerful system for recombinant protein expression that can exploited for structural biology. In this chapter, CFPS will be introduced: the history, major development and current state of the art of CFPS will be presented, together with the fundamentals of CFPS and the advantages of the system for protein expression. As the basis of this thesis project, the exploitation of CFPS for structural biology, along with the model proteins CXCR4, its ligand SDF1 and the closely associated CD4 molecule will be introduced. Finally, CFPS represents an immense potential for industrial application, which will be summarised.


## INTRODUCTION

### 1.1. Early history of the cell-free expression system

The cell-free system, as a means for protein expression, began its development in the 1950s. The system was, however, not initially used specifically for the aim of producing of recombinant proteins, but rather as a reductionist system tool for understanding the transcription-translation axis – the central dogma of molecular biology (Crick, 1970). Some of the earliest cell-free expression systems described were constituted of extracts derived from mice hepatic cells to study protein synthesis using radioactive labelled amino acids (Zamecnik *et al.*, 1948, Siekevitz, 1952, Littlefield *et al.*, 1955). The use of extracts from staphylococcal cells, allowing the incorporation of $^{14}$C-labelled glutamic acid during protein synthesis when nucleic acids were added to the system, provided early indications of the fundamental link between DNA, RNA and protein expression (Gale and Folkes, 1954). Some years later, the critical role of ribosomes in protein synthesis was demonstrated with the use of microsome extracts derived from either human or rabbit reticulocytes to produce haemoglobin *in vitro* (Schweet *et al.*, 1958, Bank and Marks, 1966). Further results confirming the importance of ribosomes, mRNA and polysomes were established using cellular extracts from wheat germ (Marcus and Feeley, 1966). Later on, *E. coli* ribosomes extracted from growing cells were used to direct the synthesis of specific exogenous proteins

in the system by addition of synthetic mRNA template strands (Nirenberg and Matthaei, 1961, Matthaei and Nirenberg, 1961).

With further research, critical components in the cell-free system, such as ATP, an ATP generating system, GTP, $Mg^{2+}$, potassium acetate and amino acids, were being identified (Lamborg and Zamecnik, 1960). These very important early studies using the cell-free system undoubtedly served to help to lay the foundations of molecular biology. Additionally, however, they also laid the foundations for the cell-free expression system as an alternative tool for recombinant protein expression.

### 1.1.1. Development of cell-free system for protein expression

As mentioned above, the early cell-free expression systems were originally developed as tools for understanding the basis of transcription and translation – the foundations of modern molecular biology. Its first use hinting towards recombinant protein expression, whilst still being used primarily for gaining insights in the mechanisms of the DNA to protein pathway, was described as early as the 1960s. However, its use as an efficient, high yielding and economic method for recombinant protein expression was hindered by several problems, to which decades of research was applied to overcome these limitations.

#### *Endogenous nucleic acid*

Whilst gaining insight to the fundamental role of ribosomes in protein translation, the early cell-free extracts were mostly crude extracts derived from host cells in which only endogenous proteins were being expressed (Zamecnik *et al.*, 1948, Siekevitz, 1952, Littlefield *et al.*, 1955). These crude extracts still contained the endogenous nucleic – genomic DNA and mRNA, from the host cells which saturated the ribosomes, thereby limiting recombinant protein expression. As such, a critical development came from the establishment of new protocols which involved production of endogenous nucleic acid-free lysates. *E. coli* cells, harvested whilst in the early logarithmic growth phase, were lysed and the extract was treated with 30,000 g centrifugation steps, 35$^{o}$C incubation periods and extensive dialysis – thus creating the basis for the still popular modern-day S30 lysate (Nirenberg and Matthaei, 1961, Matthaei and

Nirenberg, 1961). These critical steps not only allowed the recovery of ribosomes, but enabled the riddance of endogenous *E. coli* nucleic acids that would otherwise compete with the ribosomes during recombinant protein expression when exogenous synthetic RNA templates were added to the cell-free system (Nirenberg and Matthaei, 1961, Matthaei and Nirenberg, 1961). In the case of rabbit reticulocyte derived lysates, the treatment of the extracts with calcium-dependent micrococcal nucleases and their subsequent removal, resulted in functional lysates devoid of endogenous mRNA (Pelham and Jackson, 1976).

### *Coupled transcription-translation*

The first cell-free systems being exploited for recombinant protein expression used exogenous mRNA added directly to the mixture to drive the protein synthesis (Nirenberg and Matthaei, 1961). Whether it was natural or synthetic mRNA, it had to first be produced and purified; a particularly challenging and time-consuming task in the early days of molecular biology further complicated by the unstable nature of mRNA.

The introduction of the DNA-driven coupled transcription-translation *E. coli* S30 based system was therefore a major boon in the development of CFPS as it allowed the production of protein in directly in a single batch reaction mix (Chen and Zubay, 1983, Cenatiempo, 1986, DeVries and Zubay, 1967). In essence, the system replicated the central dogma of molecular biology in a living cell: DNA to mRNA to protein. Furthermore, as opposed to its mRNA counterpart, DNA being more stable and less liable to degradation is easier and cost-effective to produce, handle and store. Additionally, with the development of powerful phage promoter driven systems in DNA plasmid encoding and the supplementation of the reaction with RNA polymerases, the *in vitro* DNA transcription step was even further enhanced thus opening the way for greater protein yields across both prokaryotic and eukaryotic CFPS systems (Stueber *et al.*, 1984, Craig *et al.*, 1992, Krieg and Melton, 1987). More specifically, the introduction of the well-known T7 RNA polymerase (RNA polymerase gene of bacteriophage T7 ) in the S30 *E. coli* lysate based reaction marked a five-fold increase in CFPS protein yields (Kigawa *et al.*, 1995).

### *Short-lived CFPS system*

CFPS is an open system where the cell barriers have been removed and the cytosolic portion of the cell, containing the transcription and translation machinery, is preserved. The strongest asset of the system can however also be its weakness for the lack of the external barrier does not allow the system to eliminate waste products which ultimately accumulate and inhibit protein synthesis. A major culprit for the inhibition of CFPS systems was deemed to be the accumulation of inorganic phosphate by-products derived from ATP metabolism which interferes with DNA transcription (Kern and Davis, 1997, Akama *et al.*, 2012). Additionally, CFPS rapidly extinguishes critical components of the ATP metabolism pathway and certain amino acids that are required for protein production (Kim and Swartz, 2000).

The use of the continuous exchange cell-free (CECF) mode served to overcome these issues of batch CFPS in *E. coli* S30 based lysate (Kim and Choi, 1996, Spirin *et al.*, 1988). CECF mode involves a cell-free reaction (reaction mix, RM) and a feeding chamber (feeder mix, FM) separated by a semi-permeable membrane with a defined cut-off to allow the diffusion of small molecules (Figure 1.1). The RM, as a classic batch reaction, contains the S30 lysate where transcription and translation take place, which requires energy and consumes essential molecules, whilst generating waste products. The FM enriched in low molecular weight resources (amino acids, nucleotides, energy-generating molecules) serves to continuously replenish the consumed molecules in the RM by diffusion. Simultaneously, waste products in the RM are allowed to diffuse down a concentration gradient through the dialysis membrane into the FM, thereby limiting their accumulation and inhibitory effects. The CECF mode has been reported to extend the CFPS active process from 1-2 hours to 48 hours (Kim and Choi, 1996).

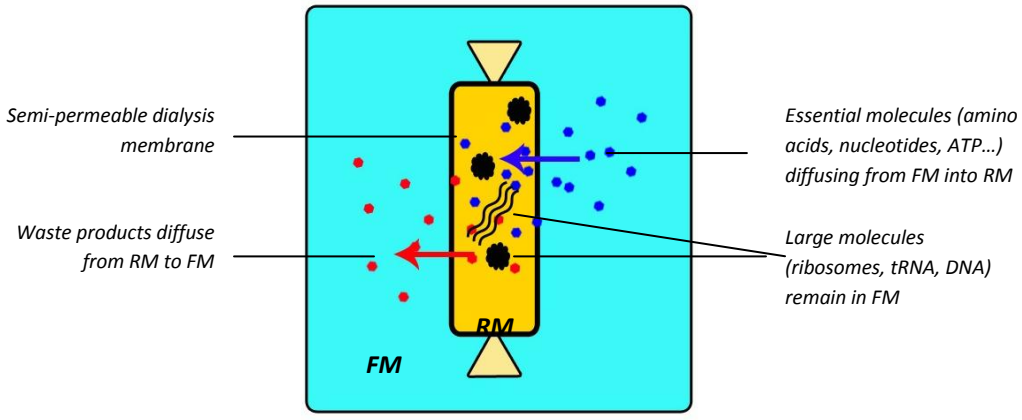



***Figure 1.1:*** *Diagram illustrating the CECF reaction; RM: reaction mix, FM: Feeder mix.*

Since depletion in the ATP generation pathway and the accumulation of inorganic phosphates are inhibitory, the developments of reaction methods that limit these problems are solutions for extending the life-time of CFPS and increasing yields of synthesised proteins (Kim and Swartz, 1999). Phosphoenolpyruvate (PEP), as well as being key component in the energy generation pathway in CFPS, is liable to degradation into inorganic phosphates and, moreover, is very costly (Calhoun and Swartz, 2005). The replacement of PEP with the more stable and cheaper glucose, glucose-6-phosphate or fructose-1,6-bisphsophate therefore not only improved the efficiency of CFPS, but also reduced its costs (Kim and Swartz, 2001, Calhoun and Swartz, 2005, Anderson *et al.*, 2015, Kim *et al.*, 2007). With the reported production of a maximum of 2.3 mg/mL of protein, the implementation of maltose- and maltodextrin-dependent recycling pathways enabled removal of the inhibitory inorganic phosphates, thus enabling longer lived and higher yielding batch CFPS reactions (Caschera and Noireaux, 2014, Caschera and Noireaux, 2015).

### 1.1.2. Current cell-free protein synthesis system

#### *The principle of CFPS*

With these developments cell-free expression systems are, today, a viable method for recombinant expression of proteins. The basis of CFPS relies upon harnessing the fundamental components of the transcription and translational machinery of the cell to produce the *de novo* protein of interest *in vitro* (Figure 1.2). As such, the cytosolic portion from highly metabolic cells containing the required protein machinery is extracted and used as the lysate. This machinery includes mainly polymerases, ribosomes, aminoacyl-tRNA synthetases, transcription and translation initiation, elongation and termination factors. Additionally, the system must also contain the critical ATP generating-recycling pathway to support the coupled protein transcription-translation. CFPS is then activated by the addition of the encoding nucleic acid template and small molecule reagents: nucleotides, amino acids, energy substrates, cofactors and salts. The reaction is sustained until depletion of one of the substrates, or inhibited by accumulation of by-products.

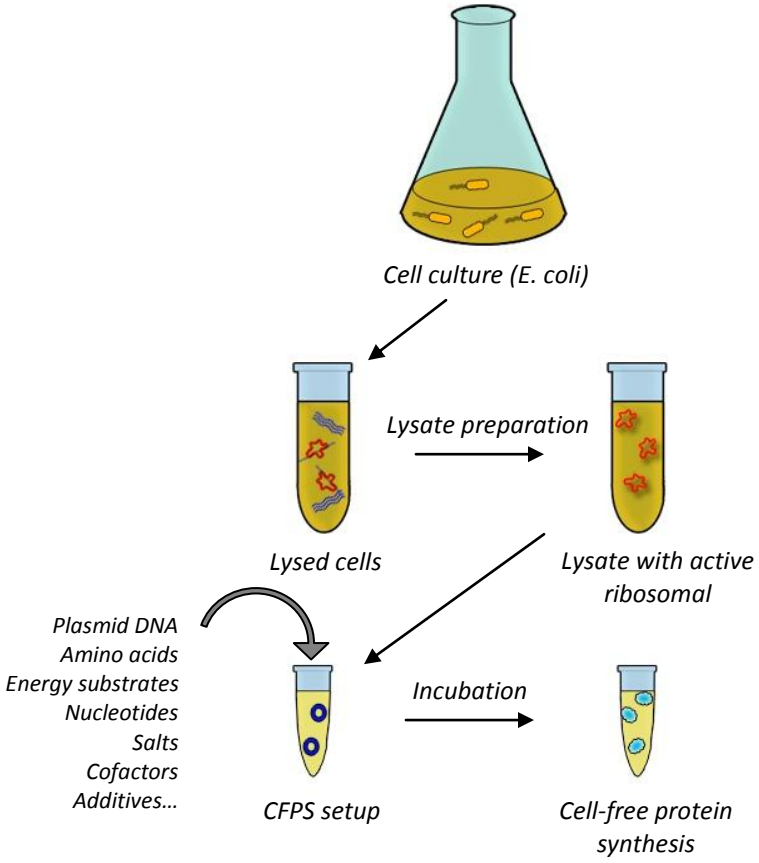



***Figure 1.2:*** *Typical cell-free reaction using E. coli lysate. E. coli cells are first cultured to adequate* $OD_{600}$ *(A). The culture is harvested, lysed and processed to eliminate genomic DNA, cellular membrane and other insoluble debris in order to generate cell-free lysate, which can be stored at -80*$^{o}$*C until further use (B). A CFPS reaction is prepared by adding plasmid DNA of target protein, nucleotides, amino acids, energy substrates, cofactors, salts, stabilising agents or other additives (C). The reaction is incubated at optimum temperature and time for production of protein, which can subsequently be purified.*

### 1.1.3. Types of cell-free lysates

#### *Prokaryotic CFPS*

In theory, the lysate of the cell-free can be extracted from any organism (Zemella *et al.*, 2015). The most popular lysate however remains the *E. coli* based system due to the relative simplicity of preparation and setup, high protein synthesis rate and yield, and economical aspects (Matthaei and Nirenberg, 1961, Pratt, 1984, Sun *et al.*, 2013). Expression yields in *E. coli* CFPS, however, are highly dependent upon the protein being produced and typically range between a few micrograms to milligrams per millitre of reaction. For instance, a study making use of dialysis CECF method reported the production of approximately 6 mg/mL of chloramphenicol acetyltransferase per millilitre of reaction (Kigawa *et al.*, 1999). Furthermore, as described in a later section, *E. coli* systems have demonstrated their potential for adequate scaling up for large production of protein (Zawada *et al.*, 2011a).

Of interest as well, archaeal extracts from thermophilic *Thermococcus kodakaraensis* were also used successfully in CFPS for the production of thermostable proteins which occur at high temperatures ($80^{o}$C) (Endoh *et al.*, 2007). However, prokaryotic derived lysates only allow limited post-translational modifications of the synthesised recombinant protein, which occasionally is essential for their structure and function (Braun and LaBaer, 2004).

#### *Mammalian CFPS*

CFPS developed from eukaryotes on the other hand permits post-translational modifications in expressed proteins. One of the very first proteins expressed in the cell-free system was radio-labelled haemoglobin which made use of extracts from rabbit reticulocyte microsomes (Schweet *et al.*, 1958). A disadvantage of the rabbit reticulocyte lysate system for large-scale CFPS remains the impracticality of animal tissue manipulation involving the extraction and purification of microsomes and, despite ingenious techniques such as the use of viral enhancers of translation to improve expression yields (Anastasina *et al.*, 2014), it remains a limited and costly method. Chinese ovary hamster (CHO) cell extracts, as an alternative for mammalian-based CFPS, having the capacity of coupled transcription-translation batch reaction to produce glycosylated human erythropoietin was recently described (Brodel

*et al.*, 2015, Brodel *et al.*, 2014). In these studies, performance of the CFPS system was increased through the use of expression vectors which carried viral internal ribosome entry sites. These vectors promote mRNA-ribosome interactions during the limiting translation initiation step – approximately 50 µg of active luciferase, for instance, can be produced per millitre of reaction (Brodel *et al.*, 2014). Finally, human derived cell extracts have also been used for coupled transcription-translation CFPS, which enables the production of correctly folded, glycosylated human proteins (Brodel *et al.*, 2015, Mikami *et al.*, 2010, Mikami *et al.*, 2008). Whilst the yields of the expensive recombinant mammalian based CFPS remains low, interestingly these systems were found to be well-adapted for production of entire viral particles in pathogenic studies on viral replication and assembly (Kobayashi *et al.*, 2014, Kobayashi *et al.*, 2007).

### *Yeast CFPS*

Yeast extracts from *Saccharomyces cerevisiae* have been an alternative to the *E. coli* S30 and mammalian CFPS systems since the 1970s (Sissons, 1974, Gallis *et al.*, 1975). The rationale for this fungal based approach was to provide a more efficient and cost-effective system for protein production bearing some post-translational modifications (Rothblatt and Meyer, 1986). The system has been improved over the decades: from yeast cell cultivation and extraction methods (Hodgman and Jewett, 2013) to the transcription-translation process (Hodgman and Jewett, 2014) and complementary energy regeneration (Anderson *et al.*, 2015) and removal of inhibitory by-products (Schoborg *et al.*, 2014). Although offering a less costly alternative to mammalian CFPS, and even with innovations, the yields of yeast CFPS remain low – less than 10 µg of luciferase per millilitre of reaction (Hodgman and Jewett, 2014, Anderson *et al.*, 2015). Furthermore, despite yeast CFPS enabling protein post-translational modifications, these do not correspond necessarily to those of mammalian cells (Garcia *et al.*, 2007).

### *Insect CFPS*

*Spodoptera frugiperda* insect cell extracts have been used for CFPS and, of particular interest in this work, resulted in the production of about 35 µg of the glycosylated form of HIV glycoprotein GP120 per millilitre of reaction (Tarui *et al.*, 2001). Insect cell based CFPS, as well as conferring post-translation modifications and enhancing correct folding of the expressed protein, is advantageous due to the preserved endoplasmic reticulum residues occurring as microsomes. These structures offer an intrinsic lipid membrane or protein translocation and embedding – of importance for membrane protein expression (Fenz *et al.*, 2014). Additionally, allowing for translocation of proteins, approximately 20 µg/mL of functional antibody fragments per millilitre of reaction was accumulated in microsomes – which could subsequently be relatively easily purified by centrifugation (Stech *et al.*, 2012). Pushing the limits of insect cell CFPS, by using the continuous exchange system, resulted in the yields of active recombinant tissue-type-plasminogen, a human protein enriched in disulphide bonds, to reach about 60 µg per millilitre of reaction. Whilst the costs are high and the yields of proteins expressed remain relatively low compared to *E. coli*, future developments in insect cell CFPS hold potential for the expression of active mammalian proteins.

### *Wheat-germ CFPS*

Wheat germ extracts for the cell-free expression, developed in the 1970s for studying protein translation (Hunter *et al.*, 1977, Roberts and Paterson, 1973), was soon used for the expression of mammalian albumin (Tse and Taylor, 1977). Following development in the optimisation of wheat-germ based CFPS (Madin *et al.*, 2000, Spirin *et al.*, 1988), Endo and colleagues reported the production of 10 mg of green fluorescent protein from only 1 mL of dialysis reaction which was sustained by a continuous supply of mRNA for over 14 days (Sawasaki *et al.*, 2002). Through further refinement of reaction conditions to enhance disulphide bond formation, wheat-germ based CFPS was successfully used to produce variable antibody fragments (Kawasaki *et al.*, 2003, Stech *et al.*, 2014). Wheat-germ CFPS was also eventually used the expression of functional malarial proteins, which are particularly complicated to produce through recombinant methods (Tsuboi *et al.*, 2008, Arumugam *et al.*, 2014). Whilst the wheat-germ extract provides a very powerful method for eukaryotic CFPS, it should be noted

that the lysate preparation step is however a tedious, laborious and rather costly procedure (Madin *et al.*, 2000). As a result, the large-scale implementation of wheat-germ CFPS reactions is restricted. Furthermore, certain post-translation modifications, such as glycosylation, do not occur.

### *Protozoan CFPS*

Protozoan extracts, which have been for cell-free expression since the 1960s to gain insights in the translation mechanism of the organisms (Chesters, 1968, Hoshino and Sugiyama, 1974), have been used more recently for expression of a family of mammalian GTPases proteins using proteozan *Leishmania tarentolae* lysate (Kovtun *et al.*, 2011). This system could hold the potential for further development in CFPS as more research is carried out.

### *PURE CFPS*

As can be seen from the series of CFPS systems described above, CFPS relies typically on the recovery of the transcriptional and translational machinery of a host cell. However, the extracted lysate includes a great deal of proteins which are not required for protein biosynthesis. These proteins can inhibit CFPS or hinder protein purification following expression. To ensure that only the vital components of transcription-translation are present, the "Protein Synthesis using Recombinant Elements" (PURE) system was developed (Shimizu *et al.*, 2001). The system comprises of only 32 proteins and was created by combining the individually expressed recombinant components of transcription-translation to enable CFPS. All of the proteins of the PURE system are histidine-tagged to facilitate their removal by nickel affinity and thereby leaving only the *de novo* purified protein behind. This system is currently commercialised as easy-to-use kits and is popular in the field of synthetic biology (Garamella *et al.*, 2016). However, the PURE system is very costly and protein yields remain low. As a result, scaling-up to large yields using the PURE system is completely impractical.

| CFPS Host | Advantages | Disadvantages |
|---|---|---|
| ***E. coli*** **(Prokaryotic)** | 1. Efficient and high protein yields<br>2. Relatively simple, cost-effective and high-yielding lysate production<br>3. Most cost-effective CFPS system<br>4. Easily modifiable system<br>5. Supports protein labelling<br>6. Well-established and studied system<br>7. Proven capability of scaling up CFPS reactions | 1. Limited post-translational modifications<br>2. Lack of intrinsic chaperones to permit folding of non-native proteins<br>3. Lack of intrinsic membrane protein supports |
| **Protozoan cell** | 1. Less expensive lysate production compared to other eukaryotic systems<br>2. Supports post-translational modification<br>3. Potential in scaling-up reaction | 1. Not well-established and poorly documented system<br>2. Limited knowledge on post-translational modifications |
| **Mammalian cell** | 1. Native system for the production of correctly folded mammalian proteins<br>2. Allows complete native post-translational modifications | 1. Low protein yields<br>2. Difficult and very costly upkeep for lysate preparation<br>3. Costly CFPS reactions<br>4. Highly sensitive system liable to CFPS failure |
| **Yeast cell** | 1. Capable of post-translational modification, including glycosylation<br>2. Relatively simple and high-yielding lysate production | 1. Low protein yields<br>2. Post-translational modifications on the proteins do not correspond to native mammalian modifications |
| **Insect cell** | 1. Easy and high-yielding lysate production<br>2. Supports post-translational modification<br>3. Presence of endogenous microsomes for support of membrane proteins | 1. Expensive cell cultivation and running costs for lysate preparation |
| **Wheat-germ** | 1. Highly productive for synthesis of complex proteins<br>2. Allows for post-translational modification<br>3. Permissible to correct protein folding | 1. Laborious and expensive lysate preparation<br>2. Post-translational modifications on the proteins do not correspond to mammalian native modifications<br>3. Relatively low protein yields<br>4. Lack of intrinsic membrane protein supports |
| **PURE system** | 1. Avoids lysate proteins contamination during protein purification<br>2. Histidine-tagged components facilitate purification<br>3. Enables resources of CFPS to be directed only towards transcription-translation of the recombinant protein | 1. Very expensive CFPS<br>2. Very laborious lysate preparation<br>3. Impractical for scaling-up |

***Table 1.1:*** *Summary of the advantages and disadvantages of using the various host cells for CFPS.*

### 1.1.4. Advantages of CFPS for protein expression for structural biology

Of particular interest in this thesis is the use of CFPS for the production of difficult-to-express and deuterated proteins for use in small angle neutron scattering (SANS). An extended description to the principle of SANS, and the relevancy of CFPS in its application, is described in the following chapter. Typically however, deuterated proteins for the purposes of structural biology are produced *in vivo* in *E. coli,* and if required can be expressed in eukaryote *Pischia pastoris* (Dunne *et al.*, 2017b). As described in the sections below, CFPS presents several advantages over the classic *in vivo* system for recombinant expression of both unlabelled and labelled proteins.

### *Simplicity and ease of use*

CFPS provides a very simple and easy means for the expression of recombinant protein once functional lysate has been acquired. Large batches of lysate can be produced within a week and stored at -80$^{o}$C for an extended period of time until required. As opposed to *in vivo* systems, well-established CFPS protocols describe quick setup procedures capable of achieving maximum protein expression within hours, and which can subsequently be immediately purified. For *in vivo* expression of proteins however, cells require to be transformed, cultivated and induced. Furthermore, unless secreted into the media, the proteins need be extracted from the cells and therefore require cell-lysis protocols that do not denature sensitive proteins. Additionally, since it does not rely on cellular viability, CFPS avoids the laborious and time-consuming task of maintaining healthy cells alive. For instance, the expression of protein in mammalian cells demands skilled personnel and highly specialised laboratories which need to conform to strict safety regulations, and have high associated running costs (Khan, 2013). *In vivo E. coli* expression systems, whilst being the most popular method for producing deuterated proteins, require an extensive adaptation period which can nevertheless be detrimental for protein production (Dunne *et al.*, 2017b, Haertlein *et al.*, 2016). CFPS, as this thesis will describe, presents a very effective and quick method for obtaining stable, deuterated proteins by the addition of commercially available deuterated amino acids to the reaction mixture.

### *Expression of difficult proteins*

Since CFPS does not rely on cellular viability, the system can be used for the expression of difficult-to-express or toxic recombinant proteins which would otherwise harm the host cell when expressed *in vivo.* Recombinant expression of onconase, a ribonuclease showing potential in cancer treatment, which was proving to be problematic due to its cytotoxic properties *in vivo*, could be successfully expressed using CFPS (Salehi *et al.*, 2016). Toxic pore forming thermostable direct haemolysin (TDH), which would normally puncture its host cell leading to death, was expressed in CFPS (Dondapati *et al.*, 2018). Wheat-germ based CFPS was not only used to produce the highly toxic PabI restriction endonculease, but was in addition used to produce selenomethionine labelled protein, which was crystallised and used for X-ray crystallography (Watanabe *et al.*, 2010). CFPS is also very well adapted for the co-expression of proteins which, as demonstrated for the production of the yeast tRNA methyltransferase heterodimer,

is only functional as a complex when co-translated (Matsumoto *et al.*, 2008). The difficulty for co-translation of protein *in vivo* arises from the need to co-transform and accurately control the expression of the various components. Due to the openness of the system, this is more readily performed using CFPS.

### *Modulation of microenvironment*

The lack of cellular boundaries in CFPS is one of its greatest strengths as it allows the direct alteration of the chemical environment for fine-tuning protein expression and folding. For instance, disulphide bond formation during recombinant protein expression in *E. coli* is limited due to the highly reducing cellular microenvironment (Baneyx and Mujacic, 2004). By treating the lysate with iodoacetamide and supplementing the reaction with the oxidising agent glutathione, the redox potential of the system was modified to produce active urokinase with correctly paired-disulphide bonds (Kim and Swartz, 2004). Directly adding chaperones to CFPS also provides an elegant support for the expression of correctly folded proteins. For example, the chaperone disulphide bond isomerase (Dsb) when added to *E. coli* based CFPS was used for disulphide bonded tissue plasminogen activator (Yin and Swartz, 2004). Additionally, synthetic chaperones, as described by Welsh and colleagues, were added to *E. coli* CFPS to enhance protein folding by simulating eukaryotic chaperone-assisted protein folding in the endoplasmic reticulum (Welsh *et al.*, 2011). Acting as artificial chaperones, amphiphilic polysaccharide nanogels when added to CFPS allow the synthesis of correctly folded protein (Sasaki *et al.*, 2011). Finally, as described in a later section, the addition of detergents or liposomes to CFPS helps to provide the hydrophobic support for the expression of challenging membrane proteins.

### *Protein labelling and unnatural amino acid incorporation*

The supplementation of the system with labelled amino acids instead of conventional amino acids to produce labelled protein, was used since the very early days of cell-free expression (Zamecnik *et al.*, 1948, Siekevitz, 1952, Littlefield *et al.*, 1955, Matthaei and Nirenberg, 1961). Nevertheless, due to inefficiency in exogenous amino acid incorporation and to poor protein yields, the labelled cell-free system was only really used for understanding the fundamentals of transcription-translation.

With the increase in performance however, CFPS was first used for the production of milligrams of stable $C^{13}/N^{15}$ isotopic-labelled model GTPase RAS protein which was first characterised by NMR spectroscopy (Kigawa *et al.*, 1999). The NMR profile of *in vitro* synthesised RAS produced protein showed very high consistency with its *in vivo* produced counterpart as it indicated uniformity in tertiary structure. Furthermore, CFPS represents a very efficient system for protein labelling since only the *de novo* synthesised target protein is being labelled. Suggesting its potential for high-throughput analyses, this feature was exploited when un-purified CFPS expressed labelled protein was identified by NMR spectroscopy directly in its extract (Guignard *et al.*, 2002).

Additionally, CFPS can be used for the incorporation of non-natural amino acids into *de novo* expressed proteins. For example, fluorotryptophan residues were successfully incorporated into cell-free expressed retinol-binding protein and gave highly sensitive NMR spectroscopy results (Neerathilingam *et al.*, 2005). Combining the advantages of CFPS, the toxic and unnatural amino acid canavanine (analogous to arginine) was incorporated into *de novo* synthesised protein (Worst *et al.*, 2015). Furthermore, selective labelling of proteins can also be applied in CFPS which could allow the identification of specific residues in the protein (Staunton *et al.*, 2006).

Deuteration of proteins, as a focus of this thesis and described extensively in the next chapter, is a labelling technique which has been performed advantageously using CFPS. Reaching approximately 95 % of deuteration, the large 800 kDa GroEL chaperone protein was successfully expressed using CFPS, with impressive yields of 300 µg per millilitre of reaction (Etezady-Esfarjani *et al.*, 2007b). CFPS for protein deuteration, as opposed to *in vivo* expression, can be optimised to avoid isotopic scrambling and produce homogenously labelled samples. Stereo-array isotope labelling (SAIL) is a CFPS technique that was developed to purposely enhance the resolution of NMR spectra derived from deuterated proteins as it allows the incorporation of deuterium in amino acid residues to be controlled accurately (Kainosho *et al.*, 2006). The success of the CFPS-based system was demonstrated by subsequently coupling it with automated NMR structure determination software (Takeda *et al.*, 2007).

### *CFPS for membrane protein production*

Accounting for nearly 20 – 30 % of the prokaryotic and eukaryotic genome (Wallin and von Heijne, 1998), membrane proteins play critical roles in physiology and represent the target of more than 50 % of modern drugs (Overington *et al.*, 2006). Elucidation of membrane protein structures is therefore of major importance with the growing field of rational drug design (Cooke *et al.*, 2015). Nevertheless, the number of membrane protein structures determined remains limited: in 2018, of the total 140,591 structures deposited in the RCSB PBD database, 781 of these are membrane proteins – a mere 0.55 % (White, 1998 - 2018). In spite of the increase in the number of membrane protein structures determined in the past few years, structural analysis of membrane proteins remain very challenging.

Extracting membrane proteins from natural sources is usually difficult and sufficient amounts of material are rarely obtained for structural studies. Over-expression of recombinant membrane proteins, the obvious solution to the problem, however remains challenging due to the hydrophobic nature of membrane proteins and their correct folding (Wagner *et al.*, 2006). Eukaryotic membrane proteins follow a complex folding pathway through the Golgi apparatus where it is stabilised by the lipid bilayer prior to insertion into the plasma membrane (Grisshammer, 2006). Recombinant expression of these membrane proteins *in vivo* using *E. coli* is generally not possible due the reducing microenvironment, the lack of post-translational folding apparatus and stabilising lipid bilayers. The hydrophobic protein is therefore liable to undergo unfolding, aggregation, proteolysis and is occasionally toxic to the host cell. The accumulation of the aggregated proteins results in the formation of inclusion bodies, from which protein can only be extracted after an extensive refolding process, of which yields and protein integrity can still be poor (Yamaguchi and Miyazaki, 2014). *In vivo* over-expression of eukaryotic membrane proteins in mammalian cells whilst being of good quality is generally restrictive in terms of yields and it remains a costly technique. In the case of CFPS, the translation machinery is typically about 20-fold more dilute than in the cell, decreasing the rates of protein production. CFPS decreases the probability of undesirable, non-specific inter-peptide chain contacts, thereby increasing the probability of proper folding and decreasing the probability of aggregation. Considering the advantages of CFPS, the system presents itself as a potentially useful tool for the expression of recombinant membrane protein for the purposes of structural biology.

Termed as precipitate cell-free (P-CF) expression (Figure 1.3, A), the optimised expression of a membrane protein directly into the reaction is possible and as expected, due to the lack of stabilising

hydrophobic support, precipitation occurs (Schwarz *et al.*, 2008, Liguori *et al.*, 2007). Since cellular viability is not a concern in CFPS, the occurrence of the precipitate in not an issue. However, as opposed to the inclusion bodies formed during *E. coli* expression, the precipitate can be more easily solubilised and refolded using mild detergents. To demonstrate this $^{15}$N labelled *E. coli* multidrug transporters have been optimised for expression as precipitate in CFPS, solubilised using mild detergents and analysed by both CD and NMR spectroscopies for structural insights (Klammt *et al.*, 2004b). The key benefit of CFPS for membrane proteins is the openness of the system which allows inclusion of a hydrophobic support for membrane protein folding and stabilisation (Schwarz *et al.*, 2008, Liguori *et al.*, 2007). In the detergent cell-free (D-CF) (Figure 1.3, B), mild detergents which are tolerated by CFPS, can allow the direct solubilisation of the membrane protein as translation proceeds. Following extensive screens to identify suitable reaction conditions, the detergent type and its working concentration, a functional and solubilised Rho-tagged GPCR could be expressed and subsequently purified directly from the reaction mixture using antibody affinity (Kaiser *et al.*, 2008). Nevertheless, it is possible for detergent solubilised membrane proteins from both P-CF and D-CF modes of expression to be reconstituted into liposomes (Klammt *et al.*, 2004b). However, in the liposome cell-free (L-CF) (Figure 1.3, C), the inclusion of liposomes into the reaction provides the support such that the membrane protein can be inserted directly into the lipid bilayer and purified as proteoliposomes (Soranzo *et al.*, 2016, Niwa *et al.*, 2015). The hepatitis C viral porin protein p7 was produced using this method and coupled to neutron reflectivity enabling *in situ* observation of the porin insertion into the lipid bilayer was made (Soranzo *et al.*, 2017). However, the recuperation of proteoliposomes from CFPS is problematic due to the non-homogeneity of the proteoliposomes and the co-purification of many non-specific proteins from the lysate (Nozawa *et al.*, 2011).

Furthermore, highly homogenous nanodiscs, which consist of lipid bilayers bound by ring proteins, can also be used in CFPS systems to allow co-translational insertion of the *de novo* synthesised membrane protein (Figure 1.3, D) (Roos *et al.*, 2014). Similarly, bicelles, which are detergent-stabilised lipid bilayers can also provide the structural support for the insertion of CFPS membrane proteins (Figure 1.3, E) (Lyukmanova *et al.*, 2012, Laguerre *et al.*, 2016). With the development of chemically engineered polymers such as amphipols, which directly provide the support for membrane proteins, (Figure 1.3, G) (Bazzacco *et al.*, 2012) and Lipodisqs®, which consist of a polymer-stabilised lipid bilayer, (Figure 1.3, F) (Dorr *et al.*, 2014), further membrane mimetics can be used in CFPS to provide support for the *de novo* synthesised membrane protein. Additionally, if successful in CFPS, all of these novel mimetics can facilitate purification of highly homogenous CFPS membrane proteins.

Nevertheless, with the expression of CFPS membrane proteins as proteoliposomes being the expertise of Synthelis, and protocols suitably adapted for the purpose, the work of this thesis considers mostly the L-CF format mode.

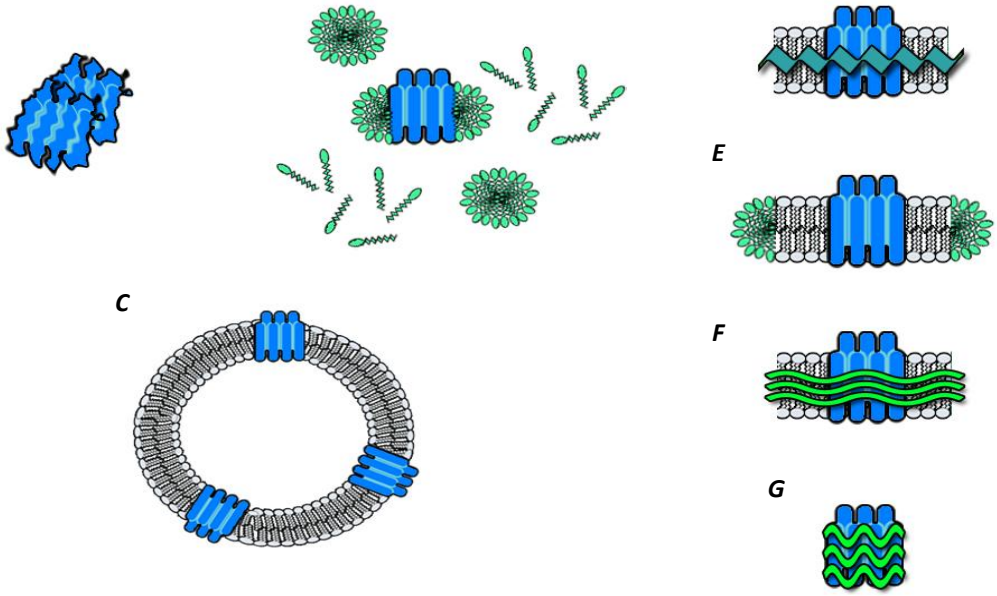



***Figure 1.3:*** *The various modes of membrane protein expression in CFPS (not to scale). Precipitate cell-free (P-CF), involving the direct synthesis of the membrane protein in CFPS without any hydrophobic support, results in the precipitation of the target protein (A). Detergent cell-free (D-CF) includes mild detergents in reaction mix, which allows solubilised in-situ expression of the membrane protein (B). Under optimised condition, lipid cell-free (L-CF), including liposomes (C, spherical lipid bilayer), nanodiscs (D, circular lipid layer surrounded by ring of protein), bicelles (E, circular lipid layer surrounded by ring of detergent), or Lipodisqs® (F, circular lipid layer surrounded by ring of styrene-maleic acid copolymer), allows for the direct incorporation of the folded target protein into the lipid membrane. Still under study, the use of CFPS-compatible amphipols provides direct support for membrane protein folding without the need for the lipid bilayer (G, synthetic amphiphilic polymer surrounding membrane protein).*

## 1.2. Model CFPS target proteins for structural studies

CFPS, as presented above, is a powerful tool that remains to be fully exploited for the purposes of structural biology. CFPS systems are capable of producing high amounts and high quality proteins. Taking advantage of the system openness, the chemical environment can be modified to optimise and enhance the system, which is limited *in vivo*. Although capable of generating any protein, CFPS is well-adapted for the production of difficult to express proteins which fail *in vivo*. Furthermore, with a direct application in structural biology, CFPS can be easily adapted for labelling *de novo* synthesised protein.

In this work three model proteins were used to further develop and test CFPS as a routine tool for production of proteins in quantity and quality for structural biology work. Particularly in this context, a further aim was to demonstrate that the CFPS-produced proteins have the same functionality and structure as those produced using conventional *in vivo* methods. The model proteins used were variants of CXCR4, its ligand SDF1-α and CD4. CXCR4 is a membrane protein which is a difficult to express *in vivo* in *E. coli* since it is typically occurs in low amounts and as insoluble bodies. SDF1-α although expressed in relatively large amounts in *E. coli*, occurs strictly as insoluble inclusion bodies that require denaturation followed by refolding (Sadir *et al.*, 2001). CD4 is a transmembrane protein which serves as the primary receptor for HIV. Due to the implication of mostly the first two domains of CD4 with respect to HIV binding, a truncated variant of CD4, referred to as 2dCD4, consisting of these two domains has been considered in the majority of studies (Ryu *et al.*, 1994, Kwong *et al.*, 1998, Cerutti *et al.*, 2010). Nevertheless, the expression of 2dCD4 in *E. coli* results in insoluble inclusion bodies. Despite occurring in large amounts, the recovery of 2dCD4 after denaturation and extensive refolding procedure remains very low (Cerutti *et al.*, 2010, Cerutti *et al.*, 2014a). As such, these proteins represent an array of moderately to very challenging to express protein *in vivo*.

The use of CFPS here, as it will be described in this thesis, serves to demonstrate the system is not only an alternative for facilitating the production of difficult-to-express proteins, but can also be used routinely for the expression of any protein.

### 1.2.1. Roles of CXCR4 and SDF1-α

G-protein – coupled receptors (GPCRs) are the largest family of membrane proteins and are the products of almost 800 human genes (Bjarnadottir *et al.*, 2006). GPCRs play a diverse variety of roles in physiology which include vision, smell, taste, immunity, neurological, cardiovascular, reproductive and endocrine functions (Katritch *et al.*, 2013). Located on the plasma membrane, GPCRs are activated by the binding of their effector ligands, which subsequently leads to signal transduction and activation of cellular functions. As such, GPCRs are the target of approximately 30 % of all drugs globally, which between 2011 – 2015 was estimated to have a market value of almost $ 890 billion (Overington *et al.*, 2006, Hauser *et al.*, 2017). Driven by both the therapeutic and economic importance, the rational design of new drugs targeting GPCRs requires detailed understanding of the structure, modulation and function of the receptor. In 2007, the first crystal structure of GPCR was finally determined using recombinant human GPCR β2 adrenergic receptor (Rosenbaum *et al.*, 2007), and which eventually culminated in its authors winning the Nobel Prize for Chemistry in 2012. Since then, there has been a rapid increase in the number of GPCR X-ray crystal structures which is leading the way towards rational drug design in this class of proteins (Cong *et al.*, 2017). However, structural studies with membrane proteins remain very challenging due to the difficulties in producing the hydrophobic, flexible and lipid-associated proteins in the first place. Indeed, in order to obtain crystal structures of GPCRs, the underlying protein sequences were heavily mutated. For example, to determine the β2 adrenergic receptor the structure was stabilised by the replacement of its third intracellular loop with a lysozyme sequence and by truncating the C-terminal (Rosenbaum *et al.*, 2007).

The chemokine receptor CXCR4 (fusin or CD184) is a GPCR that plays a critical role in chemotaxis – the migration of cells towards a gradient of its ligand, stromal cell-derived factor 1 (Figure 1.4, A) (SDF1-α or CXCL12) (Domanska *et al.*, 2013). A canonical role of CXCR4, present on lymphocytes and monocytes, relates to the movement of the white blood cells towards areas of inflammation where SDF1-α is secreted from stromal cells (Bleul *et al.*, 1996, Loetscher *et al.*, 2000). This is also the case during neural inflammation where white blood cells are able to cross the blood-brain barrier (Engelhardt and Ransohoff, 2012, Guyon, 2014). Other than lymphocytic action in the brain, CXCR4 and SDF1-α are constitutively expressed in neuronal cells which enable their function in nerve transmission and plasticity (Guyon, 2014). During embryological development, SDF1-α and CXCR4 are responsible for the proper migration of a number of precursor cells which is indispensable for foetal life. For instance,

mutant mice embryos lacking the CXCR4 and SDF1-α could not sustain normal lymphopoiesis, myelopoiesis, neural and cardiac development which eventually lead to prenatal death (Nagasawa *et al.*, 1996, Zou *et al.*, 1998).

CXCR4 – SDF1-α is also involved in cancer biology (Domanska *et al.*, 2013). Cancer cells over-express CXCR4 which remains constantly activated by SDF1-α - secreting tumour associated stromal cells (Figure 1.4, B). This situation results in a positive feedback form of paracrine signalling which maintains tumour survival and growth (Orimo *et al.*, 2005). The localised over-expression of SDF1-α in the tumour furthermore leads to the recruitment of native CXCR4 haematopoietic which leads to the development of feeder blood vessels to sustain cancer growth (Jin *et al.*, 2006). With the tumour enriched with blood vessels, metastasis is promoted as detached cancer cells are then able to enter the blood stream and circulate the body until they eventually latch onto a distant organ (Muller *et al.*, 2001). The drug AMD-3100, which is a potent antagonist of CXCR4, targets this CXCR4, disrupting the cycle and interrupting cancer proliferation (Hendrix *et al.*, 2000). Another pathological role of CXCR4, and of interest for further work, is the involvement of CXCR4 as the secondary receptor required for HIV infection (Figure 1.4, C and Figure 1.5, C) (Murakami and Yamamoto, 2010). Whilst this work was focused on the expression of CD4 as the primary receptor for HIV, it is important to appreciate the prospect to demonstrate CD4-GP120 complex binding to CFPS CXCR4.

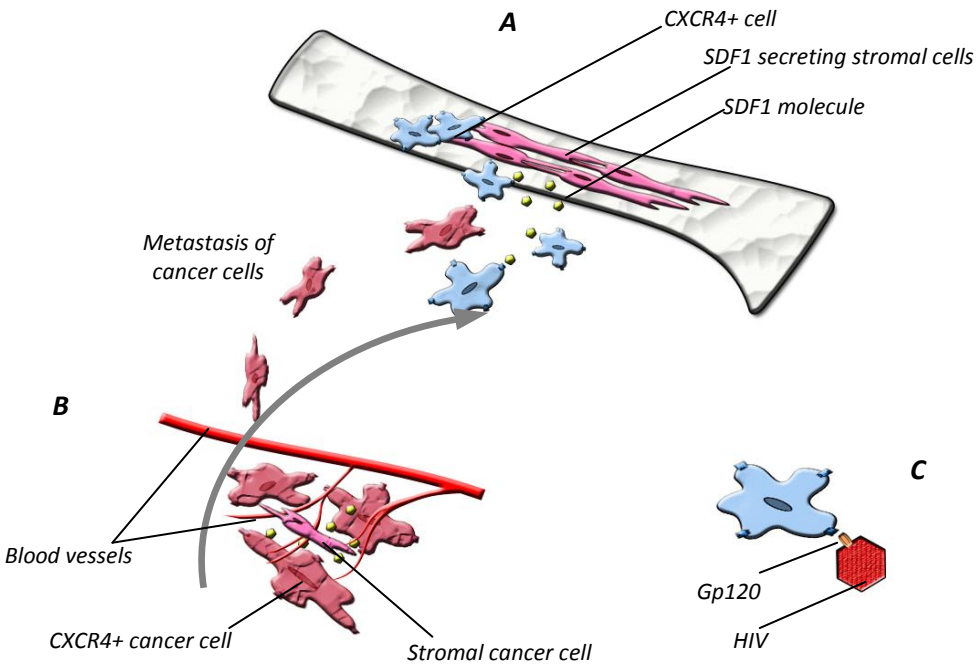



***Figure 1.4:*** *The roles of CXCR4. CXCR4, and its natural ligand SDF1, are typically involved in chemotaxis, homing and sequestration of circulating CXCR4-bearing cells to its corresponding niche – for instance bone marrow (A). Pathological CXCR4-bearing and stromal cancer cells drive the proliferation of the tumour cells and promote aberrant angiogenesis at the tumour site (B). By means of chemotaxis, circulating CXCR4-bearing cancer cells can infiltrate distant organs, encouraging cancer metastasis (C). During of HIV infection, viral gp120 protein uses CXCR4 as its secondary co-receptor for gaining entry into CXCR4-bearing cells (C).*

### 1.2.2. Roles of CD4

The human T-cell surface glycoprotein cluster of differentiation 4 (CD4; previously referred to as T-cell surface antigen T4/Leu-3) is canonically located on CD4+ T-cells, but is also present on certain macrophages, dendritic cells and monocytes. Under normal physiology, the CD4 receptor plays a fundamental role in the formation of the immunological synapse between the antigen presenting cells (APCs) and T-cells (Figure 1.5) (Sweet *et al.*, 1991, Bromley *et al.*, 2001, Yin *et al.*, 2012). During infection, as part of the innate immune response, APCs such as macrophages phagocytose the invading pathogen. The phagocytosed material is digested intracellularly, with constituent proteins being cleaved to small peptides. These peptides associate with major histo-compatibility complex class II (MHCII) molecules, which are transported to the cell surface where the MHCII is embedded/inserted into the cell surface membrane. The MHCII thereby presents the peptide to circulating T-cells. T-cells bind to peptide-MHCII complexes via the T cell receptor (TCR), and recognise the peptide as foreign. This interaction and the resulting transduced signal is however weak. The immunological synapse is stabilised and the transduced signal reinforced by CD4, which binds to MHCII following the initial, kinetically ‘’loose” association between the TCR and MHCII, thereby activating tyrosine kinase and a molecular cascade which leads to the release of cytokines.

Of consideration in this project however, is the implication of CD4 in HIV infection, with regards to its binding to GP120 (Figure 1.5). The first step of HIV infection of host cells involves the attachment of the viral envelope glycoprotein GP120 to CD4. Following the binding of CD4 to its corresponding binding site, GP120 undergoes conformational changes leading to exposure of several conserved sites on the GP120 termed CD4-induced binding sites (CD4i), including the highly-conserved co-receptor binding site - either CXCR4 or CCR5 (Rizzuto *et al.*, 1998). Conformational changes also occurring in CD4 lead to the approach and re-orientation of the GP120 co-receptor binding site to its target (Yachou and Sekaly, 1999). Furthermore, the protruding N-terminus and extracellular loops of the transmembrane co-receptor procure the contact points for the GP120 interaction (Lee *et al.*, 1999). The binding of CD4 to GP120 therefore provides strong indication in the structural and functional integrity of the CFPS-produced CD4.

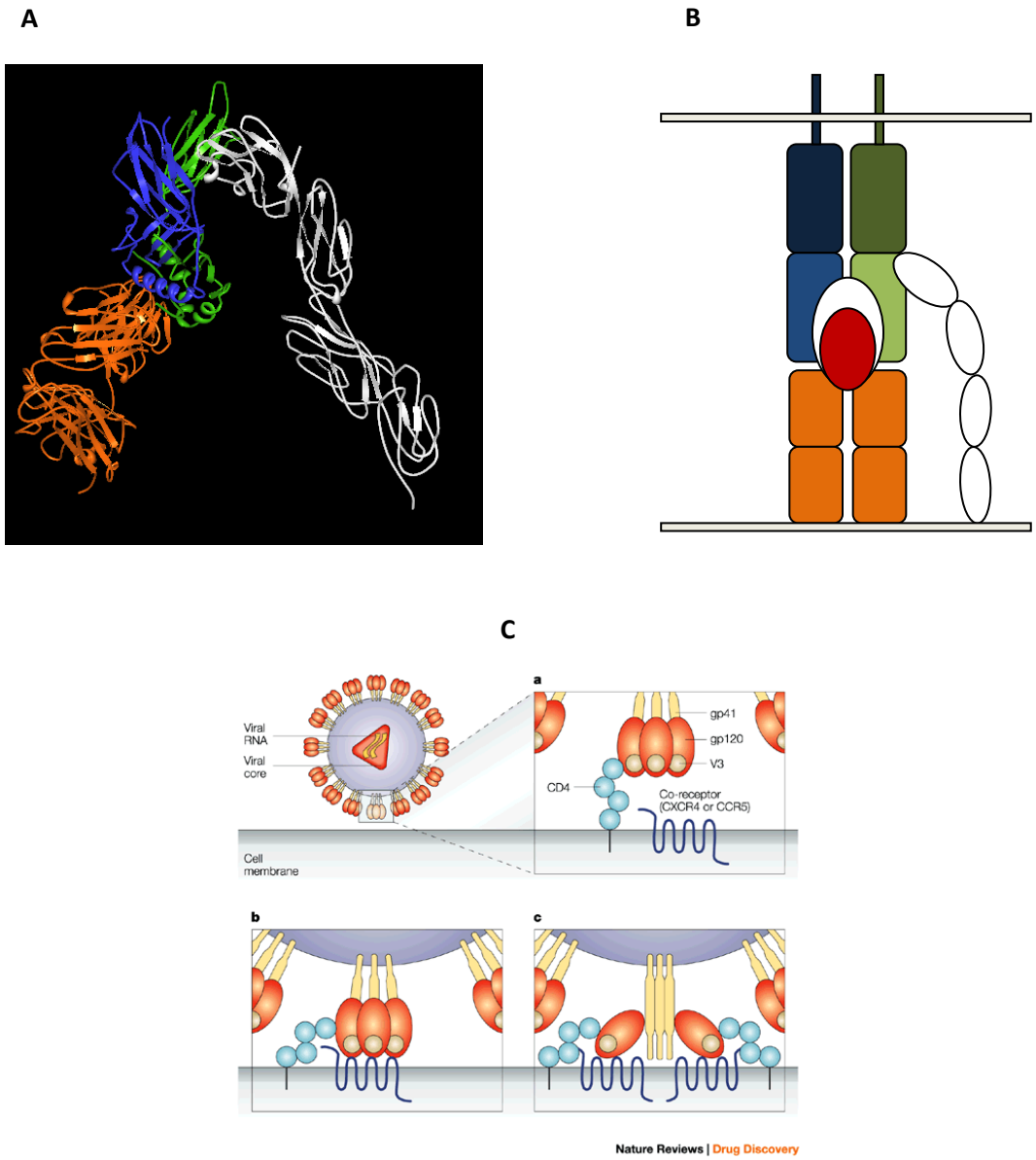



***Figure 1.5:*** *The interactions of CD4. Crystal structure of CD4, MHCII and TCR complex, the cornerstone of the immunological synapse is shown in (A) (From Yin et al., 2012 (PDB 3T0E)). The α2 domain of MHCII is shown in green; β2 domain of MHCII in blue; 2dCD4, in white; and T-Cell Receptor in orange. A schematic of the immunological synapse comprising the TCR-MHCI-CD4 complex is shown in (B). The elongated CD4 makes contact with MHCII as it presents the antigen (red circle) to the TCR. This synergy allows the reinforcement of the transduction signal to initiate the T-cell response. The interaction of CD4 and CXCR4 with viral GP120 during HIV cell-entry is shown in (C) (https://hivmedicalbreakthrough.weebly.com/a-problem.html).*

## 1.3. CFPS development for industrial application

The cell-free expression system, since first developed by Matthaei and Nirenberg in 1961 and standardised by Pratt and his colleagues in 1984, has come a long way (Matthaei and Nirenberg, 1961, Pratt, 1984). With the continuous improvement in the most popular *E. coli* based CFPS system, boosting yields and reducing running costs over the past decades (Figure 1.6), the move of CFPS towards industry, as a highly advantageous and profitable alternative to *in vivo* expression, becomes a natural step.

The first hints towards industrial application of CFPS began with the establishment of standardised protocols for the production of robust lysate (Liu *et al.*, 2005). Preparation of the cell-free extract, the most labour intensive and time-consuming process of CFPS, is also perhaps the most important step since high-quality lysate ensures high-quality protein production. The steps to improve and simplify lysate production, whilst reducing the costs and the time required for its preparation are of obvious importance for CFPS application in industry (Krinsky *et al.*, 2016). To this date, Krinsky and colleagues describe the most simplified, cost – and time – effective protocol for lysate production. For instance, by omitting centrifugation, pre-incubation and dialysis steps, extract could be obtained in about an hour following *E. coli* cell harvest. The cost of the CFPS was further reduced as they also demonstrated that their new protocol could sustain recombinant expression of proteins despite not supplementing the reaction with several reagents, including the expensive tRNA, phosphoenol phosphate (PEP) and AMP. As such, Krinsky and his colleagues claim a cost of $ 22 per millilitre of reaction, which represents approximately an 80 % reduction running costs compared to previous typical CFPS reactions. Even further reduction is cost can also be brought about with further development of CFPS which use glucose instead of costly PEP as the energy source for driving the reaction (Anderson *et al.*, 2015, Calhoun and Swartz, 2005).

Scaling up the CFPS reaction was the next task for industrial application, which was impressively demonstrated by Zawada and colleagues over the past years, and which culminated in the large-scale expression and purification platform at SutroBiopharma Inc (Zawada *et al.*, 2011a). As an example, model cytokine human granulocyte-macrophage colony-stimulating factor (hGM-CSF) was expressed using *E. coli* lysate based CFPS. As opposed to *in vivo E. coli* expression, the open CFPS system can be modulated to enable redox conditions for the formation of critical disulphide bonds that permit correct protein folding (Baneyx and Mujacic, 2004). Firstly, by using fermentors for culturing the *E. coli* cells, large amounts of lysate could be produced (Zawada *et al.*, 2011a, Liu *et al.*, 2005, Zawada and Swartz,

2005). Bioreactors of 200 L capacity were used for culturing large amounts of *E. coli* cells, which were subsequently successfully harvested and lysed using industrial scale centrifuges and homogenisers. With about 140 g of cells (wet weight) per litre of culture, approximately 1.1 L of lysate could be produced per kilogram of cells. Additionally, fermentors allow the careful monitoring and control of bacterial growth in order to obtain high quality, robust and reproducible lysates. In parallel, large scale production of plasmid DNA required for CFPS was also established (Zhang *et al.*, 2003). Once lysate and DNA were produced, starting from 250 µL, CFPS reactions were then successfully and sequentially setup in bioreactors to attain 100 L volume. Following confirmed expression of soluble protein, the hGM-CSF was purified directly from the reaction mix with yields of 700 mg per litre of reaction.

CFPS also represents a powerful technique for the large-scale synthesis of protein libraries for high-throughput structural and functional analyses. The lack of cellular boundaries, direct use of polymerase chain reaction (PCR) templates, relatively cheap multi-well reactions and the potential of miniaturisation and automation are characteristics which are readily exploited (Endo and Sawasaki, 2004, He and Taussig, 2007).

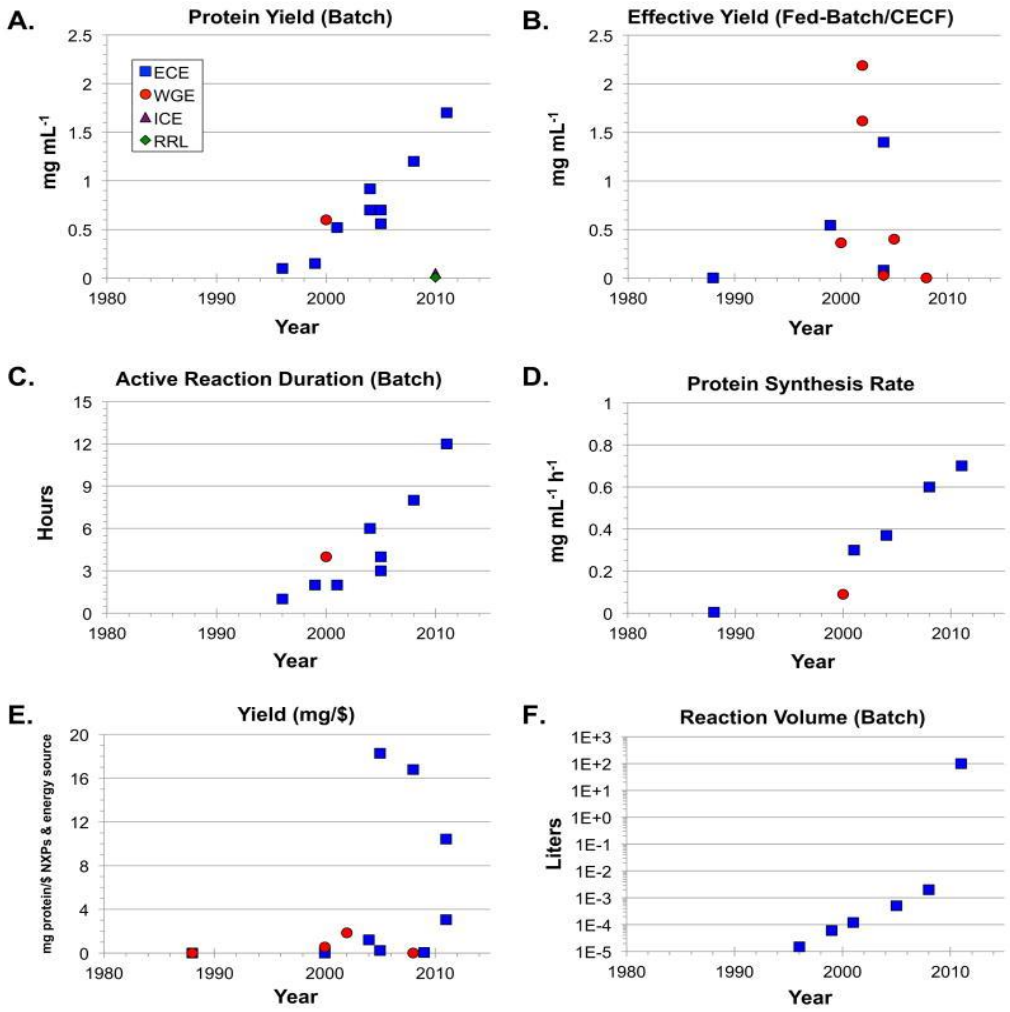



***Figure 1.6:*** *Historical trends of CFPS leading towards industrial application, from Carlson et al. "Cell-free Protein Synthesis: Applications come of Age" Biotechnology Advances, 2012 (Carlson et al., 2012a). Blue squares, E. coli lysate (ECE); red circles, wheat germ lysate (WGE); purple triangles; insect cell lysate (ICE); green diamonds, rabbit reticulocyte lysate (RRL). (A) Cell-free protein synthesis yields for a batch reaction. (B) Cell-free protein synthesis yields for a fed-batch or continuous exchange cell-free (CECF) reaction, relative to the total volume of reaction and feeding solutions. (C) Reaction length for active protein synthesis in a batch cell-free reaction. (D) Rate of protein synthesis during a cell-free reaction. (E) Protein yield per dollar of substrate reagents (F) Scale of cell-free reaction volumes.*

## 1.4. Overview and aims of the project

This project was funded by the CIFRE (Convention Industrielle de Formation par la Recherche), which aims to foster the collaborative relationship between industrial and academic partners. As such, the project was driven by the research interests of the small biotech company Synthelis (France), a spin-off from the University of Grenoble Alps (UGA, France), which specialises in the expression of membrane and other difficult-to-express proteins using patented CFPS technology. Over the course of the years, Synthelis has been driven towards the structural characterisation of its CFPS-generated proteins. Primarily, this would provide confirmation of the quality of the CFPS-generated proteins as equivalent to proteins produced using classical *in vivo* methods. Thus, backed by advanced structural characterisation, this would lend confidence to the Synthelis CFPS system and its products, thereby reinforcing customer interest. Synthelis also has interest to demonstrate the application of CFPS for routine protein production. This could have strong application in the context of the production of functionally active soluble proteins which if produced under traditional *in vivo* expression systems require complex protein recovery and purification steps (e.g. refolding).

To attain its goal in structural characterisation, the project involved extensive collaboration with, and use of resources available at the Institut Laue-Langevin (ILL, France) and the European Synchrotron Radiation Facility (ESRF, France). Additionally, with the close association of CXCR4 and CD4 to HIV entry, collaboration with the HIV Pathogenesis Research Laboratory (HIVRPL) from the University of Witwatersrand (South Africa) was established.

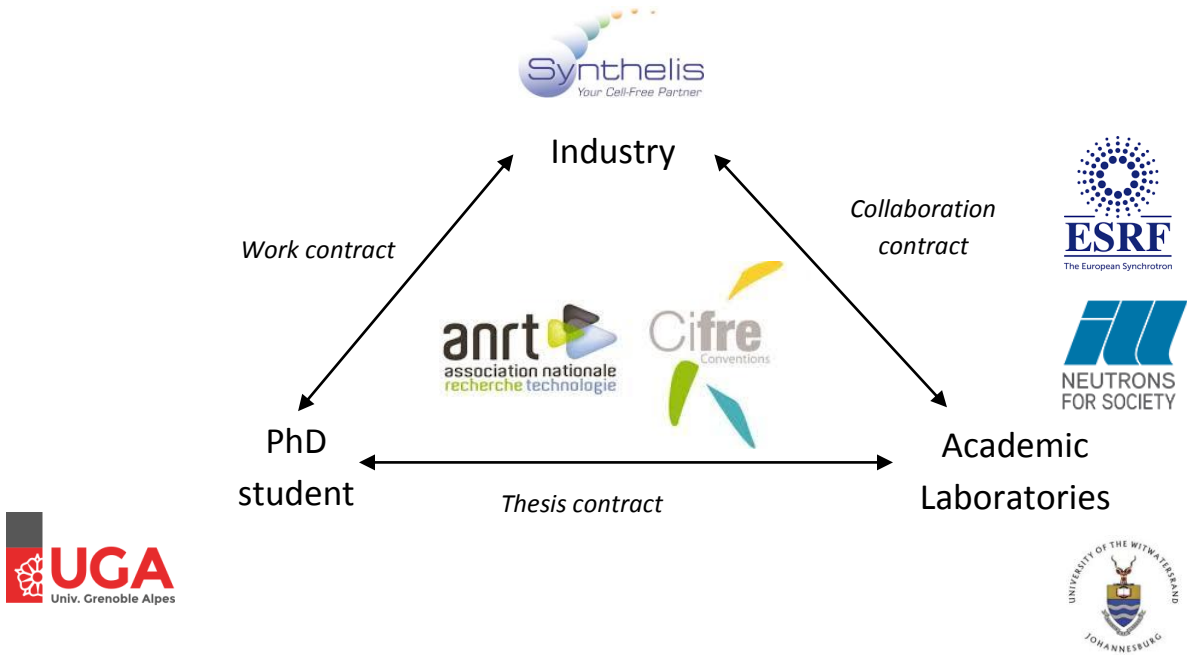



***Figure 1.7:*** *The collaboration and engagements of the industry, academia and PhD student as part of the ANRT/CIFRE scheme.*

The overarching aims of this project were therefore

i) To optimise the Synthelis CFPS methods and procedures for the production of proteins in suitable quantity and quality for advanced structural biology.

ii) To demonstrate the application of CFPS as a potential tool for routine protein production.

iii) To perform biophysical characterisation and advanced structural validation of selected model proteins as produced by the optimised CFPS to confirm their structural equivalence to those produced under recombinant *in vivo* expression conditions.

To achieve these overall aims, the detailed objectives of the project are given below in the context of the thesis chapters:

### *Chapter two: Bacterial culture and production of lysates*

- *Establishment of growth conditions for novel bacterial strains for the use as CFPS lysates, and production of lysates using Synthelis-based protocols*
- *Optimising, validating and confirming the CFPS expression of model membrane proteins in liposomes using novel lysates*
- *Scaling-up the production of lysate by developing high-density fermenter cultures*
- *Adapting protocol from initial high-density fermenter cultures for production of deuterated lysate in view of use for SANS analysis*
- *Optimising, validating and confirming the CFPS expression of model membrane proteins in liposomes using density fermenter-based lysate*
- *Overall comparison of the yields of expression using the various lysate*

### *Chapter three: Optimisation of protein expression and purification*

- *Optimising, validating and confirming the CFPS expression of CXCR4 in liposomes using novel lysates*
- *Overall comparison of the yields and quality of expression using the various lysates for CXCR4 as proteoliposomes using novel lysates and selection of optimum lysate*
- *Understanding and optimising methods for increasing yields and purity of CXCR4 as proteoliposomes*
- *Optimising, validating and confirming the CFPS expression of full-length CD4 in liposomes*
- *Optimising, validating and confirming the soluble CFPS expression of CXCR4 in detergents*
- *Optimising and selection of methods for increasing yields and purity of soluble CFPS CXCR4 in detergents*
- *Scaling-up production of soluble CFPS CXCR4 in detergents and purification*
- *Optimising, validating and confirming the soluble CFPS expression of SDF1-α*
- *Optimising, validating and confirming the soluble CFPS expression of deuterated SDF1-α*

- *Small-scale purification of soluble CFPS SDF1-α*
- *Scaling-up production and purification of soluble CFPS SDF1-α*
- *Optimising, validating and confirming the soluble CFPS expression of 2dCD4*
- *Optimising, validating and confirming the soluble CFPS expression of deuterated 2dCD4*
- *Small-scale purification of soluble CFPS 2dCD4*
- *Scaling-up production and purification of soluble CFPS 2dCD4*
- *Culturing, expression and purification of GP120 from mammalian cells*
- *Forming the complex of deuterated 2dCD4-GP120 for the purpose of SANS analysis*
- *Establishment of work and standard operating protocols for the CFPS expression of protein within Synthelis CFPS technology*

### *Chapter four: Biophysical and functional characterisation CFPS proteins*

- *Understanding the changes in liposomes and CXCR4-proteoliposomes using the NanoSight™ Particle Tracker*
- *Transmission electron microscopy (TEM) of CXCR4-proteoliposomes*
- *ELISA and SPR to confirm antibody binding to CXCR4-proteoliposomes*
- *ELISA to confirm antibody and GP120 binding to CD4-proteoliposomes*
- *Mass-spectrometry of CFPS unlabelled and deuterated SDF1-α to confirm state of disulphide bonds and deuteration levels*
- *Chemotaxis assay of unlabelled and deuterated SDF1-α to confirm functionality*

### *Chapter five: X-ray crystal structure of cell-free SDF1-α*

- *Crystallisation of CFPS SDF1-α*
- *X-ray diffraction of SDF1-α crystal at the MASSIF-1 instrument (ID30A-1, ESRF)*
- *Validating the structural state of CFPS SDF1-α compared to previously published E. coli SDF1-α*

### *Chapter six: Small-angle neutron scattering of CFPS-produced 2dCD4 in complex with GP120*

- *Preparation of 2dCD4-GP120 samples at match-out conditions*
- *SANS analysis of 2dCD4-GP120 samples at KSW-2 instrument (FRMII, Munich)*
- *Analysis and interpretation of SANS scattering data*
- *Generation of models for 2dCD4-GP120*

### *Chapter seven: Feasibility study of direct SANS structural analysis of a model protein within the unfractionated CFPS reaction mixture*

- *Preparation of samples containing model proteins in the presence of increasing lysate amounts at match-out conditions*
- *SANS analysis of samples at D22 instrument (ILL)*
- *Analysis and interpretation of SANS scattering data despite the presence of lysate*
- *Verifying the ability to generate of models despite the presence of lysate*

# CHAPTER TWO

## BACTERIAL CULTURE AND PRODUCTION OF LYSATES

## ABSTRACT

A critical task in CFPS is the production of the lysate. In order to improve the intrinsic qualities of the CFPS lysate with the goal of improving the quality and yields of the CFPS proteins, lysate was produced using various bacterial strains: Nico21, C43 and Rosetta, Rosetta-Gami and SHuffle. With the needs of standardisation and protocol reproducibility, the growth cultures and biomass yields of these new strains were recorded and compared with each other. To cope with an increased demand of protein produced by CFPS, and with the need for deuterated lysate for the purpose of SANS, lysate production using high-cell density cultures (HCDC) in fermenters was achieved. Lysates thus produced were hydrogenated cell-free lysate (LH), perdeuterated cell-free lysate (PL) and matchout cell-free lysate (LM). As opposed to classical growth in flasks where a maximum of 6 L of relatively low-density culture was processed, this achievement demonstrates the feasibility of automated and controlled scaled-up bacterial culture for the purposes of lysate production. Following optimisation of the CFPS reaction, the functionality of these lysates were then verified by confirming the expression of model membrane proteins VDAC and BAK in liposomes, which are used as standards by Synthelis. Subsequently, the lysates with improved CFPS qualities were then selected for production of CXCR4, SDF1-α and CD4, as described in chapter three.


## 2.1. INTRODUCTION

### 2.1.1. The need for quality CFPS lysate

At the core of CFPS is the lysate, the quality of which is critical for high yields of useful protein. The preparation of robust and high quality *E. coli* lysate begins with optimised *E. coli* cultures and lysate production techniques. The routine *E. coli* lysate preparation protocol at Synthelis is an established standard operating procedure (SOP) which ensures robust lysate output. In this chapter, lysates from different *E. coli* strains were explored for enhanced protein production, with respect to improved yields, reduced contaminants and folding correctness. Prior to implementation into routine protocols, the use of these new *E. coli* strains for the purposes of lysate production needs to tested, studied and validated in order to produce optimised standard protocols. By adapting the established lysate protocol from

Synthelis, lysates from different *E. coli* strains were produced and their CFPS capacity was compared with each other.

### 2.1.2. Improving the lysate properties by using different bacterial strains

Various bacterial strains have been used for producing CFPS lysates (Krinsky *et al.*, 2016, Kwon and Jewett, 2015, Kim *et al.*, 2006a, Kigawa *et al.*, 2004, Soranzo *et al.*, 2016, Liguori *et al.*, 2008). For instance, Rosetta™ BL21 and BL21 codon-plus strains, both over-expressing tRNAs for rare codons, have been described to improve efficiency of eukaryotic protein expression in CFPS (Kigawa *et al.*, 2004). Origami™ BL21, which contains mutations in thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) to enhance disulphide bond formation by lessening the cytoplasmic reducing microenvironment, have been carried out in CFPS to help recombinant protein expression (Michel and Wuthrich, 2012, Kang *et al.*, 2005).

Nico21 BL21 (DE3), suited to improve protein purification by reducing the level of certain contaminating proteins from co-purifying during histidine-tagged immobilised metal affinity chromatography (IMAC), has been an efficient strain for producing robust *E. coli* lysates in our laboratories. Additional *E. coli* strains, such as C43, Rosetta-Gami B and SHuffle-K12 strains were further explored here in their use as lysates for CFPS. C43 (DE3)-BL21, a survivor of toxic recombinant membrane protein experiments, has increased intracellular membranes making it favourable for membrane protein expression (Arechaga *et al.*, 2000, Miroux and Walker, 1996). SHuffle, which constitutively expresses the disulphide bond isomerase C (DsbC), can improve disulphide formation and protein folding (Lobstein *et al.*, 2012). The Rosetta-Gami™ BL21 (DE3) strain, which combines the advantages of Rosetta™ and Origami™, has associated published work demonstrating higher expression of soluble, folded protein as compared to that from the parent Rosetta and BL21 host cells (Danping Zhang, 2010). A summary of the strains used here, and their specific characteristic of concern, is shown in the table below.

| Bacterial Strain | Specific Characteristics |
|---|---|
| **Nico21** | Expression of mutant endogenous proteins which minimises binding to IMAC (Glms) and which can be removed by chitin affinity chromatography (SlyD, ArnA and Can) |
| **Rosetta** | Carries plasmid for the over-expression of tRNAs for rare codon usage in *E. coli* |
| **Rosetta- Gami** | Improvement of Rosetta strain, with in addition, expression of mutant thioredoxin reductase and glutathione reductase to improve disulphide bond formation in *E. coli* |
| **C43** | Over-production of intracellular membranes that support membrane proteins insertion |
| **SHuffle** | Constitutive expression of disulphide bond isomerase (DsbC) for improved disulphide bond formation and protein folding in *E. coli* |

***Table 2.1:*** *Summary of the bacterial strains, and their specific characteristics, used for lysate preparation.*

### 2.1.1. Understanding bacterial growth curves: critical in lysate production

In the context of CFPS used in an industrial setting, adhesion to quality control (QC) regulations is an engagement that ensures high-quality protein products, which in turn bolsters customer confidence. Realistically however, lysate preparation tends to be a labour-intensive, multi-step and complex procedure, which is liable to variations which can be difficult to control. A particularly important source of variations in lysate preparation lies in the bacterial culture. Thus, the establishment of defined growth curves for each strain would enable detection of deviations from the norm. The sigmoid-shaped growth curve, also termed as Verhulst or logistic growth curve, is the canonical profile for population growth (Verhulst, 1845), which includes bacterial cultures. After being introduced to a new environment, the small population initially increases slowly, then rapidly increases at an exponential rate and finally attains a stable population which is dependent upon the carrying capacity of the environment. When resources are depleted and when toxic by-products accumulate, death ensues and the population quickly drops. These are correspondingly identified as the lag, log or exponential, stationary and death phases (Zwietering *et al.*, 1990) and are in turn dependent upon several factors, both environmental and

intrinsic, which affect bacterial metabolism and the overall bacterial growth. In the case of lysate production, the bacterial curve serves to identify and predict the ideal time for harvesting the culture to maximise the yield of functional lysate. This point in the curve, located in the exponential phase and prior to decline in bacterial growth, contains highly active transcription-translation machinery which is recuperated in the lysate. Here, the growth curves of the novel bacterial strains cultures to be used as sources of lysate for CFPS were therefore compared. Whilst there are several papers that provide advanced mathematical descriptions and modelling equations of bacterial growth, here the goal was to assess the differences and reproducibility of these novel bacterial strains during cell culture.

### 2.1.2. High-cell density cultures (HCDC) in fermenters and scaling up lysate production

Typically, *E. coli* for lysate production for CFPS is cultured in 2YPTG (yeast extract-phosphate-tryptone-glucose) media, which sustains high-density bacterial growth and population. However, with increased industrial growth in the use of CFPS methods and an aim to exploit CFPS for protein production for structural biology where, in both cases, the absolute yield of protein is important, the lysate supply to cope with this increased demand of CFPS production will be a limiting factor. Furthermore, with increased lysate production, manually maintaining and controlling the growth of large volumes of bacteria in flasks becomes more and more difficult. Therefore, it is imperative that reproducible measures for scaling up the lysate production be established.

High-cell density cultures (HCDC) in fermenters for lysate production is a well-established procedure that allows cultures to attain very high biomass levels, with fine manual and automated control over the culture growth and conditions (Kwon and Jewett, 2015, Zawada and Swartz, 2005, Zawada *et al.*, 2011a). This is possible since limiting factors, such as nutrients, oxygen, stirring, temperature or pH changes, are monitored and they can be supplied, altered or maintained in order to sustain continued bacterial growth and avoid detrimental by-product accumulation (Zawada and Swartz, 2005). This chapter is therefore concerned with the optimisation of lysate production in preparation for large-scale protein production to feed into structural biology techniques, as well as setting the foundations towards scaling up production and automation in the context of industrial use of the Synthelis CFPS technology.

### 2.1.3. The need for deuterated proteins and lysate for SANS purposes

Simultaneously, the setting-up of HCDC in fermenters for the production of unlabelled lysate served as a feasibility experiment to provide the basis for the production of deuterated lysates, which were to be used for SANS experiments as described extensively in chapter six. Deuterated *E. coli* culture is costly (300 €/L for $D_2O$). Therefore, in order to avoid wasteful consumption of resources, confirming the feasibility of producing lysate using HCDC in fermenters with unlabelled media was a required initial step.

An introduction to the basic principle of SANS and the need for deuterated proteins is provided in the next sections.

#### *Basic principle of small-angle scattering*

Small-angle scattering (SAS) is an analytical technique that can be used to provide low-resolution structural information on the structure of proteins and protein assemblies in solution. Despite having lower resolution than typically obtained using crystallography, its application in structural biology is advantageous since it can provide insights into the structure of bio-molecules in the absence of any effects that may occur from crystal packing forces. SAS provides structural information on the shape, size and interactions of proteins and other biomolecules in solution. Moreover, being an *in situ* method of measurement, SAS can be used to undertake parametric studies to track changes in conformation and interactions of the protein of interest.

SAS is divided into two categories: small-angle neutron scattering (SANS) and small-angle X-ray scattering (SAXS). In the case of neutrons, the neutrons interact with the nuclei of the atoms in the sample, whereas in the case of X-rays, the interaction is with the electrons of the atoms. The scattering powers of atoms for neutrons and X-rays are markedly different and provide scope for powerfully complementary structural studies. The fundamentals of the two techniques are, however, essentially the same. In SAS, an incident collimated, monochromatic neutron or X-ray beam ($\mathbf{k}_i$) passes through the protein sample of interest (Figure 2.1). The interaction of the beam with the sample results in a scattered wave pattern ($\mathbf{k}_o$) that is recorded on a detector.

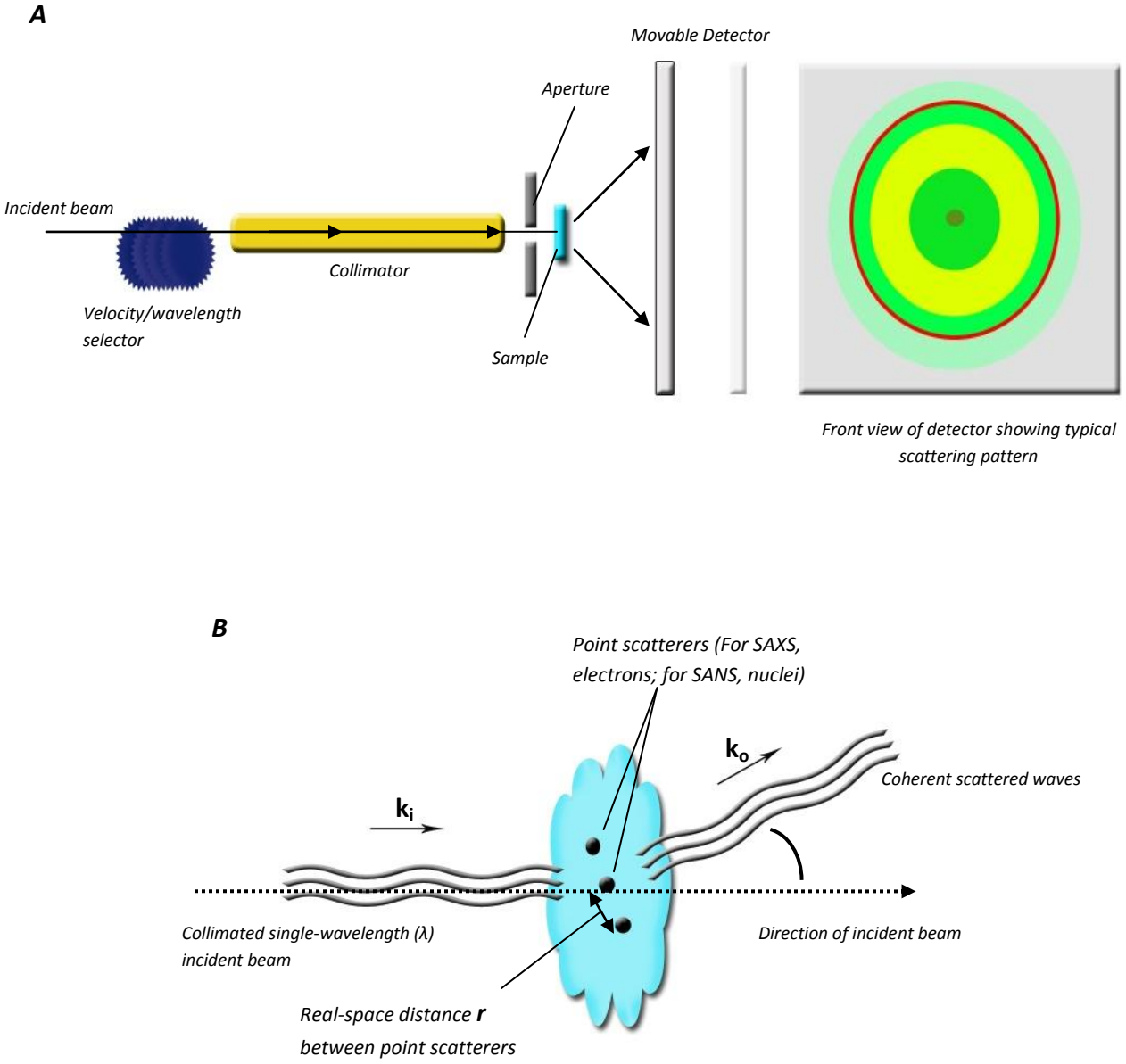



***Figure 2.1:*** *Schematic summary of a typical small-angle scattering (SAS) experiment (A) and simplified representation of the scattering event (B).*

In both neutron and X-ray scattering, the relationship between the incident and scattered wave is given by:

$$\boldsymbol{q} = \boldsymbol{k_i} - \boldsymbol{k_o} = \frac{4\pi \sin \theta}{\lambda}$$

*Equation 2.1*

Where **q** is the momentum transfer, or vector difference, between the incident wave vector, $\mathbf{k_i}$, and the scattered wave vector, $\mathbf{k_o}$. The scattered angle is given by 2θ, and the wavelength of the incident beam is λ. The equation holds true provided that the scattering is elastic: that is, there is no change in the energy between the incident and scattered wave, $|\mathbf{k_i}| = |\mathbf{k_o}| = k$. Changes between $\mathbf{k_i}$ and $\mathbf{k_o}$ are due to the direction of the wave vector. The elastic scattering arises due to the interference of scattered waves produced by point scatters (atoms of the sample) which have correlated positions in space, within the limits of coherent scattering volume. Given by **q**, the equation provides reciprocal-space coordinates, which are related to the real-space coordinates, **r** (Figure 2.2B), by Fourier transform.

However, unlike in a crystal, where protein particles are static and arranged uniformly in a lattice structure, the protein particles in SAS studies are typically in solution and are therefore randomly oriented. The scattered pattern therefore represents an average of the scattering arising from all of the possible orientations of the proteins in an isotropic solution. The resulting 2D images of the scattered pattern are integrated about the beam centre to give one-dimensional curves of scattering intensity (*I)* against *q*, where *q* is given in units of inverse Ångstroms ($Å^{-1}$) or inverse nanometre ($nm^{-1}$).

### 2.1.4. The use of SANS for structural characterisation: scattering length density, solvent contrast variation and "match-out deuteration"

SAS experiments involve the scattering of waves brought about by the sample under investigation. In the field of structural biology, the sample is typically a solution (≈300 µL) comprising of the biomolecule(s) of interest and a solvent, which are contained in an inert high-grade quartz cuvette. The scattering is influenced by inhomogeneities in the sample due to its chemical composition. In the case of SAXS, different locales of the sample having different the electron densities result in differentially

scattered X-rays, which can be distinguished from each other. In the case of SANS, the process is the same but scattering of the neutrons is brought about by the nuclei present in the sample. The neutron scattering property of the sample is termed as the scattering-length density (SLD). Thus, a sample consisting of two components having different electron densities/SLDs (for SAXS and SANS respectively) can produce scattering signals which are distinguishable from each other (Figure 2.2A). On the other hand, if the components of the sample have increasingly similar electron densities/SLDs, the signal arising from the protein sample of concern also becomes gradually indistinguishable from the background – leading to a drop in the contrast (Figure 2.2B, C). When electron density/SLD of the protein of interest is identical to that of the background, the signal between the protein and background become indistinguishable and there is no contrast difference (Figure 2.2D). Effectively, the protein component is rendered "invisible" and is termed to be "matched-out" against the background.

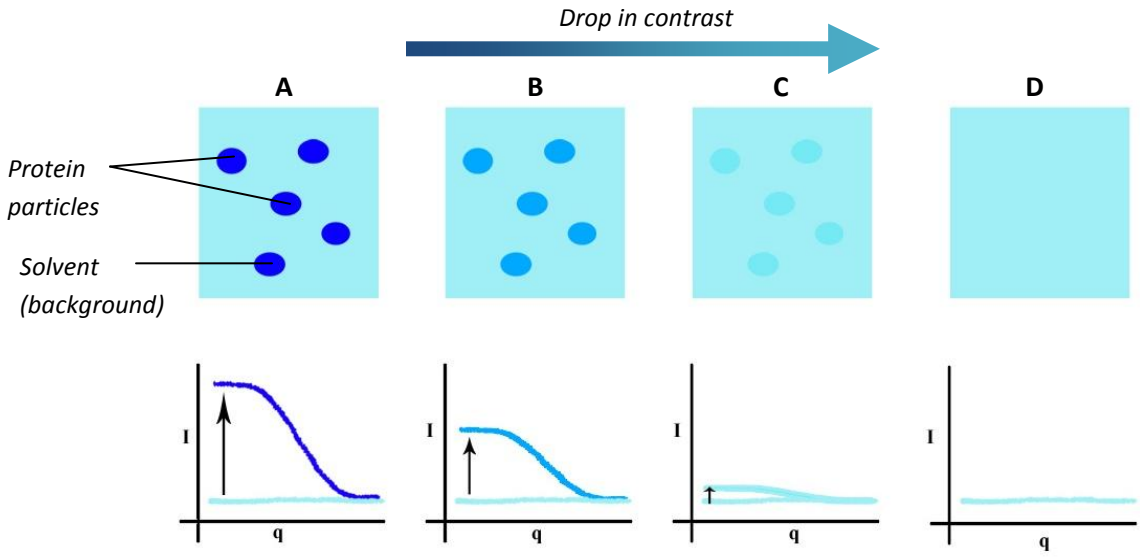



***Figure 2.2:*** *Scattering profiles showing drop in contrast and effectively resulting in the concealment of the protein of interest against the solvent background. Dark blue scatter profile = protein scatter + buffer scatter; pale blue scatter = buffer scatter only.*

Therefore, relying on the difference between the scattering profile of the protein of interest and the background, SAXS and SANS measurements reflect the contrast between the protein and the solvent in which it is immersed. In the case of SANS, $H_2O/D_2O$ solvent mixtures allow the use of a very wide range of SLDs to match that of the component required (Figure 2.3). For X-rays this can be achieved by adding components to the buffer to make it denser and therefore more electron "rich" (such as glycerol or by dissolving sugars). However, only minor variations are realistically feasible. For neutrons the options are far more powerful.

Hydrogen ($^1$H) has a neutron coherent scattering length $b_c = -3.742$ fm whilst its heavier isotope, deuterium ($^2$H or D) has a neutron coherent scattering length $b_c = 6.675$ fm, twice as large and positive. This results in very different SLDs for solvents composed of $H_2O$ ($b_c = -0.562 \times 10^{10}$ cm$^{-2}$) and $D_2O$ ($b_c = 6.404 \times 10^{10}$ cm$^{-2}$). Therefore, it is possible make solutions having specific SLDs within this range by adjusting the amounts of $H_2O$ and $D_2O$ in the buffer (Figure 2.3). When the SLD of the solution matches that of a biomolecule of interest, the scattered signals are indistinguishable from each other and the solvent and biomolecule sample mixture is said to be "matched". The bio-molecule is then referred to as being "matched-out", and is effectively rendered invisible. Specifically, at 10 %, 42 %, and 62 % $D_2O$ buffer composition, unlabelled phospholipids, proteins and DNA are respectively are matched out (Figure 2.5). Furthermore, with hydrogen atoms constituting a major portion of proteins, it is possible to synthesise deuterated recombinant proteins, which if required, can be matched out (with 85 % deuteration) in 100 % $D_2O$ buffer (Dunne *et al.*, 2017b, Haertlein *et al.*, 2016). The ability to produce deuterated proteins therefore provides a means to distinguish between different parts of a bimolecular complex. This technique is particularly useful for the study of specific proteins in complexes and for the modelling of conformational changes that arise from the interaction (Figure 2.5) (Callow *et al.*, 2007, Niemann *et al.*, 2008, Compton *et al.*, 2014, Appolaire *et al.*, 2014).

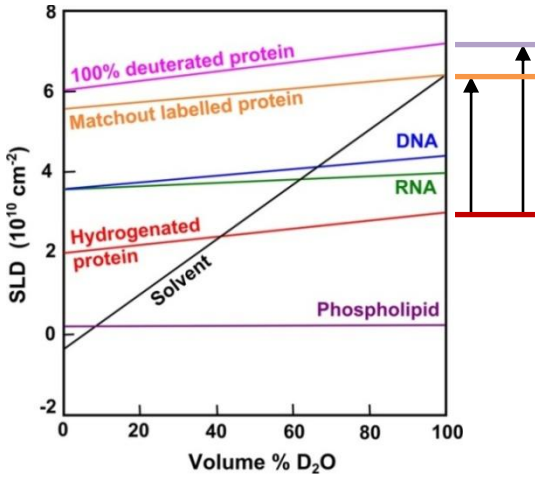



***Figure 2.3:*** *The scattering length densities (SLDs) of different biomolecules as a function of the volume percentage of $D_2O$ (A) (Dunne et al., 2017b). The black line shows the variation in SLD of the solvent as a function of $D_2O$ content. The match-point of each bio-molecule corresponds to the intersection of the solvent SLD with that for each bio-molecule. Arrows illustrate the SLD difference between 100 % and hydrogenated protein, and 85 % deuterated protein and hydrogenated protein in 100 % $D_2O$.*

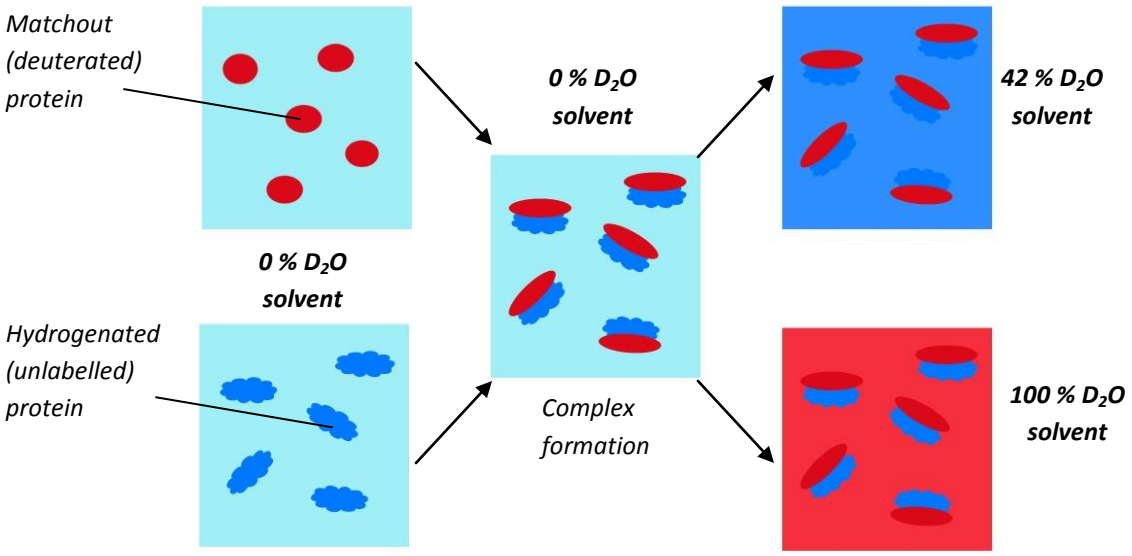



***Figure 2.4:*** *Deuterated and unlabelled proteins forming complexes can be individually matched out by altering $D_2O$ content of the solvent. This allows the observation of specific protein components in a complex, and permits detection of conformational changes induced as a result of the interaction with its binding partner.*

As well as exploring the techniques for large-scale and controlled cell culture, the use of fermenters allowed the establishment of cell culture conditions for the high-density bacterial growth in 85 % and 100 % deuterated medium for the production of respectively perdeuterated cell-free lysate (PL) and matchout cell-free lysate (LM). With proteins being the major constituent of the lysate, the LM corresponds to "matchout" conditions whereby in 100 % $D_2O$ the lysate should theoretically be rendered "invisible", leaving only the hydrogenated protein of interest "visible" (Figure 2.4). LP, which is fully deuterated CFPS lysate, serves as a reference to compare the adequate matching out conditions of the matchout lysate. These lysates were to be used in the feasibility study whereby model hydrogenated proteins, as components of the unfractionated lysates, were analysed by SANS. Described extensively in chapter six, the novel analytical technique, exploiting the power of contrast variation and matchout, therefore serves as the first step towards the ability to structural characterise un-purified CFPS proteins still in the presence of its reaction mix.

### 2.1.5. Interpretation and information obtained from SANS intensity plots

The scattering function results from the sum contributions from all of the scattering particles present in the sample. This therefore includes contributions from the proteins of interest, and of the buffer in which they are dissolved. If other components are present in the sample, these too will contribute to the scattering profile. It is necessary to subtract the scattering due to the background arising from the buffer from the total scattering profile obtained. The background subtracted scattering profile, ΔI vs q, can then be interpreted to provide structural data on the protein. This plot is the product of the form factor, P(q), and structure factor, S(q), which are described in the next sections.

#### *Form factor, P(q)*

The buffer subtracted scattering profile of a protein, ΔI vs q, is in fact the product of the form factor, P (q), and the structure factor S(q). P(q) is due to the overall scattering from **one protein particle**, which is in turn due to the summed elastic coherent from all nuclei (SANS) or electrons (SAXS) present in the protein particle. The square of the sum of all of these wave amplitudes equals P(q). P(q) therefore provides the essential information about the shape of the protein particle. The ideal conditions to

measure accurate P(q) require that the protein particles are homogenous - identical in shape and size, and that they are far away from each other (to avoid contribution of S(q), described below). P(q) derived from ideal elementary particles shapes are depicted below (Figure 2.5). However, proteins are not ideal, but are varied in shape and constantly undergoing conformational changes. SAS provides information as the average of all of the particles in a sample. The experimentally determined curves can be compared with calculated scatter functions to gain insight into the shape of the proteins under study.

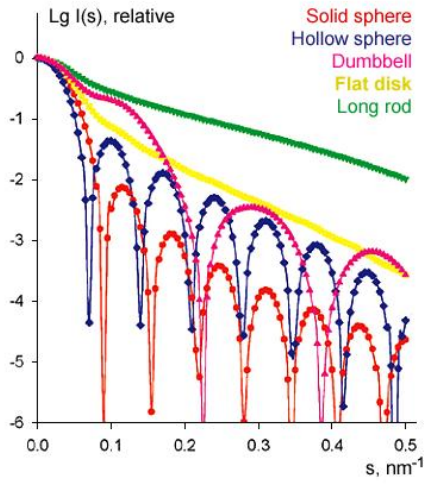



***Figure 2.5:*** *Scatter plot of simple, ideal elementary shape particles, showing the P(q) function (Svergun and Koch, 2003).*

## *Structure factor, S(q)*

The square sum of wave amplitudes arising from the interaction **between protein particles** equals S(q). Unlike in a crystal where proteins are packed closely in a lattice, protein particles in the typical SAS experiment are usually studied in relatively dilute solutions. Thus, for dilute and monodisperse solutions, the protein particles are far from each other, and S(q) is usually negligible. Throughout the scattering profile, S(q) is ideally equal to 1, such that the product of P(q) and S(q) gives only P(q). However, the effect of S(q) becomes visible at low q values. Therefore, S(q) can serve as a measure of the interactions, attractive or repulsive, occurring between protein particles. An increase in the slope of the scattering

curve of at low q (equating to interactions across short distances), for instance, indicates attractive interaction, whereas a decrease implies repulsive interaction. Repulsive interactions can be expected as the concentration of the sample is increased. Aggregation of protein particles is also marked by an increase in the slope at low q.

### *Interpretation of data in the low q range*

As mentioned above, the low q region can provide information about the S(q) of the sample and the presence of aggregation. This region corresponds to the Guinier region, for which the Guinier approximation, which enables determination of *Rg* for globular particles, holds true (Guinier and Fourmet, 1955). The *Rg* of a protein is defined as the root-mean-square distance of the atoms that constitute the protein from their common centre of gravity. The *Rg* value is therefore a measure of the overall spread of the protein atoms, and thus its compactness and indicates folding state (Lobanov *et al.*, 2008). The *Rg* value is a fundamental piece of information obtained from small-angle scattering of macromolecules in solution (Putnam, 2016).

The Guinier approximation is given by the equation:

$$I(q) \cong I_o e^{(-qRg)^2/3}$$

***Equation 2.2***

Where $I(q)$ is the scattering intensity, $I_o$ is the forward scattering intensity, $q$ is the scattering vector from equation 5.1.

Rearrangement of the equation gives:

$$ln\ [I(q)] = ln\ [I_o]\ -\ (\frac{Rg^2}{3})\ q^2$$

***Equation 2.3***

Therefore, from the plot of ln [$I(q)$] against $q^2$, as $q^2$ tends to 0, the "reciprocal space" $Rg$ value is provided. However, the condition holds true only for data in the low q region, where $qR_g < 1.0 - 1.3$.

$I_o$ value is also an important parameter since it can provide an estimation of the molecular mass (MW) of the protein. The equation for MW (g/mol) is:

$$MW = \frac{I_o N_A}{C(\Delta\rho v)^2}$$

***Equation 2.4***

Where $N_A$ is Avogadro's number ($6.02 \times 10^{23}$ /mol), C is the concentration of the protein in g/cm$^3$, $\Delta\rho$ is the contrast in $10^{10}$ cm$^{-2}$ and $v$ is the average partial specific volume of a protein, which is equal to 0.74 cm$^3$/g. Prior to interpretation however, the $I_o$ value needs to be corrected for background scattering arising from the sample cuvette, the transmission of the sample and normalised to the concentration of the protein.

### *Interpretation at intermediate q range*

The pair-distance distribution function, P(r), is a real-space representation of the shape and size of the protein particle and is interpreted from mostly the intermediate q range. The P(r) function, calculated by an indirect Fourier transform, is given by:

$$I(q) = 4\pi \int_0^{Dmax} P(r) \frac{\sin(qr)}{qr} dr$$

*Equation 2.5*

Where *r* is the distance between pairs of scattering components of one protein particle, and $D_{max}$ is the maximum inter-particle distance in the protein particle. Below, a typical depiction of P(r) plot against r for ideal elementary particles in shown (Figure 2.6).

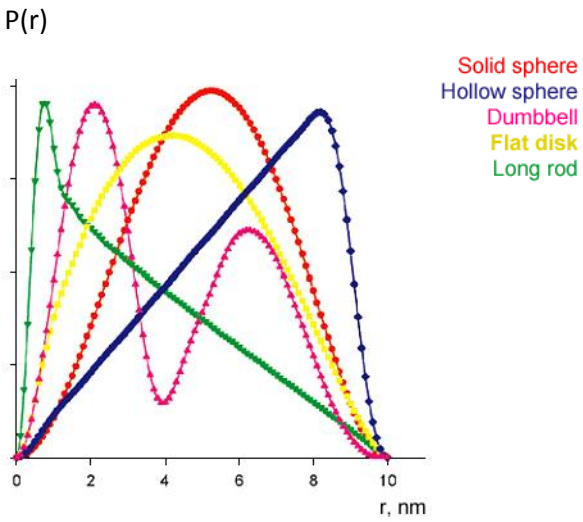



***Figure 2.6:*** *P(r) against r plot of simple, ideal elementary shape particles, showing the P(r) function (Svergun and Koch, 2003).*

From P(r), it is also possible to obtain "real-space" *Rg* values as given by:

$$Rg^2 = \frac{\int_0^{Dmax} r^2 P(r)\mathrm{d}r}{2\int_0^{Dmax} P(r)\,\mathrm{d}r}$$

***Equation 2.6***

The advantage of *P*(*r*) for determination of the *Rg* value is that it considers a much larger section of the scattering curve, and not just the lowest resolution data. Nevertheless, values of *Rg* calculated from the P(r) function and Guinier approximation should be close to each other for good quality and well-behaving samples.

### *Interpretation at high q range*

The high q range provides information about the surface of the protein particle at the protein-solvent interface. Porod's law states that at high q, the scattered intensity plot of a spherical globular protein particle in solution displays a decaying $q^{-4}$ relationship (Kratky and Porod, 1949) as defined by the equation:

$$I(q) \approx \frac{K}{q^4}$$

***Equation 2.7***

Where *K* is a constant dependent upon the surface of the protein particle. To ease visualisation, so-called Kratky plots, given by $I(q) \times q^2$ vs q, are used (Figure 2.7A). A globular protein particle will produce a bell-shape curve which tends to zero by with increasing q. Kratky plots of unfolded proteins deviate from the typical bell-shape profile by producing plateaus as q increases. However, as shown in the figure 2.7A, the Kratky plot can also serve to indicate multi-domain and highly flexible proteins.

The normalised (dimensionless) Kratky plot, which is plot of $I(q)/I_o$ x $(q \cdot Rg)^2$ vs q·Rg (Durand *et al.*, 2010), relates the Kratky plot to the Guinier approximation (Equation 2.2), and provides semi-quantitative method for assessing the state of the biomolecule in solution. As such, by mathematical derivation, it has been shown that for monodisperse and globular particles obeying Guinier approximation, a peak at $q^*Rg = \sqrt{3}$ (x-axis) and $I(q)/I_o$ x $(q \cdot Rg)^2 = 3.e^{-1}$ (y-axis) should be present (Receveur-Brechot and Durand, 2012, Durand, 2010), which can be represented on the normalised Kratky plot by a cross-hair (Figure 2.7B). Non-globular, disordered or unfolded proteins which deviate from the Guinier approximation have peaks which deviate from the cross-hair. Furthermore, I(q) being divided by the experiment's $I_o$, the particle is therefore normalised with respect to mass and concentration.

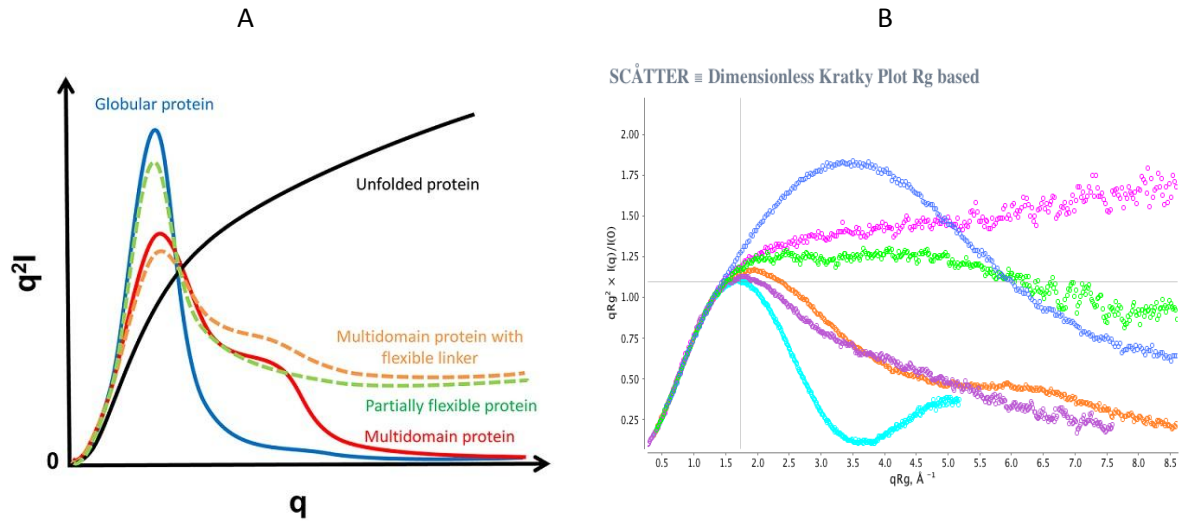



***Figure 2.7:*** *Examples of Kratky plot and normalised Kratky plot. (A) Kratky plot exemplifying plots of various types of proteins (from: www-ssrl.slac.stanford.edu/~saxs/analysis/assessment.html.). (B) Example of normalised Kratky plot (taken from: http://www.bioisis.net/tutorial/21) of a 21 kDa globular protein, xylanase (cyan), shows a peak maxima at √3. For asymmetric particles, P4-P6 RNA domain (orange) and SAM-1 riboswitch (purple), the peak shifts right with a maxima > 1.104. For an intrinsically disordered protein (RAD51-AP1), we get the classic hyperbolic plateau (magenta). Attaching RAD51-AP1 to the well-folded maltose binding protein (blue), the peak shifts right ($q \cdot R_g = 3.5$) and reaches a new maxima.*

### 2.1.6. BAK and VDAC: established model proteins to be used for confirming the activity of lysates

Human BAK, located in the outer mitochondrial membrane, is a protein that plays a critical role in the modulation of cellular survival and death. Also known as Bcl-2 homologous antagonist/killer, BAK upon being activated by pro-apoptotic signals, oligomerises to form pores which permit the release of further cascade-inducing pro-apoptotic proteins from the mitochondria into the cytosol (Reed, 2006). Voltage-dependent anion channels (VDAC), the most abundant protein in the outer mitochondrial membrane, are porin ion channels that are canonically known to regulate the transport of calcium, potassium and sodium ions, and energising molecules such as ATP into the mitochondria (Blachly-Dyson and Forte, 2001, Lemasters and Holmuhamedov, 2006). VDAC is however implicated in apoptosis, where it not only allows the leakage of pro-apoptotic proteins, but also can interact directly with BAK (Shore, 2009, McCommis and Baines, 2012, Galluzzi and Kroemer, 2007, Cheng *et al.*, 2003). By adding liposomes to the cell-free reaction, functional, histidine-tagged variants of these membrane proteins inserted into proteoliposomes, have been successfully produced by cell-free expression system (Liguori and Lenormand, 2009, Liguori *et al.*, 2008, Liguori *et al.*, 2015, Deniaud *et al.*, 2010). Indeed, the use of CFPS is well-adapted for the expression of membrane proteins as, it firstly provides a very easy method of producing high amounts of normally difficult and laborious to obtain proteins. Due to their hydrophobic and toxic nature, unless mutated, these proteins express poorly and typically form insoluble, inclusion bodies *in vivo* (Moldoveanu *et al.*, 2006, Goping *et al.*, 1998, Koppel *et al.*, 1998). In the work of this chapter, these well-known and highly-expressing proteins, of which CFPS production was optimised and standardised by Synthelis in the reference lysate Nico21, were used as models to confirm the functionality and efficiency of the lysates.

## 2.2. MATERIALS AND METHODS

### 2.2.1. Bacterial strains used for lysate preparation

The following chemically competent *E. coli* cells were used: Nico21 (DE3) (New England BioLabs, C2529H), OverExpress™ C43 (DE3) (Sigma-Aldrich, CMC0019), SHuffle® RNA T7 (New England BioLabs, C3026J), Rosetta™ (DE3) (Novagen, 70954) and Rosetta-Gami™ B (Novagen, 71135).

### 2.2.2. Media composition and preparation

The composition and preparation of media which were used in lysate preparation and which are recurrently mentioned in the materials and methods are described below:

- Diethyl polycarbonate (DEPC) – treated water (autoclaved)
- 2YPTG (autoclaved, sterile filtered 0.22µm): Sterile 2YT media (31.6 g/L), 2 % glucose, phosphate buffer
- S30 A/B (50X stock, sterile filtered 0.22µm): 500 mM Tris-acetate, 700 mM $Mg(OAc)_2$, 30 mM KCl, pH 8.2 with acetic acid
- S30 C (50X stock, sterile filtered 0.22µm): 500 mM Tris-acetate, 700 mM $Mg(OAc)_2$, 30 mM KOAc, pH 8.2 acetic acid
- Trace metal stock solution (sterile filtered 0.22µm): 0.5 g/L $CaCl_2.2H_2O$, 16.7 g/L $FeCl_3.6H_2O$, 0.18 g/L $ZnSO_4.7H_2O$, 0.16 g/L $CuSO_4.5H_2O$, 0.15 g/L $MnSO_4.4H_2O$, 0.18 g/L $CoCl_2.6H_2O$, 20.1 g/L EDTA, 5 g/L glycerol, made up with distilled water. For preparation of deuterated stocks, mineral salts were dried and labile hydrogren atoms exchanged for deuterium by dissolving in a minimal volume of $D_2O$ and then freeze drying. Mineral salts were then dissolved in $D_2O$. $D_8$ – glycerol was used instead.
- H-Enfors medium (sterile filtered 0.22µm): 6.86 g/L $(NH_4)2SO_4$, 1.56 g/L $KH_2PO_4$, 6.48 g/L $Na_2HPO_4.2H_2O$, 0.49 g/L $(NH_4)_2HC_6H_5O_7$ (diammonium hydrogen citrate), 0.25 g/L $MgSO_4.7H_2O$, 5 g/L glycerol, 1.0 mL of trace metal stock solution, made up with distilled water. For preparation of deuterated stocks, 100 % $D_2O$ (100 % D-Enfors) or 85 % $D_2O$/15 % $H_2O$ (85 % D-Enfors) was used to dissolve salts, deuterated mineral stock solution used and $D_8$ – glycerol was used instead.

### 2.2.3. *E. coli* culture in flasks

The protocol that was developed by Synthelis was used and adapted for the culture of the various bacterial strains. Due to confidentiality, a brief overview of the protocol is described here. Bacterial cultures, of the various strains mentioned in section 2.2.1, were inoculated into 1500 mL of 2YTPG media baffled 3 L flasks at an initial $OD_{600}$ of 0.2. Four flasks of cultures for each strain were used for the generation of a total of 6 L of culture. Cultures were incubated at 37 $^{o}$C at 170 rpm. Regular $OD_{600}$ readings were taken and growth curves plotted. The culture was stopped in the exponential phase. The cells were harvested by centrifugation and the cell paste was then collected.

### 2.2.4. *E. coli* culture lysis and lysate processing

The cell paste in ice-cold S30 A buffer (1 X S30 A/B buffer, 6 mM β-mercaptoethanol, distilled water) (Figure 2.8, III) using a cell homogeniser (Ultra Turrax T25, S25N-18G) (Figure 2.8, IV). The cell paste was washed and recuperated by centrifugation (Figure 2.9, V). The supernatant was discarded and cell paste was resuspended in fresh ice-cold S30 A buffer (Figure 2.9, VI). The resuspension was centrifugated and the collected cell paste was resuspended again into fresh ice-cold S30 A buffer. The resuspended cell paste was washed a final time and collected by centrifugation. The supernatant was discarded and the combined wet cell mass derived from 6 L of culture was weighed. The cell pastes were then stored at -80 $^{o}$C until further processing, or used immediately.

The cell pastes were resuspended and homogenised in ice-cold S30 B buffer (1 X S30 A/B, 1 mM DTT, 1 mM PMSF, distilled water) (Figure 2.8, VII, VIII). The cell paste was lysed at 1500 PSI using a high pressure homogeniser (AVESTIN, Emulsiflex-C3) (Figure 2.8, IX). The lysed cells were then centrifuged. The supernatant was then collected into fresh tubes (Figure 2.8, X). Stock solution of 4 M NaCl in DEPC-water was added to the supernatant up to a final concentration of 400 mM. The tubes were then incubated in the dark at 42$^{o}$C for 45 minutes under very light agitation (50 rpm) in order to produce the run-off (Figure 2.8, XI). The run-off was then transferred into dialysis tubings (12–14 kDa molecular weight cut-off) and incubated in 100 X volume of cold S30 C buffer (1 X S30 C, 0.5 mM DTT, distilled water). Dialysis was carried out extensively with several changes into fresh S30 C buffer dialysis (Figure 2.8, XII). The dialysed lysate was then recuperated into fresh tubes and centrifuged (Figure 2.8, XIV).

Aliquots of the supernatant, representing the final lysate produce, were collected and stored at -80 $^{o}$C until use for CFPS. The frozen lysate can be stored effectively for extended periods of time (Synthelis unpublished data).

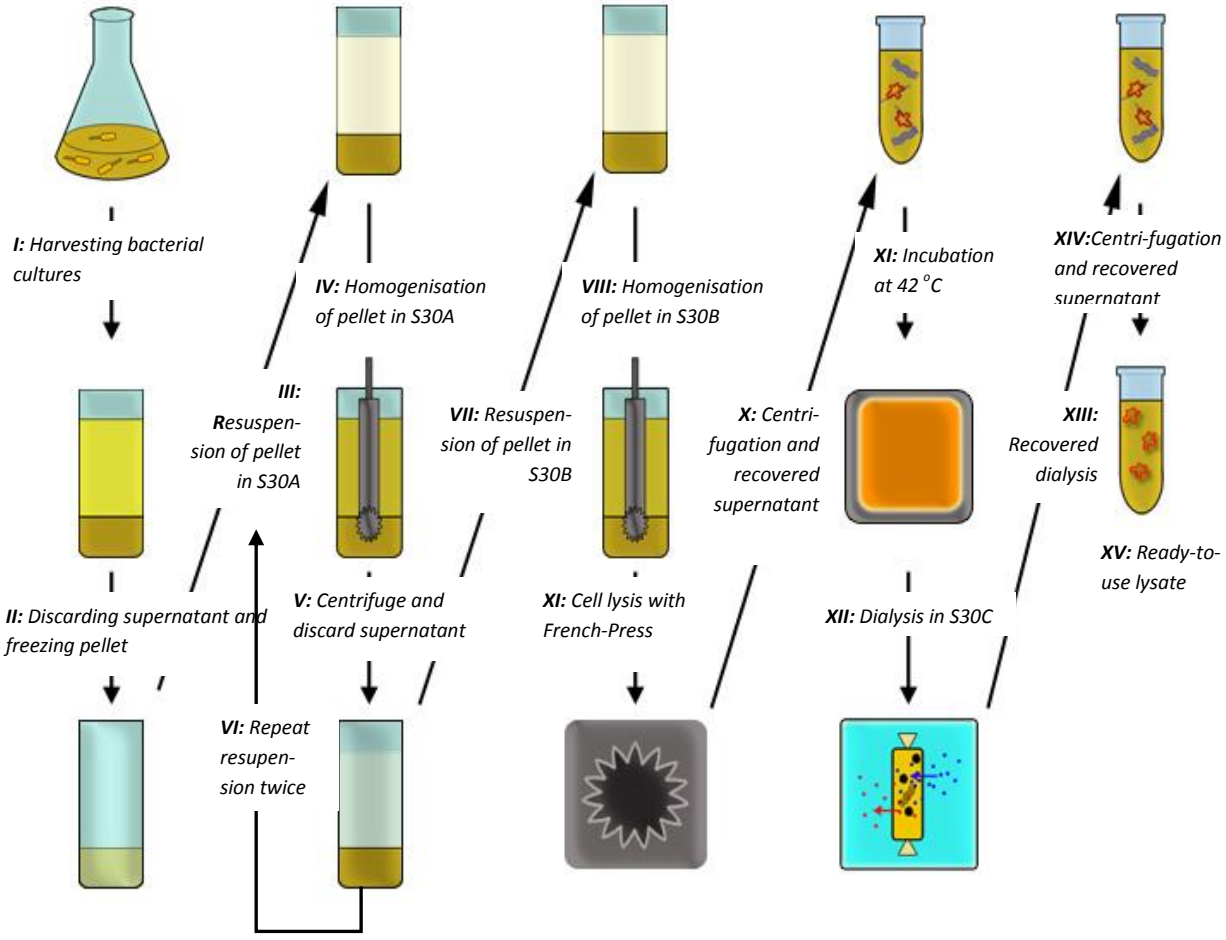



***Figure 2.8:*** *Diagram depicting the steps involved in lysate preparation.*

### 2.2.5. Lysate validation using reference proteins

To validate the functionality of the lysate, batch CFPS reactions using either highly-expressing reference histidine-tagged Bcl-2 homologous antagonist/killer (BAK) or histidine-tagged voltage-dependent anion channel (VDAC) was setup. Magnesium ion, potassium glutamate and lysate concentrations are critical for successful CFPS. Reactions in a 96-well plate were therefore setup to screen for the optimum concentrations of these components.

#### *Plasmid preparation for use in CFPS*

Polyhistidine-tagged BAK and polyhistidine-tagged VDAC into specific expression vectors were provided by Synthelis (Liguori *et al.*, 2015, Deniaud *et al.*, 2010).

The plasmid was transformed into chemically competent DH5α *E. coli* cells (New England BioLabs, C2987H) using the heat shock method. Stocks of 25 µL of DH5α *E. coli* cells were thawed on ice. Plasmid DNA was added to the cells up to a final concentration of approximately 2 ng/µL. The cells were incubated on ice for 20 minutes and incubated for 60 seconds at 42$^{o}$C on a heating block. The cells were incubated on ice for 2 minutes and 180 µl LB or SOC media (without antibiotic) were added to the bacteria then the culture was grown in a shaking incubator for 45 minutes at 37°C. After growing in SOC medium, the cells were plated on LB agar with ampicillin at 100 µg/ml. The plates were incubated overnight at 37$^{o}$C and the plates were then stored at 4$^{o}$C for a maximum of two weeks. Single transformed colonies present on the plates were picked and inoculated into 3 mL of starter ampicillin (100 µg/ml) LB culture in sterile capped 5 mL polypropene tubes. The starter culture was incubated at 37$^{o}$C at 200 rpm agitation for at least 6 hours. The culture was then tipped into 300 mL of ampicillin (100 µg/ml) containing LB in a sterile 500 mL baffled flask. The culture was incubated overnight at 37 $^{o}$C at 200 rpm. The culture was harvested by centrifugation at 7000 g at 4$^{o}$C for 15 minutes and the cell paste was collected. A plasmid maxi-prep kit (Qiagen) was used to extract and purify the plasmid from the cell paste. The purified DNA was quantified using a NanoDrop ND-1000 spectrophotometer. Purified plasmids were stored at -20$^{o}$C.

### *Reaction setup and screen in 96-well plate*

A reaction mix for batch CFPS consisting of invariable reagents was made. The invariable reagents consisted of 55 mM HEPES-KOH buffer (pH 8.1) containing 2 µM folinic acid, 330 µM nicotinamide adenine dinucleotide (NAD), 2.4 mM AMP, 1.7 mM GMP, 1.7 mM UMP, 1.7 mM CMP, 1.32 mM spermidine, 30 µM 3-phosphoglyceric acid (3PG), 0.5 mg/ml tRNA, 1 x cOmplete® EDTA-free protease inhibitor cocktail, 2.1 mM sodium oxalate and amino acid mix at 2.0 mM per amino acid. All solutions were made using DEPC-treated water. Plasmid DNA was used at a final concentration of 15 ng/µL of reaction. Table 3 provides a description of the reagents used for setting up the CFPS reactions. Liposomes, artificial phospholipid vesicles, are critical components in the CFPS reaction that provide the hydrophobic matrix for the insertion of the *de novo* expressed membrane protein. An extensive description on the preparation of liposomes can be found in the following chapter. For a final volume of 50 µL per well, 26.2 µL of the mix was then dispensed into a 96-well flat-bottom plate.

| Reagents | Volume per 50 µL reaction (µL) |
|---|---|
| **Hepes 2 M KOH 1.2 M pH 8.1** | 1.4 |
| **Folinic acid 2 µM** | 0.5 |
| **NAD 33 mM** | 0.5 |
| **AMP 240 mM** | 0.5 |
| **GMP 170 mM** | 0.5 |
| **UMP 170 mM** | 0.5 |
| **CMP 170 mM** | 0.5 |
| **Spermidine 300 mM** | 0.2 |
| **3PG 1.5 M** | 1.0 |
| **tRNA 25 mg/mL** | 1.0 |
| **CIP 50X** | 1.0 |
| **Amino acid mix** | 8.3 |
| **Sodium oxalate 210 mM** | 0.5 |
| **Liposome 30 mg/mL** | 8.3 |
| **DNA 500 µg/mL** | 1.5 |
| **TOTAL** | 26.2 |

***Table 2.2**: List of invariable reagents mix and volumes required for 50 µL cell-free reaction.*

The estimates for the optimum amounts of magnesium and potassium glutamate were determined using a preliminary dot blot screen. As shown in Figure 2.9, a range of 1.6 – 9.6 mM magnesium glutamate (L-glutamic acid hemimagnesium salt tetrahydrate, Mg/Glu) and 200 – 300 mM potassium glutamate (K/Glu) were screened the in the 96-well flat-bottom plate. The amount of lysate was varied between 20 – 35 % of the total final volume in each quadrant of the plate. The final volume of 50 µL was made up with DEPC-treated water. The plate was sealed with a plate cover and incubated overnight (16 hours) at 30 °C at 150 rpm.

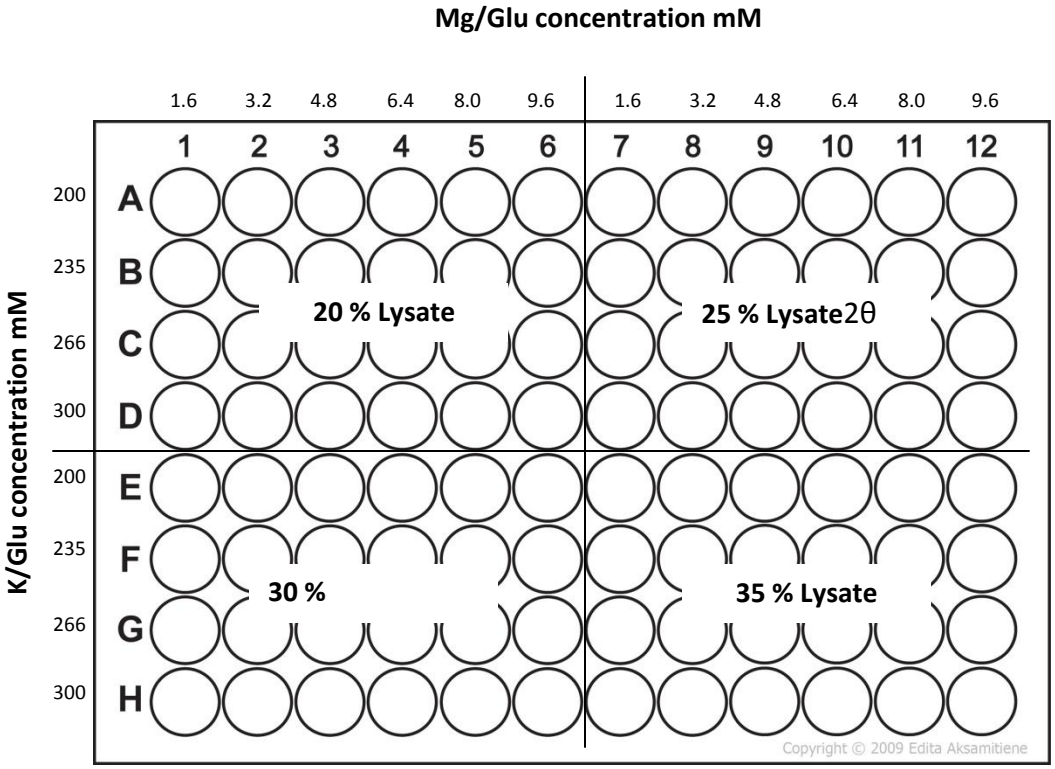



***Figure 2.9:*** *Setup for magnesium glutamate, potassium glutamate and lysate screen. The screen was carried out in a 96-well plate for lysate validation and determination of optimum conditions for CFPS reactions.*

### *Dot blot of 96-well plate CFPS reaction*

The reaction plate was removed from the incubator, the seal removed and the CPFE reaction of each well mixed thoroughly. Dot blot reactions were then carried out. In a new 96-well plate, 200 µL of dot blot loading buffer (99:1 distilled water to gel loading buffer) was dispensed in each well. To each well, 1.5 µL of each CFPS reaction from the reaction plate was transferred to the corresponding well onto the new plate. The contents of the wells were mixed adequately. Into an assembled dot blot cassette, 100 µL of each well was dispensed into corresponding wells onto a pre-soaked nitrocellulose membrane (Hybond C, Amersham Bioscience, UK). The pump was switched on and mix from each well was transferred onto the membrane. The dot blot cassette was disassembled and the membrane was incubated into blocking buffer (5 % non-fat milk powder in 1 X TBS-T) for 1 hour with light agitation. The membrane was then incubated with monoclonal anti-polyhistidine peroxidase conjugate antibody (1 µg/mL antibody in blocking buffer) for 1 hour with light agitation. The membrane as then rinsed for 10 minutes three times with 1 X TBS-T buffer. The membrane was incubated with detection solution (ECL™ Prime western Blot, GE HealthCare) and polyhistidine chemiluminescence detected on Chemi-Doc imaging instrument (Bio-Rad, CA, U.S.A).

### *SDS-PAGE, Coomassie blue gel staining and western blot analysis*

From selected wells, 10 µl of sample was added to 2.5 µL of 5 X gel loading buffer (250 mM Tris. HCl, pH 6.8, 10 % SDS, 30 % (v/v) glycerol, 10 mM DTT, 0.05 % (w/v) bromophenol blue) in 1.5 mL Eppendorf tubes. Without boiling, 5 µL samples were loaded onto a 12 % SDS-PAGE gel and gel electrophoresis was carried out at 150 V until the dye-front exited the gel.

For Coomassie blue gel staining, the resolved gel was incubated in Coomassie blue stain and subsequently destained by incubation in destaining buffer.

For western blot analysis, the resolved gel was transferred onto a transfer buffer soaked nitrocellulose membrane (Hybond C, Amersham Bioscience, UK) and sandwiched between six layers of transfer buffer soaked filter paper. Western blotting was carried out using the recommended settings from the Trans-Blot Turbo semi-dry blotting transfer system (BioRad). The system was then disassembled and the

membrane was incubated into blocking buffer (5 % w/v non-fat milk powder in 1 X TBS-T) for 1 hour with light agitation. The membrane was then incubated with monoclonal anti-polyhistidine peroxidase conjugate antibody (1 µg/mL antibody in blocking buffer) (Sigma-Aldrich) for 1 hour with light agitation. The membrane as then rinsed for 10 minutes three times with 1 X TBS-T buffer. The membrane was incubated with detection solution (ECL™ Prime Western Blot, GE HealthCare) and polyhistidine chemiluminescence detected on Chemi-Doc imaging instrument (Bio-Rad, CA, U.S.A).

### 2.2.6. Fermenter culture of bacteria in hydrogenated and deuterated Enfors medium

#### *Cell adaptation and growth in hydrogenated Enfors medium*

For high-density fermenter cultures (HDFC) the Rosetta strain of *E. coli* was used. A single colony of Rosetta *E. coli* was picked from the agar plate and inoculated into 10 mL of chloramphenicol (33 µg/ml) LB in a sterile 25 mL flask. The starter culture was incubated overnight at 37 °C at 200 rpm. The culture was then stored at 4 °C until further use.

For the production of deuterated recombinant proteins, which were used subsequently in SANS structural studies (see chapter six), *E. coli* was grown in deuterated Enfors medium (D-Enfors). The Rosetta cells therefore first had to be adapted for growth in non-deuterated Enfors medium – termed as hydrogenated Enfors medium (H-Enfors). This was achieved by a step-wise passaging process in which the starter culture grown was successively inoculated into fresh H-Enfors. 1 mL of starter culture was diluted into 9 mL of chloramphenicol-containing (33 µg/mL final concentration) H-Enfors in a sterile 125 mL a sterile, dry, capped flask. The culture was incubated for 24 hours at 37°C at 200 rpm. 1 mL of the culture was then added to 9 mL of fresh chloramphenicol-containing (33 µg/mL) H-Enfors in a sterile 25 mL flask. The culture was then incubated for 24 - 36 hours at 37 °C at 130 rpm in a 25 mL flask. Passaging was repeated at least a further six times in order to ensure that Rosetta cells had been successfully adapted to H-Enfors growth. The final culture of cells in H-Enfors could then be stored at 4°C. 2 mL of the final culture was used to inoculate 200 mL of fresh chloramphenicol-containing (33 µg/mL) H-Enfors in a 500 mL flask. The culture was incubated for 24 hours at 37°C at 130 rpm. The culture was then stored at 4°C and used within a week for high-density fermenter culture.

### *Cell adaptation and growth in deuterated Enfors medium*

For growth in 100 % D-Enfors, passaging was repeated as described above but inoculation was instead performed in chloramphenicol-containing (33 µg/mL) 100 % D-Enfors. The H-Enfors adapted cells were used as starter culture. Growth being slower in 100 % D-Enfors requires incubation of cultures at 37 $^{o}$C at 200 rpm for at least 48 hours. Passaging was repeated once again at least six times in order to ensure that the cells had been successfully adapted to D-Enfors growth.

### *Fermenter culture of bacteria in hydrogenated Enfors medium*

In 3 L fermenters (Labfors, Infors), 1.7 L of H-Enfors was inoculated with 120 mL pre-culture of adapted cells. During the batch and fed-batch phases, the pH was adjusted to 6.9 by addition of 10 % ammonium hydroxide. The temperature of the culture was maintained to 30 °C. The gas-flow rate of sterile filtered air was 0.5 L/min. Stirring was adjusted to ensure a dissolved oxygen tension of 30 %. The fed-batch phase with 12 % (w/v) H-glycerol enriched D-Enfors, was initiated when the oxygen consumption spike was detected which reflects depletion of nutritive carbon source (glycerol) in the culture. Cultures were harvested at $OD_{600}$ = 19.0.

### *Fermenter culture of bacteria in deuterated Enfors medium*

For fermenter growth in 100 % D-Enfors, 1.8 L of 100 % D-Enfors (made using 100 % $D_2O$) was inoculated with 120 mL respective pre-culture of adapted cells. Growth conditions identical to those described above were used. The pD of 6.4, equivalent to a pH of 6.9, was however maintained using 10 % NaOD that was made with 100 % $D_2O$. The feed solution used consisted of 12 % $D_8$-glycerol in 100 % D-Enfors. Culture was harvested at $OD_{600}$ = 15.8.

For fermenter growth in 85 % D-Enfors, 1.2 L of 85 % D-Enfors (made with 85 % $D_2O$: 25 % $H_2O$) was inoculated with 150 mL respective pre-culture of adapted cells. Identical growth conditions as described above. The pD of 6.4, equivalent to pH of 6.9, was however maintained using 10 % NaOH that was made with 85 % $D_2O$. The feed solution used consisted of 12 % (w/v) H-glycerol in 85 % D-Enfors. Cultures were harvested at $OD_{600}$ = 18.3.

For each fermenter run, $OD_{600}$ measurements were taken throughout. When the mid-exponential growth phase was reached, 1 M IPTG made in the corresponding culture media was added to a final concentration of 1 mM. The culture was stopped and harvested as soon as the growth plateau was reached. Harvesting was carried out by centrifuging the culture in 1 L centrifugation bottles at 7000 g for 15 minutes at 4$^{o}$C. The supernatant was discarded and the cell paste was then stored at -80$^{o}$C. Lysate processing was carried out as described in section 2.3.4. Lysates derived from H-Enfors, 100 % deuterated D-Enfors and 85 % deuterated D-Enfors growth conditions were respectively termed as hydrogenated lysate (HL), perdeuterated lysate (PL) and matchout lysate (ML). The lysates were then validated using BAK or VDAC model protein using western blot as described previously.

## 2.3. RESULTS

### 2.3.1. Bacterial culture for lysate preparation

#### *Comparing growth curves and rate of various E. coli strains*

Lysate preparation is perhaps the most important task when setting up effective cell-free expression systems. It is also the most time-consuming and labour-intensive task in the process. Once prepared however, the lysate can be conserved for extended period of times at - 80 $^{o}$C. However, the lysate and its preparation are the most likely sources of variability which can affect the quality of CFPS and resulting protein yields. Indeed, from experience, it is not uncommon that two different batches of lysate production do not have the same CFPS quality. Therefore, efforts to develop robust, effective and reliable lysates are required.

Each lysate preparation begins, in this case, with *E. coli* bacterial culture. This procedure requires accurate and standard bacterial culturing, induction and harvesting protocols that have been developed by Synthelis. In this thesis, these protocols were adapted to establish new, standard protocols for culturing different *E. coli* strains. Nico21, a derivative of the BL21 (DE3), has been a favoured bacterial strain for producing robust *E. coli* lysates at Synthelis. The optimised protocol was then transposed for other strains to validate the robustness of the process. Various other *E. coli* strains having specifically suited advantages have been developed and are currently commercially available. To investigate whether these advantages can be transcribed to CFPS with respect to increasing the quality and yields of expressed protein, several lysates derived from various *E. coli* strains were produced and tested.

The growths of these strains in batches of 1.5 L of 2YPTG media in 3 L baffled flasks at 170 rpm were compared and analysed in order to understand and develop adapted growth protocols for lysate preparation. Starting off with $OD_{600}$ 0.2, it was observed that all of the *E. coli* strains follow the well-known bacterial growth profile, comprising the initial lag, middle exponential and final plateau phases (Figure 2.10A). Since cultures were diluted during spectrophotometry readings to ensure $OD_{600}$ values (<0.5) suitable for accurate results, $OD_{600}$ values recorded are directly proportional to cell density, or population, *N* (Koch, 1968, Koch, 1970, Stevenson *et al.*, 2016). The lowest and highest $OD_{600}$ values therefore are equivalent to the initial inoculum ($N_o$) and maximum ($N_{max}$) *E. coli* cell populations.

It can therefore be observed from Figure 4A that C43, Nico21, Rosetta, SHuffle and Rosetta-Gami attain respectively $N_{max}$ of 9.15, 9.13, 9.28, 8.70 and 7.48. The recovered wet weight biomasses of these were in turn respectively 16.3, 17.7, 16.8, 13.0 and 9.7 g/L of 2YPTG culture media.

Bacterial growth is known to follow logarithmic growth, whereby there is doubling of the population at regular time intervals, which is defined as the doubling time ($t_d$). Measuring $t_d$ provides therefore indication on the rate of growth of bacterial growth during the exponential phase. As such, exponential bacterial growth follows the growth equation:

$$N = N_o e^{\mu t}$$

***Equation 2.8***

Where $N$ is the bacterial population, $N_o$ is the initial inoculum population, $\mu$ is the growth rate and t, is the time, in hours, after inoculation. $\mu$ is equal to $ln\ 2/t_d$.

This can be re-written as equation:

$$ln\ \frac{N}{N_o} = \mu t$$

***Equation 2.9***

The graph of $ln\ \frac{N}{N_o}$ against $t$ was plotted and linear portion of the curve (Figure 2.11B), which corresponds to the exponential phase, was estimated to be between one and three hours (Figure 2.11C). Gradients of the lines of best fit were calculated in order to determine $t_d$ of the corresponding *E. coli* culture (Table D). C43 was observed to have the lowest $t_d$ of 33 minutes. Nico21 and Rosetta, having almost identical growth profiles, had $t_d$ at 36 and 38 minutes respectively. SHuffle and Rosetta-Gami had the longest $t_d$ at respectively 40 and 43 minutes. This order matches the order for cellular mass production noted above.

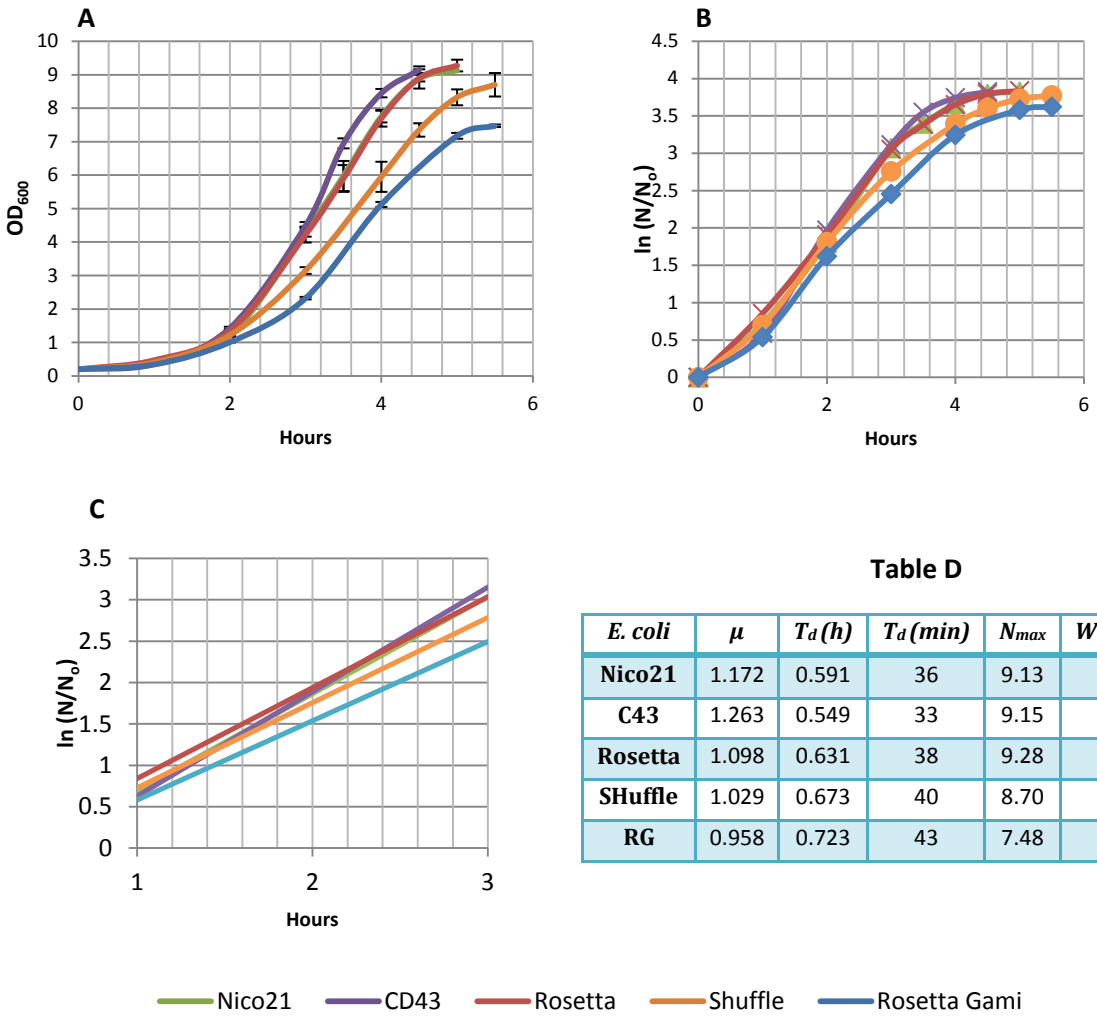



**Table D**

| *E. coli* | $\mu$ | $T_d$ (h) | $T_d$ (min) | $N_{max}$ | WW (g/L) |
|---|---|---|---|---|---|
| **Nico21** | 1.172 | 0.591 | 36 | 9.13 | 16.3 |
| **C43** | 1.263 | 0.549 | 33 | 9.15 | 17.7 |
| **Rosetta** | 1.098 | 0.631 | 38 | 9.28 | 16.8 |
| **SHuffle** | 1.029 | 0.673 | 40 | 8.70 | 13.0 |
| **RG** | 0.958 | 0.723 | 43 | 7.48 | 9.7 |

***Figure 2.10:*** *Growth curves of different bacterial strains to be used for lysate production, measured as $OD_{600}$ against time. The growth curves of the different E. coli strains as four batches each in 1.5 L of 2YPTG media in a 3 L baffled flask is displayed (A). With $N_o$ at 0.2, $N_{max}$ of C43, Nico21, Rosetta, SHuffle and RG were respectively 9.15, 9.13, 9.28, 8.70 and 7.48. Logarithmic curve fitting shows the exponential phase of bacterial growth where a linear relation is observed between 1 – 3 hours (B). Lines of best fit, following $\boldsymbol{Ln\ \frac{N}{N_o} = \mu t}$ equation, were drawn between these points (C). The gradient of the lines provides growth coefficients, μ, from which doubling times, $T_d$, of the bacteria were determined (Table D). Also shown are the end $OD_{600}$ readings and wet weights (WW) of cell pastes collected. In each case, four independent flask cultures were used (n=4).*

### *Assessment of bacterial culture in fermenter*

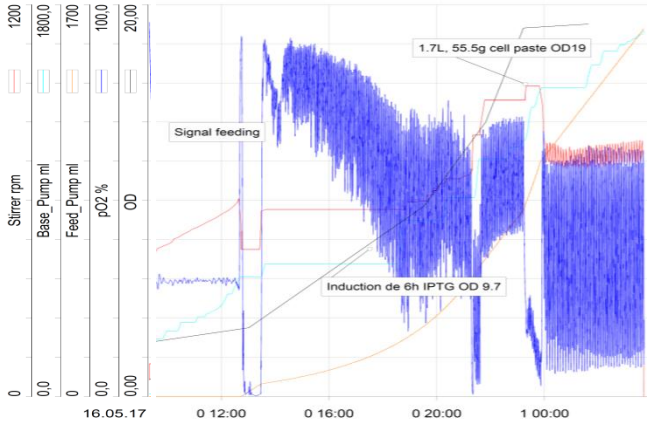
The growth of bacteria in deuterated media can be severely compromised. In order to maximise the production of deuterated lysate for the purpose of SANS, high-density fermenter culture (HDFC) was opted for instead of flasks for growing the bacteria. Furthermore, HDFC provides the basis for scaling up and automation of growth conditions of bacteria for the purposes of lysate production.

Due to the high costs of deuterated media ($D_2O$ and D-glycerol), a feasibility test prior to growth of bacteria in deuterated HDFC was carried out. This involved the HDFC growth of bacteria in unlabelled, hydrogenated media which was subsequently used to produce CFPS lysate. The feasibility of this test therefore provided the reference for setting-up and comparing against the HDFC growth in deuterated media. For HDFC, the Rosetta strain was selected. Firstly, it carries chloramphenicol-resistance. This allows the inclusion of chloramphenicol in culture which would positively select for Rosetta bacteria, whilst preventing contamination. Secondly, the Rosetta strain is an autotroph that can survive on the nutrient-limited media which is used for deuterated HDFC. In contrast, the SHuffle strain used here is an auxotroph which cannot grow without enriched media. As shown in the following section (Figure 2.15), CFPS expression using the Rosetta-derived lysate produced from flask cultures was shown to enhance expression of proteins. It was therefore expected that the high protein expressing qualities of the CFPS would be replicated as well using HCDC.

The Rosetta *E. coli* strain was adapted for growth in H-Enfors and targeting protein production for neutron-based structural studies also adapted for growth in both 100 % D-Enfors and 85 % D-Enfors, and cultured at high-density in fermenters under controlled and tracked conditions (Figure 2.11).

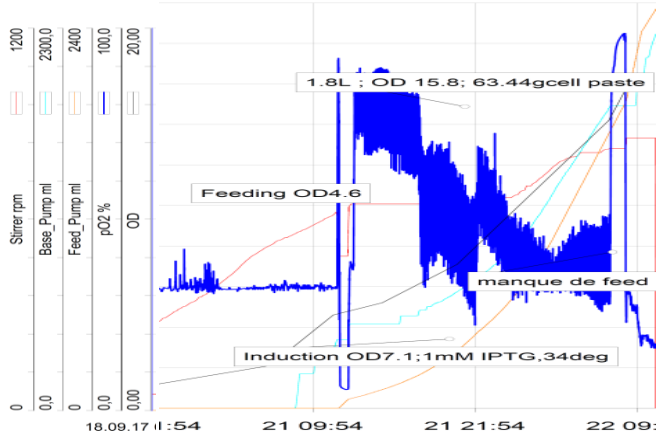
Culture of Rosetta in H-Enfors was performed first and provided the reference for subsequent culture in deuterated conditions. Rosetta culture in H-Enfors, 100 % D-Enfors and 85 % D-Enfors was induced with 1 mM IPTG respectively at $OD_{600}$ 9.7, 7.1 and 8.0, which corresponds to approximately the mid-log phase. As soon as a drop in the exponential rate of growth was detected, the culture was stopped. As such, the $OD_{600}$ in H-Enfors, 100 % D-Enfors and 85 % D-Enfors attained 19.0, 15.8 and 18.3 respectively. From the final volumes of 1.7, 1.8 and 1.2 L of H-Enfors, 100 % D-Enfors and 85 % D-Enfors respectively, 55.5, 74 and 63.4 g of wet cell paste were collected. It should be noted due to spillage; some H-Enfors was lost and therefore was not accounted for.

**Fermenter culture of in H-Enfors for producing hydrogenated lysate (HL)**



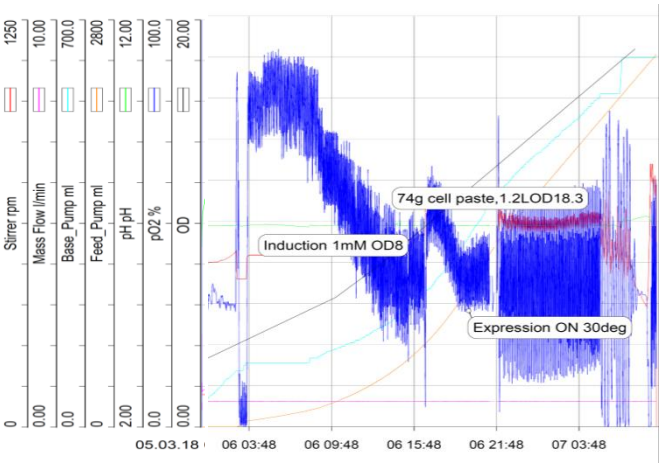

| Feature | Value |
|---|---|
| Initial $OD_{600}$ | 0.54 |
| Feed Start $OD_{600}$ | 3.8 |
| Induction $OD_{600}$ | 9.7 |
| Final $OD_{600}$ | 19.0 |
| Volume of Culture | 1.7L |
| Weight of Cell paste | 55.5g |

**Fermenter culture of in 100 % D-Enfors for producing perdeuterated lysate (PL)**




| Feature | Value |
|---|---|
| Initial $OD_{600}$ | 0.15 |
| Feed ctart $OD_{600}$ | 4.6 |
| Induction $OD_{600}$ | 7.0 |
| Final $OD_{600}$ | 15.8 |
| Volume of culture | 1.8L |
| Weight of cell paste | 63.4g |

**Fermenter culture of in 85 % D-Enfors for producing matchout lysate (ML)**




| Feature | Value |
|---|---|
| Initial $OD_{600}$ | 0.3 |
| Feed ctart $OD_{600}$ | 4.0 |
| Induction $OD_{600}$ | 8.0 |
| Final $OD_{600}$ | 18.3 |
| Volume of culture | 1.2 L |
| Weight of cell paste | 74.0 g |

***Figure 2.11:*** *Growth tracking curves of HCDC fermentation of Rosetta E. coli against time (hours) in H-Enfors, 100 % D-Enfors and 85 % D-Enfors. The parameters considered are stirrer speed, dissolved oxygen concentration, air flow volumes, feed and base volumes. Following inoculation, bacterial growth is monitored by considering oxygenation consumption ($pO_2$, in blue). Once bacterial growth has been stabilised, attaining the set-point of 30 % $pO_2$, stirring (red) is increased to increase oxygen supply and exchange in order to sustain growing cells. Sudden increase in bacterial oxygen consumption indicates the end of batch phase and depletion of resources. Corresponding to $OD_{600}$ values (black) of approximately 4 – 5, the fed-batch phase is then initiated with the supply of hydrogenated or deuterated ($D_8$) -glycerol feed (orange) to cultures. Throughout, base (light-blue) is supplied to maintain constant pH of 6.9. The culture is induced with 1 mM of IPTG during mid-log phase ($OD_{600}$ 7-9) and maintained for six hours before harvest.*

### 2.3.2. Dot blot and western blot analysis of proteins BAK and VDAC expressed in produced lysates

#### *Confirming BAK expression in Nico21 lysate by dot blot and western blot analysis*

In order to firstly validate lysate functionality and secondly, to establish working concentrations of Mg/Glu and K/Glu ions, and lysate proportion in batch CFPS, screening using dot blot analysis for these conditions was carried out for the expression of histidine-tagged BAK protein in proteoliposomes using the standard Nico21 lysate (Figure 2.12A). The expression of histidine-tagged BAK was confirmed. The expression of BAK was dependent upon Mg/Glu, K/Glu and lysate concentrations. At low concentrations of all of these components, no BAK is expressed. For 20 % and 25 % lysate, it was observed that expression of BAK was limited by low concentrations of Mg/Glu: at 1.6 mM Mg/Glu there was no protein expression. Above 30 % lysate, all of the tested Mg/Glu and K/Glu conditions permitted BAK expression. Across the conditions screened, it appears that CFPS is more sensitive to changes in Mg/Glu compared to K/Glu.

Dot blot allows rapid visualisation of expression across a wide-range of conditions. However, due to the highly sensitive detection, saturation is quickly achieved and artefacts can therefore be introduced in the interpretation of data. Furthermore, the dot blot shows the presence of all of the histidine-tagged protein, which includes full-length protein, oligomers and truncated species. Having a known molecular weight of approximately 16 kDa, the presence of the BAK was confirmed, when samples from dot blot were selected for visualisation by western blot (Figure 2.12B). Here, relative estimates on the protein yields could be determined. With approximately twice the amount of BAK present, higher amounts of Mg/Glu, K/Glu and lysate increased BAK expression. As confirmed by mass-spectrometry by Synthelis previously (unpublished data), the presence of truncated BAK species was also observed. Also present was the BAK dimer, of theoretical molecular weight 32 kDa, which corresponds to the band between 25 – 30 kDa.

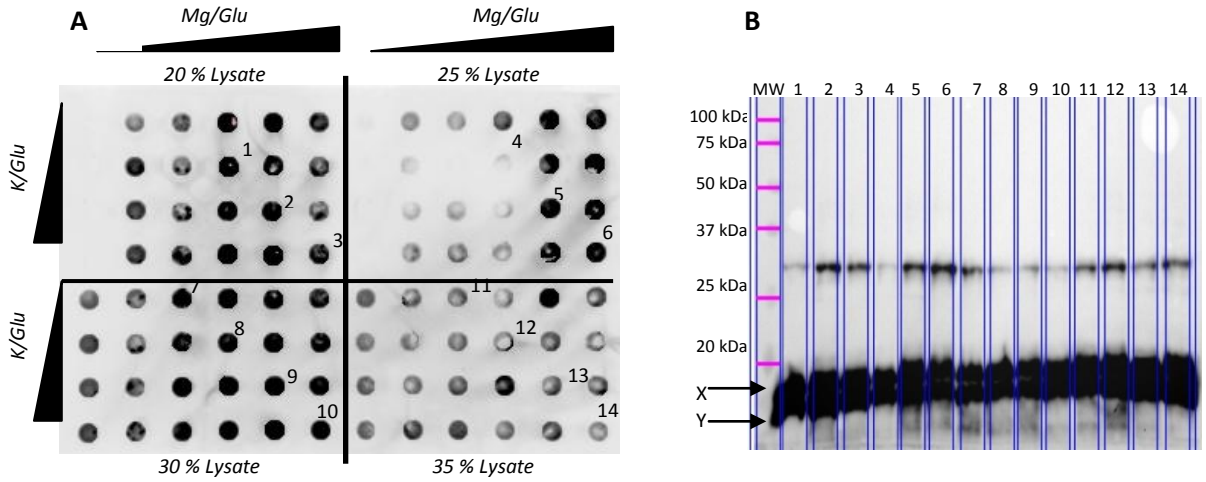



***Figure 2.12:*** *Dot blot and western blot analysis of CFPS using Nico21-derived lysate to express histidine-tagged BAK protein in liposomes. Using the dot blot, a screen of magnesium glutamate (1.6 – 9.6 mM; Mg/Glu), potassium glutamate (200 – 300 mM; K/Glu) and lysate (20 – 35 %) concentrations was carried out to validate and identify optimum conditions for CFPS (A). Numbered conditions from the dot blot were picked for western blot in order to confirm correct expression of full-length BAK. Arrows indicates full-length (X) and truncated BAK specimen (Y) (B). MW: Molecular weight.*

### *Comparison of BAK expression using other lysates by dot blot and western blot analysis*

Once preliminary dot blot analysis had been carried out on the standard Nico21 lysate, the expression of BAK in proteoliposomes using the other lysates was tested (Figure 2.13). To reduce the number of variables under investigation, the lysate and K/Glu concentrations were therefore maintained as constants at respectively 30 % and 300 mM. From the dot blot results it can be seen that BAK could be expressed in all of the lysates for the three Mg/Glu concentrations tested (Figure 2.13A). Western blot allowed the confirmation that BAK and its dimer had been expressed with molecular weights of 16 and 32 kDa, respectively (Figure 2.13B). Furthermore, from the western blot, it could be observed that there was relatively more BAK expressed when 8 mM Mg/Glu was used as compared to 9.6 mM. The highest levels of BAK and its dimer were expressed in the Rosetta and Shuffle strains, and the least amount of BAK was produced in C43 and Rosetta-Gami. BAK expressed in Rosetta lysate derived from fermenter culture in H-Enfors (HL) showed very strong expression that was comparable to Rosetta lysate that was made using the established protocol in flasks in 2YPTG media. Therefore, with the need for scaling-up production, it can be established that bacterial culture in fermenters, with the benefit of automation and control over culture conditions, can be achieved.

Further confirmation on the relative yields of BAK could be observed with the subsequent Coomassie blue gel analysis where the presence of BAK can be seen. With reference standard Nico21, expected 100 µg/mL of monomeric BAK was expressed with the use of 8.0 mM Mg/Glu. Thus, estimates on the amount of BAK produced in the other lysates, relative to that produced in Nico21, could be determined. In C43 and Rosetta-Gami, approximately 50 µg/mL of monomeric BAK was expressed with 8.0 mM Mg/Glu. In SHuffle, Rosetta and RH, approximately 200 µg/mL of monomeric BAK were produced with 8.0 mM Mg/Glu. These yields were nearly halved when Mg/Glu was increased to 9.6 mM.

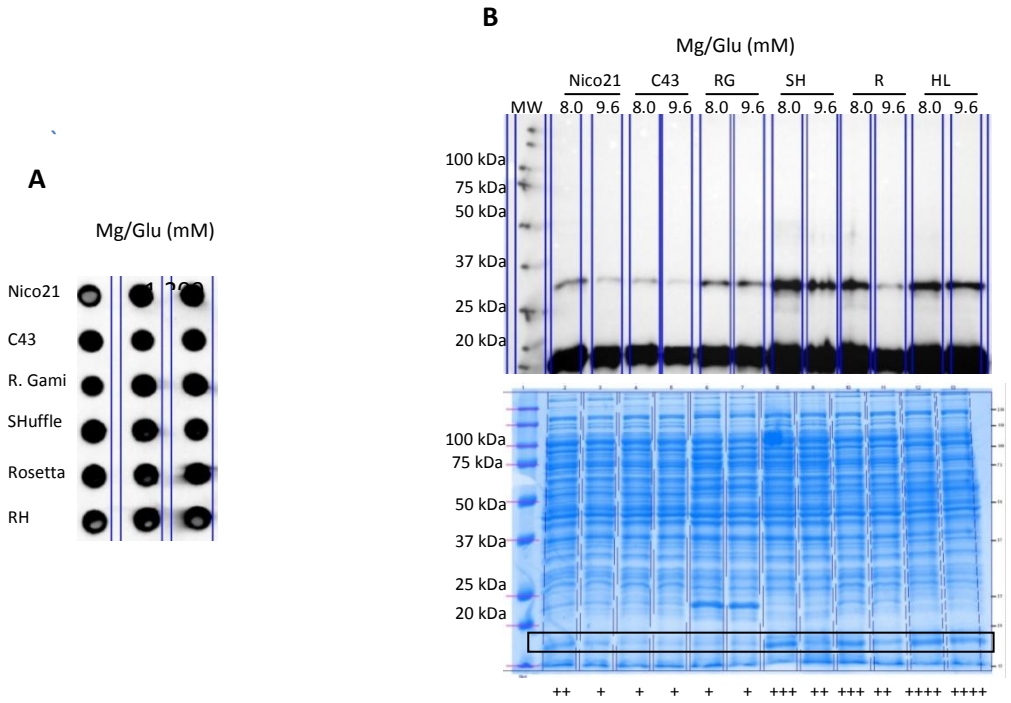



***Figure 2.13**: Dot blot, western blot and Coomassie-stained SDS-PAGE gel analyses of CFPS reactions using various lysates to express histidine-tagged BAK protein in liposomes. Keeping potassium glutamate (300 mM) and lysate (30 %) amounts as constants, a screen of magnesium glutamate concentration (6.4 – 9.6 mM) was carried out to validate and identify optimum conditions for CFPS (A). From the dot blot analysis samples were used for western blot and Coomassie analyses in order to confirm correct expression of full-length BAK (B), and to provide estimates of the protein yield. RG: Rosetta-Gami, SH: SHuffle, R: Rosetta lysate derived from flask culture, HL: hydrogenated lysate (Rosetta lysate derived from HCDC H-Enfors). Plus signs indicate the relative amount of BAK produced in each of the corresponding lanes. MW: Molecular weight.*

### *Comparison of VDAC expression using deuterated lysates by western blot and Coomassie gel analysis*

Since it resulted in the highest yields, CFPS conditions supplemented with $Mg^{2+}$ at a concentration of 8 mM was used as the basis for further screening to optimise the reactions. The functionality of Rosetta *E. coli* lysate produced in HCDC in fermenters H- (hydrogenated lysate, HL) and D- Enfors medium -

(perdeuterated lysate, PL and matchout lysate, ML)was verified by western blot and Coomassie analyses of histidine-tagged VDAC (Figure 2.14A and B). Compared to control Rosetta *E. coli* produced in flasks, CFPS reactions HL showed nearly the same high levels of expression of VDAC. However, expression of VDAC was completely lost using PL (Figure 2.14, lane PL). Very slight restoration of VDAC expression could be observed with the use of ML (Figure 2.14, lane ML). To verify that the presence of liposomes is not detrimental to the reaction, CFPS without the inclusion of liposomes was carried out in order to express VDAC as a precipitate (Figure 2.14B). Expression of VDAC was however identical whether liposomes were included or not in the CFPS reaction. This is therefore indicative of the detrimental effect of deuterated media on CFPS. Rosetta lysate produced in 2YPG media and HL had comparable expression levels to reference Nico21 lysate, which was estimated to be about 150 µg/mL of reaction.

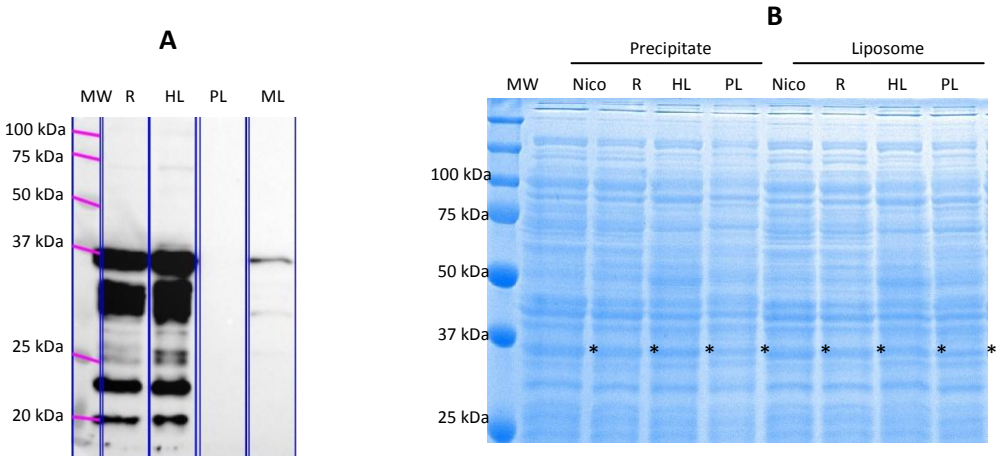



***Figure 2.14:*** *Western blot (A) and Coomassie-stained SDS-PAGE gel (B) analyses of CFPS reactions using Rosetta lysates derived from hydrogenated and deuterated HCDC to express histidine-tagged VDAC protein in liposomes, or as precipitate (without inclusion of liposomes in CFPS reaction). Expression was observed using hydrogenated-based lysate, whereas expression was compromised with deuterated-based lysate. R: Rosetta lysate derived from flask culture, RH: Hydrogenated lysate (Rosetta lysate derived from HCDC in H-Enfors), PL: Perdeuterated lysate (Rosetta lysate derived from HCDC in 100 % D-Enfors) ML: Matchout lysate (Rosetta lysate derived from HCDC in 85 % D-Enfors). MW: Molecular weight.*

## 2.4. DISCUSSION

### 2.4.1. Comparison of bacterial growth rates and biomass yields

Compared to the reference Nico21 strain (Figure 2.10), BL21-derived C43 *E. coli* had a particularly high maximum growth and growth rate, which could be explained by its origin as a survivor of intense selection pressures during toxic recombinant protein expression experiments (Miroux and Walker, 1996). The SHuffle strain had a relatively lower growth rate and lower maximum cell mass (Figure 2.10), but not due to its intrinsic mutations, since previous findings have described it as having an identical growth rate to its *E. coli* parent (Lobstein *et al.*, 2012). Rather, as a derivative of the K12 strain, the growth of SHuffle could be attributed to the difference in its glucose-acetate metabolism, and the rapid accumulation of the inhibitory acetate by-product when cultured in high-glucose medium (Marisch *et al.*, 2013, Shiloach *et al.*, 1996, Phue and Shiloach, 2004). Furthermore, instead of being directed towards increase in cell mass (and therefore the final lysate product), carbon from the glucose is wastefully converted to acetate. Indeed, it was reported that approximately 16 % of supplied carbon was lost to acetate production instead of cell mass in the K12 strain as compared to the BL21 strain (Marisch *et al.*, 2013). Here, compared to the BL21 Nico21 strain, it is interesting to note that about 20 % less of wet weight SHuffle K12 strain cellular mass was obtained. Several strategies could be considered to limit the formation of acetate in cultures, but using the commercially available BL21-derived strain of SHuffle could be the best solution (Eiteman and Altman, 2006). Rosetta had practically an identical high growth rate and cell mass yields to Nico21 (Figure 2.10). This, however, contrasts with Rosetta-Gami that experienced the lowest growth rate and yielded least cellular mass.

### 2.4.2. Large-scale hydrogenated and deuterated bacterial culture

With the need for increased lysate production in the context of coping with industrial demand and exploitation in small-angle neutron scattering (SANS) structural biology experiments, protocols typically used for recombinant protein expression in HCDC in fermenters in hydrogenated and deuterated Enfors medium (Haertlein *et al.*, 2016) were adapted for the production of Rosetta (DE3) *E. coli* lysate (Figure

2.11, 2.13, 2.14). These results therefore lead the way towards high-end and automated control on the growth of *E.* coli – a critical step in large-scale industrial CFPS development.

Typically, *E. coli* for lysate production for CFPS is cultured in 2YPTG media which sustains high density bacterial growth and population. Functional lysate was reported to have been produced when cells were growth in flasks in either glucose-devoid YT or H-Enfors (Failmezger *et al.*, 2017, Borkowski *et al.*, 2018). These studies were, however, mostly concerned with the changes in transcription-translation biochemistry of the bacteria rather than the actual production of protein. Lysate derived from fed-batch HCDC in fermenters using Enfors medium was shown to be as effective as flask cultivated 2YPTG *E. coli* for CFPS of CAT protein (Zawada and Swartz, 2006). However, the effectiveness of the lysate was determined to be dependent on the rate of growth of the bacteria from which it was derived. Extracts prepared from cultures with a specific growth rate of 0.7/hour, produced approximately 0.9 mg/mL CAT, whereas when the source culture growth rate was 0.3/hour, the resulting extract produced only 0.5 mg/mL CAT. This relates to the presence of higher ribosome content of the cell during highly active growth phases. Here, our studies demonstrated that both lysates from 2YPTG flask cultures and HCDC in H-Enfors, having a high growth rate, had comparable CFPS effectiveness (Figure 2.15). This suggests that the lower CFPS effectiveness of Rosetta-Gami lysate could be attributed to its lower growth rate, and thus, lower active ribosomal content.

### 2.4.3. Poor CFPS from deuterated media-derived lysate

Poor CFPS reactions were observed with the use of deuterated lysates (Figure 2.14). Whilst the lower growth rate in deuterated media is a factor that reduces ribosomal abundance, the intrinsic isotopic effect of deuterium can certainly hinder the transcriptional-translational machinery and as observed here, completely compromise CFPS. Indicative of the effect, functional hydrogenated lysate for CFPS which was dialysed to above 98 % $D_2O$, showed a significant 35 % reduction in translational efficacy (Etezady-Esfarjani *et al.*, 2007a). In the same fore-mentioned study, the activity of the lysate was restored with the supplementation of exogenous, purified ribosomes. However, using similar HCDC fed-batch fermenter *E. coli* culture in 100 % D-Enfors, whilst maintaining a slow growth rate of 0.3/hour, deuterated 50S-ribosomes were shown to be biologically active (Vanatalu *et al.*, 1993). Furthermore, since perdeuterated recombinant proteins are produced routinely using these methods (Haertlein *et al.*,

2016), it indicates that the transcription and translation machinery are active *in vivo*. It still remains to be tested whether the restoration of CFPS activity in deuterated lysate is possible here with further optimisation of the culture conditions and exogenous supplementation of CFPS reaction with purified active components.

### 2.4.4. Magnesium and potassium ion concentrations influence CFPS production levels

Consistent with previous observations (Shin and Noireaux, 2010), CFPS was more sensitive to changes in magnesium ions ($Mg^{2+}$) than potassium (K+). $Mg^{2+}$are indispensable ions serving as an activating cofactor for a large number of enzymes involved in metabolism and nucleic acid biochemistry (Cowan, 2002, Hartwig, 2001). As a cofactor of DNA polymerase and a stabiliser of DNA strands, limited access and depletion of $Mg^{2+}$ rapidly inhibits DNA transcription, replication and annealing, a well-known problem in the polymerase chain reaction (PCR) (Opel *et al.*, 2010). During protein synthesis, $Mg^{2+}$ is involved in amino acylation of tRNA (Thiebe, 1975), and the formation and preservation of ribosomal complexes (Bonincontro *et al.*, 1993, Favaudon and Pochon, 1976). As such, depletion of $Mg^{2+}$ in CFPS due to sequestration as inorganic phosphates and its re-supplementation caused respectively inhibition and restoration of chloramphenicol acetyl transferase (CAT) synthesis (Kim and Swartz, 2000).

In CFPS, $Mg^{2+}$ is a critical component for transcription to take place. However, due to the presence of the complex protein lysate mixtures, higher concentrations of $Mg^{2+}$ is required (Kim *et al.*, 2006b). However, as shown here (Figure 2.13), excessive $Mg^{2+}$ can also compromise CFPS. Consistent with previous studies, concentrations of $Mg^{2+}$ greater than optimum caused a decline in cell-free expressed CAT. This was because of the $Mg^{2+}$ chelating effect of the energy source phosphoenolpyruvate and the resulting accumulation of inhibitory inorganic phosphates at a reported critical value of 40 mM. Further evidence of the inhibitory effect of excessive $Mg^{2+}$ was demonstrated in transcription experiments where RNA polymerase reaction inorganic by-product pyrophosphates, rapidly reacts, precipitates and depletes $Mg^{2+}$ when the critical saturation point is reached (Akama *et al.*, 2012).

### 2.4.5. Increased protein production in CFPS using Rosetta and SHuffle lysate

*E. coli* has been a very popular choice for recombinant protein expression *in vivo* and it has undergone extensive genetic modifications to specifically suit its expressing capabilities. These advantages can be readily transferred to CFPS as yields of proteins were boosted with the use of Rosetta and SHuffle strains here (Figure 2.13). This is consistent with previous studies in which Rosetta-based lysate showed increased CFPS yields of model CAT protein (Kigawa *et al.*, 2004). Correctly folded and high yields of disulphide-containing cytokine were produced on an industrial scale using CFPS when exogenous disulphide bond isomerase C (DsbC) was added to the reaction (Zawada *et al.*, 2011a). It can be deemed that the improved protein yields from SHuffle™ lysate here is due to the presence of Dsb, which is constitutively expressed in this modified strain (Lobstein *et al.*, 2012). Obviously, this presents economic advantages as it the enzyme does not have to be purchased or produced separately. Whilst no reports on the use of Rosetta-Gami as CFPS lysate could be found, the use of Origami™ BL21 lysate, which contains mutations in thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) to enhance disulphide bond formation by lessening the cytoplasmic reducing microenvironment, have proven to yield less heterologous protein compared to the standard BL21 (Michel and Wuthrich, 2012, Kang *et al.*, 2005).

These results therefore lead the way towards the prospects of improving the CFPS lysate with the use of novel bacterial strains. Additionally, one can also consider the use of mixed prokaryotic and eukaryotic lysates, which if optimised, would mean the high production levels of correctly folded proteins with post-translational modifications.

## 2.5. CONCLUSIONS

The efficacy of CFPS is highly dependent on the quality of the lysate. Standardising and optimising the lysate production methods are particularly important for large-scale industrial applications (Zawada and Swartz, 2005, Zawada *et al.*, 2011a). By defining the growth curves of various bacterial strains, references for future cultures have been established. Albeit that only a small number of samples were considered here (n=4), it is expected that these results will help in ensuring culture reproducibility and industrial quality assurance. Furthermore, description of the curves can be used as predictive models for future cultures, and assist in ensuring cell harvest at the ideal time for highest quality lysate. It is also possible to relate the CFPS efficacy to factors such as harvest point or induction timing, in order to further optimise growth protocols and subsequent protein synthesis reaction. Furthermore, in the context of quality control in industry, where the standardisation, reproducibility and traceability of lysate batches is critical, it is important to record the growth profiles of new strains, which will serve as references for future cultures.

Bacteria for the lysate production were successfully grown using fed-batch HCDC in fermenters. Fermenter cultures provide larger amounts of lysate and, with visualisation of the growth conditions, allow both manual and automated control over the culture – which again is an important feature for an industrialised process. Furthermore, the use of HCDC in fermenters is in line with the development for large-scale CFPS. Finally, the protocol for HCDC in fermenters was adapted for the production of deuterated lysate. With further optimisation the use of deuterated lysate for CFPS, as described later (see chapter seven), can have very important implications in the field of SANS.

The functionality of several non-standard bacterial strains was confirmed and lysates with better CFPS yields were successfully identified. Namely, Rosetta and SHuffle lysate were found to be superior in terms of expression of model proteins compared to the standard Nico21 lysate. Consequently, as described in the next chapter, for the purpose of this project, where disulphide bond formation is critical for the functionality of particularly SDF1-α and CD4, SHuffle lysate was preferentially used. It can therefore be envisioned that lysate derived from other *E. coli* strains may further improve CFPS. Moreover, mixing prokaryotic and eukaryotic lysate for enhanced CFPS is an exciting prospect.

# CHAPTER THREE

## OPTIMISATION OF PROTEIN EXPRESSION AND PURIFICATION

## ABSTRACT

Following lysate preparation, CFPS was used for the recombinant expression of CXCR4, CD4 and SDF1-α. CXCR4 was expressed in two forms: embedded in liposomes and in soluble form in the presence of detergents. Normally occurring as an insoluble protein when expressed in *E. coli* and therefore requiring a refolding procedure, the canonical ligand of CXCR4, SDF1-α, was produced directly in soluble form using CFPS. With the large amounts of protein that are typically required for structural studies, the expression and purification of each of these proteins in *E. coli* CPFS had to be optimised: the impact of lysate type used, expression times, temperature and Mg/Glu content of CFPS reactions were investigated. For the expression of CXCR4 in detergents, a screen of detergents at varying concentration was carried out. Despite the large array of tests considered here, the expression and purity of CXCR4 remained unfortunately too low to be carried forwards with advanced structural studies. However, the expression of SDF1-α and 2dCD4 did allow the proteins to be carried forwards. CPFS of these two proteins was therefore optimised in order to produce them in correctly-folded and solubilised form and with the establishment of scaled-up CFPS production methods (up to 100 mL) and purification protocols, adequate amounts of these proteins were recovered to allow for structural studies. Driven by the application of SANS, optimised protocols for deuterated forms of proteins were also established. Following this work, and in the context of establishing reliable protocols for future CPFS in the industrial context, standard operating procedures (SOPs) were established.

## 3.1. INTRODUCTION

### 3.1.1. Difficulty and cost of GPCR expression

Mammalian GPCRs expressed *in vivo* in *E. coli* typically occurs as aggregated protein in the form of inclusion bodies which require extensive optimised and laborious denaturing, solubilisation and refolding protocols in order to obtain functional proteins (Sarramegna *et al.*, 2003). Mammalian cell expression, the ideal form of expression, however remains a very costly method and expertise-demanding process for protein production. As a compromise, the baculovirus-insect cell expression has instead been used for the expression of the lysozyme-mutant β-adrenergic receptor, which led to the elucidation of the first crystal structure of recombinant expressed GPRC (Cherezov *et al.*, 2007). Following this pioneering work, the system was then adapted for the expression of lysozyme-mutant CXCR4, the structure of which was then determined by X-ray crystallography (Wu *et al.*, 2010). However, similar to mammalian cell-expression, the baculovirus-expression cell system remains an expensive and skill-demanding method for protein expression. Coupled with the need to adhere to strict safety legislation for working with viruses, the baculovirus-expression cell system furthermore has low genomic stability and currently cannot sustain heterologous protein expression for long periods of time (van Oers *et al.*, 2015). These features have limited the extensive application of the baculovirus-expression cell system for large-scale application and within industry.

### 3.1.2. CFPS as an efficient technology for difficult-to-express proteins

CFPS is a powerful alternative method for recombinant protein expression. The openness of the system makes it particularly advantageous as it allows the modulation of the reaction conditions for the production of difficult-to-express proteins. Limited by supply, many proteins are in constant demand by pharmaceutical and medical research, in both industrial and academic settings. As such, CFPS is powerful potential alternative for the expression of difficult-to-express proteins, including GPCRs, and which has shown promise for its potential in large-scale protein expression for industrial exploitation (Zawada *et al.*, 2011a).

### *Previous methods for production of CXCR4 for structural studies*

Following the elucidation of the crystal structure of β-adrenergic GPCR as the first recombinant membrane protein, the crystal structure of CXCR4 (2.5 Å), albeit heavily mutated, was soon to follow (Figure 3.1) (Wu *et al.*, 2010, Qin *et al.*, 2015). As a member of the GPCR family, CXCR4 is structurally characterised by an extracellular N-terminus, seven transmembrane (TM) α – helical domains, linked to each other with three intracellular (IL) and three extracellular loops (EL), and ending with an intracellular C-terminus. CXCR4 then embeds itself into the lipid bilayer forming a barrel-like cavity.

In previous studies, CXCR4 was produced in baculovirus-infected insect cells *Spodoptera frugiperda* (Sf9) (Wu *et al.*, 2010, Qin *et al.*, 2015). The over-expressed CXCR4 was then extracted and purified from the insect cell plasma membrane using mild detergent based buffers. Using highly purified and homogenous concentrated samples (60 – 70 mg/mL of protein), CXCR4 was then successfully crystallised in the lipid cubic phase (LCP). Whilst being successful, the baculovirus-insect cell expression system can be quite costly for industrial scale application. As well as requiring expensive reagents, the system cannot be sustained for long and requires re-infection of fresh cells after a few passages. Furthermore, due to the use of viruses, the system needs to conform to strict safety and legislative protocols (van Oers *et al.*, 2015).

CFPS represents a powerful system for the expression of membrane proteins. CXCR4, and its analogue chemokine receptor CCR5, have been produced in *E. coli* based CPFS systems by Synthelis and others (Chi *et al.*, 2015, Chi *et al.*, 2016, Chadli *et al.*, 2017). This successful production was made possible by screening various lysates and constructs, and by taking advantage of the openness of the system, soluble and correctly folded chemokine receptors were produced following the inclusion of detergents directly in the reaction. Further enhancement of folding could also be achieved with the inclusion of GroEL – GroES chaperones. With the inclusion of liposomes in the reaction, active GPCR fit for immunisation purposes was produced in wheat germ CPFS (Takeda *et al.*, 2015).

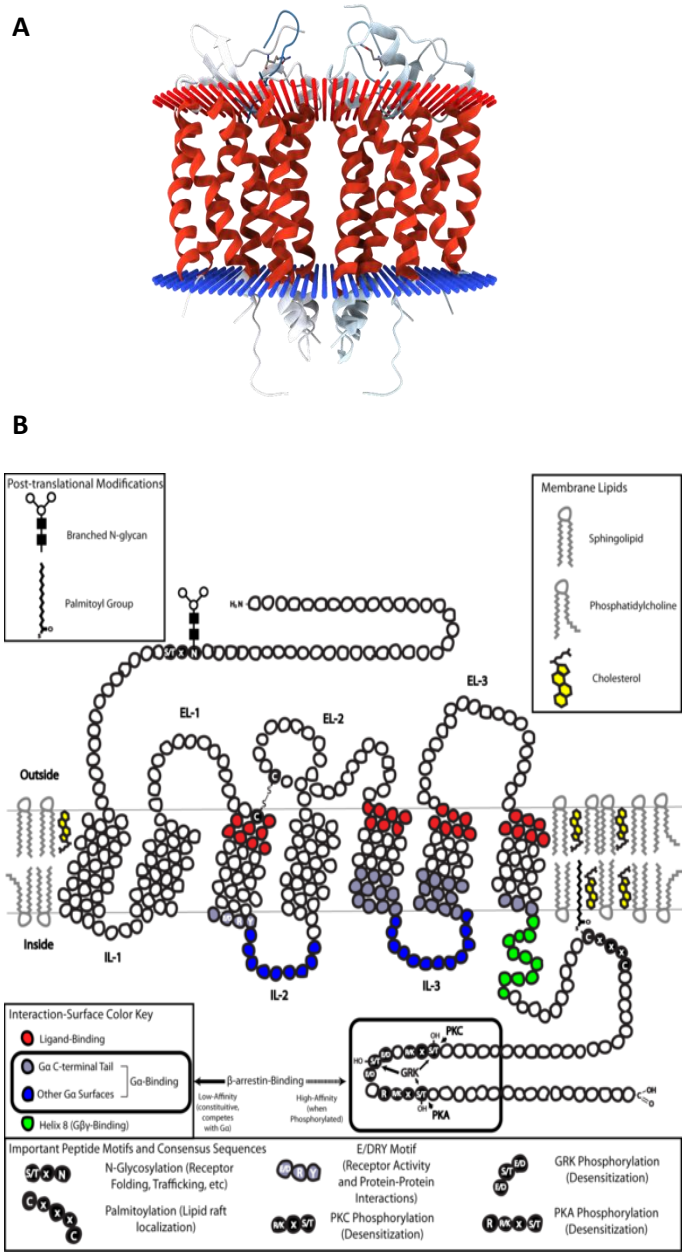



***Figure 3.1:*** *Structure of CXCR4 (PDB: 3OE0) (A) and schematic representation of a typical GPRC in the lipid bilayer (B; https://commons.wikimedia.org/wiki/File:GPCR_in_membrane.png; original author: Repapetilto).*

### *Previous methods for production of SDF1-α for structural studies*

SDF1-α, as a member of the CXC chemokine family, is characterised by the conserved C-X-C motif whereby there is the occurrence of an amino acid between the first two cysteine residues. SDF1-α also contains two disulphide bonds (Figure 3.2).

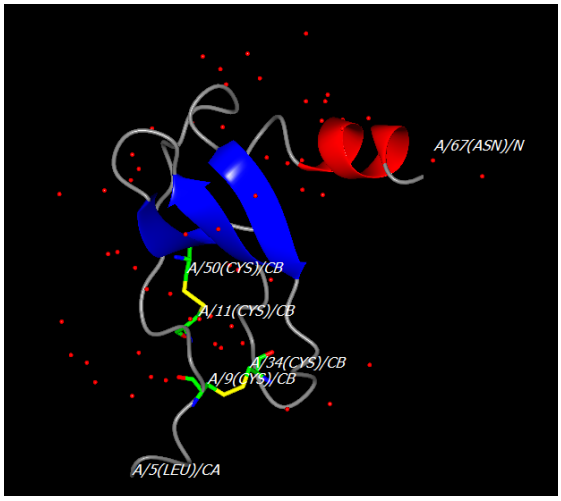



***Figure 3.2:*** *Ribbon diagram of SDF1-α (PDB: 3GV3, (Murphy et al., 2010), with α-helices in shown red and β-sheets in blue. Red dots represent water molecules. Cyteine residues are shown in green, and disulphide bonds are shown in yellow.*

Highly similar crystal structures of SDF1-α were obtained using SDF1-α made by chemical synthesis (Dealwis *et al.*, 1998), recombinant mammalian cell expression (Ohnishi *et al.*, 2000) and recombinant *E. coli* expression (Ryu *et al.*, 2007, Smith *et al.*, 2014). Chemically synthesis SDF1-α, which was made by the ligation of the N-terminal and C-terminal peptides, was possible since SDF1-α is a relatively small molecule. Chemically synthesising larger molecules is generally not possible. Furthermore, the method is costly, uses highly toxic chemical and requires highly experienced for personnel. Mammalian SDF1-α whilst being most native form of the protein, is again a costly method for production of the protein, requiring specialised laboratory units and trained personnel. SDF1-α expression in *E. coli* is therefore the cheapest and most productive method. However, SDF1-α occurred as insoluble inclusion bodies that needed an extensively optimised and labour-intensive refolding process in order to obtain folded protein. Attempts to produce soluble SDF1-α in *E. coli* fusing it to maltose binding protein (MBP) have been used, but stability of the protein was lost once the MBP removed (Picciocchi *et al.*, 2014).

Demonstrating its potential, *E. coli* based CPFE was used to produce an active soluble cytokine human (granulocyte-macrophage colony-stimulating factor) on the industrial scale (Zawada *et al.*, 2011a).

### *Production of CD4 for structural studies*

CD4 is an elongated rod-shaped molecule that can be divided into three regions: the extracellular (residues 26 – 396), transmembranal (residues 397 – 418), and cytoplasmic (residues 419 – 458) regions. Starting most distal from the cell surface membrane, the extracellular region contains four distinct domains: domain 1 (D1, residues 26 – 125), domain 2 (D2, residues 126 – 203), domain 3 (D3, residues 204 – 317) and domain 4 (D4, residues 318 – 374).

Typically using the first two domain of CD4 (2dCD4) for crystallography, the molecule is generated using mammalian expression systems (Figure 3.3) (Ryu *et al.*, 1990, Acharya *et al.*, 2014). Mammalian expression can also be used for producing full-length CD4 (Kwong *et al.*, 1990). Alternatively, 2dCD4 or full-length can also be produced in baculovirus-infected insect cells (Webb *et al.*, 1989). In any case, recombinant CD4 expressed lacks the transmembranal domain and is secreted as a soluble protein. *E. coli* expression of 2dCD4 results in the occurrence of inclusion bodies from which the protein needs extracted using harsh denaturing conditions and subsequently refolded using extensive optimised dialysis protocols (Garlick *et al.*, 1990, Cerutti *et al.*, 2010). Whilst cheaper, *E. coli* expressed CD4 does not attain the same folding quality as the mammalian expressed counterpart. Furthermore, the yields of the protein are quite low since lot of it is lost due to precipitation during refolding.

CPFE can be used as an alternate system for expression of difficult-to-express and aggregating proteins. A strong contributor for improper folding can be attributed to the reducing microenvironment of *E. coli* inhibits disulphide bonds formation. The modulation of the chemical environment in CPFE can be altered to enhance the disulphide bond formation and produce correctly folded protein (Zawada *et al.*, 2011a). In addition, once expression confirmed, labelled proteins can easily be produced. Furthermore, being a membrane protein, the addition of liposomes or detergents can be considered in order to produce full-length CD4 including transmembrane and intracellular domains.

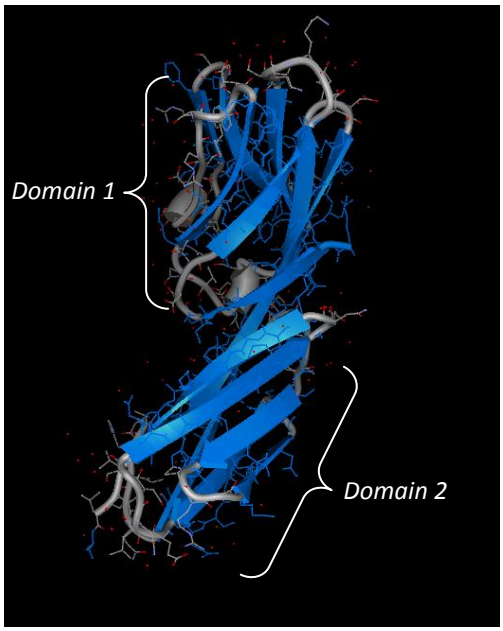



***Figure 3.3:*** *Structure of first two domains of CD4, termed 2dCD4 (PDB: 1CDJ) (Ryu et al., 1990). B-sheets are shown in blue.*

### 3.1.3. Generating standard operating procedures

In the context of industry, time is an important resource which needs to be allocated optimally in the various tasks of CFPS in order to boost production rates. Therefore, once optimised protocols have been established, these should serve as a general template to be used for producing other proteins in CPFS.

The work presented in this chapter leads the way towards the implementation of standard operating protocols (SOPs) for the efficient production of proteins to be used for analytical, large-scale or structural purposes. SOPs are defined as a set of instructions that describes all of the relevant steps and activities of a process and is typically provided as a workflow for easy interpretation. SOPs thus serve to breakdown complex routine operations, simplifying the tasks and can be applied by users, even having limited experience. As such, these protocols can help to ensure accuracy, reproducibility and performance, which are particularly critical in the industrial setting.

## 3.2. MATERIALS AND METHODS

### 3.2.1. Plasmid constructs

Three constructs of human CXCR4 were considered: wild-type (*CXCR4-wt*), wild-type strep-tag (*CXCR4-h/s*) and lysozyme mutant (*CXCR4-lys*). The CXCR4-wt sequence was obtained from the UniProtKB database (P61073). The plasmid for *CXCR4-wt* was provided by Synthelis (France). *CXCR4-strep* was identical to *CXCR4-wt*, but included a strep-tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) at the C-terminal. The *CXCR4-lys* sequence, based on previous X-ray crystallography work (Wu *et al.*, 2010), was obtained from RCSB PDB database (accession code: 3ODU). All constructs included, sequentially, a His-tag (6X histidine) and factor Xa cleavage site (Ile-Glu-Gly-Arg) at the N-terminal, and were encoded in pIVEX plasmid 2.4d between the *NcoI* and *XhoI* restriction sites. The plasmids were synthesised by Proteogenix (France).

Four constructs of human SDF1-α were considered: *NHis-SDF1*, *CHis-SDF1, SDF1-His* and *SDF1-LH.* The sequence of wild-type human SDF1-α was obtained from the UniProtKB database (accession code: P4801). *NHis-SDF1* was encoded in pIVEX 2.4d plasmid and included a His-tag and thrombin cleavage site at the N-terminal. *CHis-SDF1* was encoded in pIVEX 2.3d plasmid and included a TEV cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-Gly) and His-tag at the C-terminal. These plasmids were encoded between the *NcoI* and *XhoI* restriction sites and synthesised by Proteogenix (France). *SDF1-His* and *SDF1-LH,* encoded in pET20-b plasmid vector, were kindly provided by C. Vives and F. Fieschi. Both *SDF1-His* and *SDF1-LT* had a C-terminal His-tag. *SDF1-LH* included a double lanthanide-binding tag between the protein and the His-tag. Further details of these two constructs can be found in a previous publication (Picciocchi *et al.*, 2014).

Four constructs of human CD4 were considered: *CD4-FL-NHis*, *CD4-FL-CHis, 2dCD4-wt-NHis and 2dCD4-wt-Chis*. The sequence of wild-type human CD4 was obtained from UniProtKB database (accession number: P01730). *CD4-FL-NHis* and *CD4-FL-CHis* are full-length wild-type CD4, which include the transmembrane and cytoplasmic domains, whereas *2dCD4-wt-NHis* and *2dCD4-wt-Chis* comprised only the first two domains of CD4. *2dCD4-wt-NHis* and *2dCD4-wt-CHis* are wild-type variants (Cerutti *et al.*, 2014b, Cerutti *et al.*, 2010).

The placement of the histidine tag at the N-terminal has been used for CPFS reactions at Synthelis since it has been suggested to favour the expression of proteins, as opposed to when it is located at the C-terminal. However, in the case of CD4, the presence of the histidine tag could interfere with its binding to GP120. Therefore, CD4 variants with histidine at the N-terminal and C-terminal were considered. *CD4-FL-NHis* and *2dCD4-wt-NHis* included an enhancer sequence (Lys-Pro-Tyr-Asp-Gly-Pro) as described previously (Haberstock *et al.*, 2012), His-tag and TEV cleavage site at the N-terminal. *CD4-FL-CHis* and *2dCD4-wt-CHis* included a TEV cleavage site and His-tag at the C-terminal. The plasmids used are summarised in Table 3.1.

| Plasmid | Plasmid vector | Protein expressed | Tag | Interest for construct |
|---|---|---|---|---|
| ***CXCR4-wt*** | pIVEX 2.4d | CXCR4 wild-type | Histidine tag at N-terminal | Wild-type protein |
| ***CXCR4-lys*** | pIVEX 2.4d | Mutant lysozyme CXCR4 | Histidine tag at N-terminal | Lysozyme CXCR4 mutant successfully used for crystallisation in previous publications (Wu, Chien *et al.* 2010) |
| ***CXCR4-h/s*** | pIVEX 2.4d | CXCR4 wild-type | Histidine tag at N-terminal and strep-tag at C-terminal | Wild-type protein, with strep-tag at C-terminal to confirm completed protein translation |
| ***NHis-SDF1*** | pIVEX 2.4d | SDF1-α wild-type | Histidine tag at N-terminal | Initial construct of wild-type SDF1-α with His-tag at N-terminal which is typically better expressed in CFPS |
| ***CHis-SDF1*** | pIVEX 2.3d | SDF1-α wild-type | Histidine tag at C-terminal | Due to presence of His-tag at the N-terminal which impedes functionality, construct of wild-type SDF1-α with His-tag a C-terminal is considered |
| ***SDF1-His*** | pET20-b | SDF1-α wild-type | Histidine tag at C-terminal | Construct used for *E. coli in vivo* expression (Picciocchi *et al.*, 2014), which was used as a trial for CFPS |
| ***SDF1-LH*** | pET20-b | SDF1-α wild-type | Double lanthanide-binding tag and histidine tag at C-terminal | Construct used for *E. coli in vivo* expression (Picciocchi *et al.*, 2014), which was used as a trial for CFPS |
| ***CD4-FL-NHis*** | pIVEX 2.4d | Full-length CD4 wild-type | Histidine tag at N-terminal | Full-length CD4 with His-tag at N-terminal that is typically better expressing in CFPS |
| ***CD4-FL-CHis*** | pIVEX 2.4d | Full-length CD4 wild-type | Histidine tag at C-terminal | Full-length CD4 with His-tag at C-terminal that does not impede functionality and binding to GP120 |
| ***2dCD4-wt-NHis*** | pIVEX 2.4d | Truncated two-domain CD4 wild-type | Histidine tag at N-terminal | Soluble 2dCD4 with His-tag at N-terminal that is typically better expressing in CFPS |
| ***2dCD4-wt-Chis*** | pIVEX 2.4d | Truncated two-domain CD4 wild-type | Histidine tag at C-terminal | Soluble 2dCD4 with His-tag at N-terminal that does not impede functionality and binding to GP120 |

***Table 3.1:*** *Summary of plasmids used for CPFS.*

As described in the previous chapter, the plasmids were transformed and amplified in *DH5α E. coli* purified using MaxiPrep (Macherey-Nagel, 300 µL elution) or GigaPrep (Quiagen, 10 mL elution). Plasmid DNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific$^{TM}$). Using stocks at 500 ng/µL, the corresponding plasmids were then used in CFPS reactions at final concentrations of 15 ng/µL.

### 3.2.2. Liposome CFPS reactions

#### *Liposome Preparation*

Liposome mixture, consisting of 1,2-dioleoyl-sn-glycero-3-phosphocholine, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine, 1,2-dimyristoyl-sn-glycero-3-phosphate and cholesterol in respectively 4:2:2:1 volume ratio, was prepared in chloroform at 10 mg/mL. All lipids were purchased from Avanti Polar Lipids. Chloroform was evaporated using a vacuum concentrator (UniEquip UNIVAPO 150H). The resulting thin lipid film was rehydrated with DEPC-treated water to obtain a lipid slurry at 30 mg/ml. The slurry was sonicated using a tip sonicator (Branson Digital Sonifier 250) at 20 % for five times for 30 s each before being filtered once with a 0.22 µm PES filter. The liposomes were added to CFPS reaction mix at final concentration of 5 mg/mL.

#### *CXCR4 variants expression in various lysates*

As described previously, liposome-included CPFS reactions (30 $^{o}$C, 16 h, 200 rpm) with the various lysates (30 %) were carried out using *CXCR4-wt* and *CXCR4-lys*, with Mg/Glu concentration being varied between 6.4 – 9.6 mM and K/Glu concentration being kept constant at 300 mM. The expression results were analysed by dot blot and western blot. Based on the resulting expression yields, the best lysates were then selected for expression of the remaining CXCR4 variants.

### *Time-course expression study*

Using CXCR4-wt expression in liposomes as the model, attempts to improve expression and purify the liposome were considered. A time-course assay of CXCR4-wt expression in liposomes was carried out in order to identify when expression produces the highest full-length protein to truncated species ratio. In batches of 200 µL in sterile 1.5 mL Eppendorf tubes, samples of 10 µL were drawn from standard CFPS reactions (30 % lysate, 8.0 mM $Mg^{2+}$, 300 mM $K^{+}$, 30 $^{o}$C, 16 h, 500 rpm in Thermomix) at 5, 15, 30 and 60 min, then at 1.5, 2, 3, 4, 5, 6 and 16 h following setup. Each sample was mixed with 5 X loading buffer and used for analysis by SDS-PAGE and anti-histidine western blot.

### *Purification of proteoliposomes with buffer washes*

To wash proteoliposomes, cell-free reactions were diluted with at least five times of its volume with 50 mM HEPES at pH 7.5. The tubes were then centrifuged at 30 000 g at 4 $^{o}$C for 30 min. The supernatant was collected and the proteoliposomes (a white insoluble precipitate) were resuspended in 50 µL of 50 mM HEPES pH 7.5 buffer. Samples of supernatants were used for SDS-PAGE and anti-histidine western blot.

Washed CXCR4-wt proteoliposomes produced previously in batches of 200 µL using standard CFPS conditions, were centrifuged (30 000 g at 4 $^{o}$C for 30 min) and resuspended in NaCl (1, 3 and 5 M), ammonium sulphate (30 %, 40 % and 50 % w/v), urea (1, 4 and 8 M) or sodium carbonate (1 M) solutions. Except for the sodium carbonate sample which was incubated for 10 minutes, the other samples were incubated on ice for 30 minutes before centrifugation. The sodium carbonate pellet was then resuspended in fresh sodium carbonate and incubated on ice for another 10 minutes before centrifugation. Incubation of the pellet in sodium carbonate and centrifugation was repeated once more. This protocol for the treatment of proteoliposome with sodium carbonate was possible since proteoliposomes remain stuck to the bottom of the microtube and are therefore not lost into the supernatant, which is not the case for the other treatments. The supernatants, pellets and liposome layers (top floating layer formed with NaCl and ammonium sulphate) of the different samples were collected. The pellets and liposome layers were resuspended in a volume of 50 mM HEPES pH 7.5 equal

to that CFPS reaction (200 µL). The supernatant, pellet and liposome were used for analysis by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

### *Use of pre-centrifuged CFPS lysate for reduction in contaminants*

In order limit the presence of contaminants adhering to proteoliposomes following CPFS, the lysate was initially centrifuged before setting up the reaction. Lysate (SHuffle) was centrifuged at 30 000 g at 4 $^{o}$C for 30 min, and the supernatant, free from insoluble precipitates, was used for CPFS reaction. In batches of 50 µL in sterile 1.5 mL Eppendorf tubes, with either the exclusion (precipitate reaction) or inclusion of liposomes, CFPS with centrifuged and non-centrifuged lysate was carried out at 20 $^{o}$C and 30 $^{o}$C. As described previously, the proteoliposomes were collected by centrifugation, resuspended in 50 mM HEPES pH 7.5. The supernatant and proteoliposomes fraction were used for analysis by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

### *Purification of proteoliposomes by sucrose-gradient ultracentrifugation*

Sucrose-gradient, consisting of three-step discontinuous sucrose gradient (60 %, 30 % and 5 %), and ultracentrifugation (200,000 g for 2 h at 4 $^{o}$C) was used to purify liposomes (Figure 3.4). 200 µL of CXCR4-wt proteoliposomes CPFS reactions, derived from centrifuged and non-centrifuged lysates, were loaded directly onto the sucrose gradient. Proteoliposome pellets, derived from non-centrifuged lysate CPFE reactions, were resuspended in 200 µl 50 mM HEPES pH 7.5 and incubated on ice until sucrose-gradient ultracentrifugation. In order to promote further dissociation of contaminants and ribosome sub-units from proteoliposomes additional lots of proteoliposome pellets derived from 200 µL of CXCR4-wt cell-free reaction, were resuspended in 200 µL of 1 M NaCl (control) or 1 M NaCl/3 mM EDTA in 50 mM HEPES pH 7.5 buffer. These were incubated on ice for 30 minutes before being centrifuged at 30 000 g for 30 minutes at 4 $^{o}$C. The supernatant was discarded and proteoliposome pellets were resuspended in 200 µl of 50 mM HEPES pH 7.5. These were then loaded onto the sucrose gradient. The lipid fraction between the 30 and 15 % sucrose interface was collected. To remove the sucrose solution, the proteoliposomes were washed with at least five times volume of 50 mM HEPES pH 7.5 buffer and were then centrifuged at 30 000 g at 4 $^{o}$C for 30 min. The supernatant was discarded and the proteoliposomes (white insoluble precipitate) were resuspended in 50 mM HEPES pH 7.5 buffer. The

proteoliposomes, incubated with 5 X loading buffer, were used for SDS-PAGE, and analysed by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

.

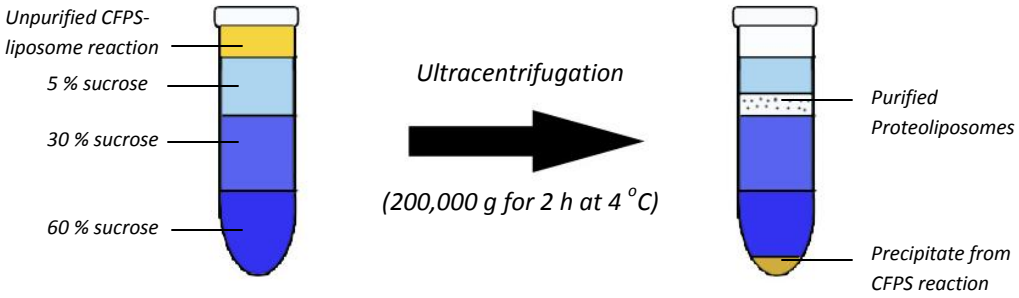



***Figure 3.4:*** *Illustration of sucrose-gradient and ultracentrifugation mode of proteoliposome purification. Liposome-based CPFS at the end of the reaction is loaded directly onto the three-step discontinuous sucrose gradient and subject to ultracentrifugation. Purified proteoliposomes, with the de novo membrane protein embedded within, are then separated from the total CPFS reaction and can be recovered.*

### *CD4 expression in proteoliposomes*

In preliminary batches of 50 µL, CD4-FL-NHis and CD4-FL-CHis CFPS reactions with liposomes were carried out using 30 % SHuffle lysate, 8 mM Mg/Glu and 300 mM K/Glu. The reactions were incubated at 17, 20 and 30 $^{o}$C for 16 h in a Thermomix at 500 rpm. As described previously, the proteoliposomes were collected by centrifugation, resuspended in 50 µL of 50 mM HEPES pH 7.5. The supernatant and proteoliposomes fraction were used for SDS-PAGE and anti-histidine western blot. Larger reaction volumes of 1 mL were carried out and purified using sucrose-gradient ultracentrifugation.

### 3.2.3. Detergent-based CFPS reactions

#### *Preliminary detergent screen for CXCR4 variants*

CXCR4-wt, CXCR4-h/s and CXCR4-lys were expressed in standard CFPS reaction (30 % lysate, 8 mM Mg/Glu, 300 mM K/Glu, 30$^{o}$C) with detergents included in the reaction mix. The detergents used in the screen were Brij35, Brij58, Brij98, digitonin, n-dodecyl β-D-maltoside (DDM), triton X-100 and CHAPSO (3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulphonate)**.**

A critical consideration in the use of detergents concerns its critical micelle concentration (CMC). The CMC is defined as the critical concentration above which the detergents form micelles which are capable of providing hydrophobic support to membrane proteins (Figure 1.3B). Detergents used at below CMC cannot form micelles and occur as monomeric detergent molecules in solution that cannot provide the hydrophobic support for membrane proteins.

Stock solutions of detergents, with corresponding critical micelle concentration (CMC) of 0.09 mM Brij35, 0.07 mM Brij58, 0.025 mM Brij98, 0.73 mM digitonin, 0.019 mM DDM, 0.55 mM Triton X-100 and 8 mM CHAPSO, were reconstituted in DEPC-treated water at 100 X or 1000 X of their corresponding CMC. DDM-cholesteryl hemisuccinate mix (DDM-CHS), which was used in the previous publication where CXCR4 lysozyme mutant was described, was also tested (Wu *et al.*, 2010). The stock solution of DDM-CHS was composed of 5 % DDM, 1 % CHS, 10 % glycerol in 50 mM HEPES at pH 7.5 with 500 mM NaCl.

For the detergent screen, CPFE reactions were carried with the various detergents at between 1 – 50 times their CMC, were carried out in 96-well plates for CXCR4-wt and CXCR4-lys (Figure 3.5). Following 16 h incubation at 30 $^{o}$C, detection of expression was achieved using anti-histidine dot blot, as described previously. Once expression was confirmed, analysis by SDS-PAGE and anti-histidine western blot, as described previously, was carried out in order to verify that CXCR4-wt and CXCR4-lys were being expressed as soluble products, and that they had the correct molecular weight. The total reaction samples from the reaction wells were taken and centrifuged at 30 000 g for 30 min. The supernatant was then extracted as the soluble portion. The total reaction and soluble portions were incubated with 5 X loading buffer for 10 minutes at room temperature, and used for analysis by SDS-PAGE and anti-histidine western blot.

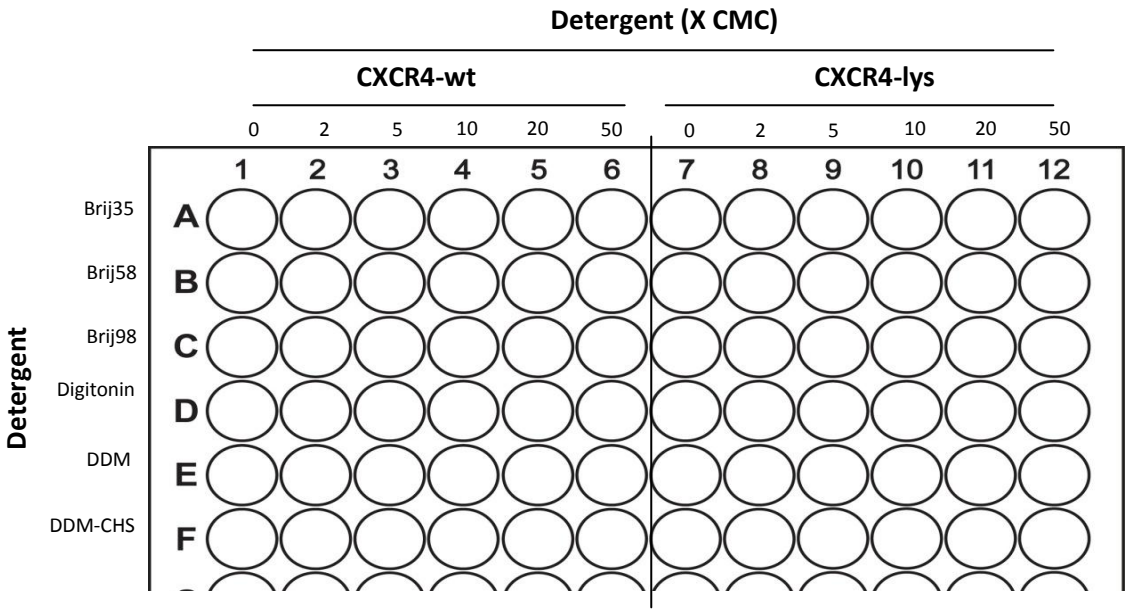



***Figure 3.5:*** *Plate setup for expression of CXCR4-wt and CXCR4-lys in the presence of detergents between 1 – 50 X CMC.*

### *Preliminary detergent screen of CXCR4 variants expressed using pre-centrifuged lysate*

With soluble expression of proteins being confirmed in the presence of detergents at relatively low CMCs, lower CMCs for the expression of soluble protein were to be then to be considered. This was considered in order to limit any inhibitory effects on expression brought about by the detergents, as well as for economic reasons.

To verify the influence of pre-centrifuged lysate in favouring soluble expression of CXCR4, reactions with detergent screen at 3 X CMC was carried out in 96-well plate for CXCR4-wt, CXCR4-lys and CXCR4-h/s, using either centrifuged or non-centrifuged Rosetta lysate (Figure 3.6). Following 16 h incubation at 30$^{o}$C, total reaction and soluble portion were produced and loaded onto a dot blot setup. Detection of expression was achieved using an anti-histidine dot blot and western blot.

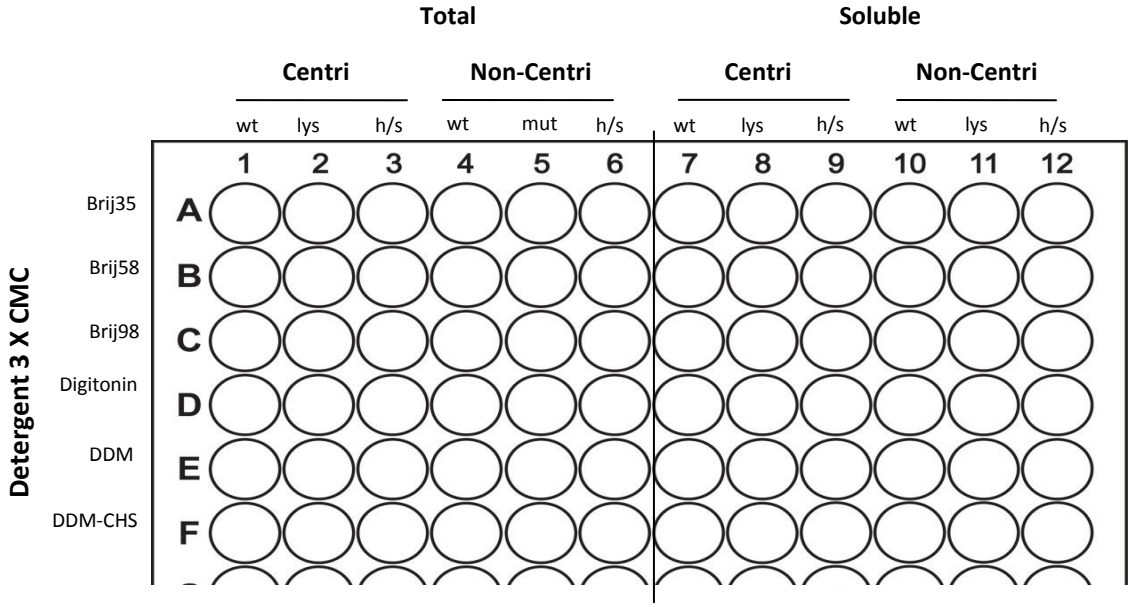



***Figure 3.6:*** *Plate setup for the verification of soluble expression of CXCR4-wt, CXCR4-lys and CXCR4-h/s in the presence of detergents at 3 X CMC, using pre-centrifuged lysate (Centri) or non-centrifuged lysate (Non-Centri).*

Whilst anti-histidine blots provide confirmation for the expression of CXCR4, it nonetheless remains to verify whether there is expression of non-truncated, full-length CXCR4. To achieve this, anti-streptavidin western blot analysis of CXCR4-h/s, which carries a strep-tag at the C-terminal, was carried out. CPFS reactions with a detergent screen, between 2 – 10 X CMC, was carried out to verify the expression CXCR4-h/s. For the purposes of carrying out an anti-strep western blot, after the blocking the membrane was then incubated with monoclonal anti-strep mouse antibody (1 µg/mL antibody in blocking buffer) for 1 h with light agitation. The membrane as then rinsed for 10 minutes three times with 1 X TBS-T buffer, and then incubated with anti-mouse peroxidase conjugate antibody (1 µg/mL antibody in 1 X TBS-T). The membrane as then rinsed again for 10 minutes three times with 1 X TBS-T buffer. The membrane was incubated with detection solution (ECL™ Prime Western Blot, GE HealthCare) and poly-histidine chemiluminescence detected on a Chemi-Doc imaging instrument (Bio-Rad, CA, U.S.A).

### *Large-scale expression and purification of CXCR4-mut*

Preliminary 50 µL batches of CFPS reactions (30 % Rosetta lysate, 300 mM K/Glu, 30 $^{o}$C), with Brij35 detergent, for the expression of CXCR4-mut was setup in 1.5 mL Eppendorfs, with a screen for Mg/Glu (4.8 – 11.2 mM). CFPS reaction with 9.6 mM Mg/Glu was then set up in 1.5 mL Eppendorf tubes, with a screen of the various detergents. After reaction, the total reaction samples were centrifuged at 30 000 g for 30 min. The supernatants were then extracted as the soluble portion. The total reaction and soluble portions were incubated with 5 X loading buffer for 10 minutes at room temperature, and used for analysis by SDS-PAGE and anti-histidine western blot.

The expression of CXCR4-mut with 5 X CMC Brij35 was scaled up to 1 mL. Small-scale purification was carried out using the MagneHis$^{TM}$ Protein purification system from Promega. An equal volume of binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 10 mM imidazole) was added to 1 mL of the reaction. 20 µl of MagneHis beads (Promega) suspension was added to the reaction, pipetted to mix and incubated at room temperature for 2 min. The supernatant was extracted and the beads were washed three times with 1 mL of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 20 mM imidazole). *CXCR4-mut* was then eluted from the beads following 2 minutes incubation in 100 µL of elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 300 mM imidazole). The presence of CXCR4-mut was verified by SDS-PAGE and anti-histidine western blot.

The expression of CXCR4-mut with 5 X CMC Brij35 was scaled up to 50 mL of batch CFPS. The reaction was vacuum filtered using 0.45 µm filter membrane and 10 mL aliquots were put into sterile capped 50 mL Falcon tubes. The flasks were incubated for 16 h at 30 $^{o}$C at 150 rpm. The completed reaction was diluted with 5 X with binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 10 mM imidazole). The reaction was filtered again using 0.45 µm filter membrane. Purification was carried out using HisTrap FF prepacked Ni Sepharose 1 mL columns (GE Healthcare) on an ÄKTAprime plus protein purification system (GE Healthcare).Binding to the column was carried at a flow-rate of 0.5 mL/min. After loading the sample, subsequent washes with binding buffer, wash buffer 1 (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 20 mM imidazole) and wash buffer 2 (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 40 mM imidazole) were carried out until no further protein was washed from the column. Elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 X Brij35, 300 mM imidazole) was used and the eluted

protein was collected in the fractions (1 mL volume) of interest. The presence of CXCR4-mut in fractions was confirmed by western blot and Coomassie-stained SDS-PAGE.

Fractions 9 – 15 containing CXCR4-mut (Figure 3.22) were pooled and concentrated to 1 mL using 100 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore). The concentrated sample was used for size exclusion chromatography (SEC) using Superdex$^{TM}$ 200 10/300 GL column (GE HealthCare). The buffer was simultaneously exchanged to degassed sterile degassed sterile DDM-CHS buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 0.05 % DDM, 0.01 % CHS, 10 % glycerol). Elutions of 500 µL fractions were collected and the presence of *CXCR4-mut* was detected using SDS-PAGE, Coomassie blue staining and anti-histidine western blot.

Purification and exchange into DDM-CHS buffer was carried out during a nickel affinity purification step. The above procedure was repeated. Following the passage of diluted 5 X brij35 reaction supernatant through a Ni column, sequential washes with DDM-CHS wash buffers 1 (25 mM HEPES pH 7.5, 500 mM NaCl, 0.5 % DDM, 0.1 % CHS, 10 % glycerol, 5 mM imidazole), wash buffer 2 (25 mM HEPES pH 7.5, 500 mM NaCl, 0.1 % DDM, 0.02 % CHS, 10 % glycerol, 40 mM imidazole), wash buffer 3 (25 mM HEPES pH 7.5, 500 mM NaCl, 0.05 % DDM, 0.01 % CHS, 10 % glycerol, 40 mM imidazole) were carried out until no further protein was being washed away at each step. Elution with elution buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 0.05 % DDM, 0.01 % CHS, 10 % glycerol, 300 mM imidazole) was carried and eluted protein was collected as 1 mL fractions. The presence of CXCR4-mut in fractions was verified by Coomassie-stained SDS-PAGE. Sample was concentrated as described above and SEC was carried using degassed sterile DDM-CHS buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 0.05 % DDM, 0.01 % CHS, 10 % glycerol). Elutions of 500 µL fractions were collected and presence of CXCR4-mut was verified by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

### 3.2.4. Soluble Expression Reactions

#### *Preliminary SDF1-α reaction*

The *NHis-SDF1* construct was the first plasmid of SDF1-α which was worked with and therefore, was used as a reference, comparison and model for further work with the remaining SDF1-α constructs. Based on previous reaction conditions for expression of either BAK or VDAC, in preliminary batches of 50 µL of CFPS reactions (30 % lysate, 8 mM Mg/Glu and 300 mM K/Glu, 30$^{o}$C), NHis-SDF1 was expressed using either Nico21 or SHuffle lysate. Samples of total reaction and soluble portion (30 000 g centrifuged supernatant) were analysed by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

#### *Lysate, $Mg^{2+}$, $K^{+}$ and temperature screen*

CFPS optimisation was carried out in order to determine the working conditions of lysate, $Mg^{2+}$, $K^{+}$ and incubation temperature. In a 96-well plate, 50 µl CPFS was carried out for the expression of NHis-SDF1, with the different lysates. Additionally, in order determine the optimum working concentration of $Mg^{2+}$, a screen of Mg/Glu of concentration 3.2 – 9.6 mM was carried out. Samples of total reaction and soluble portion (30 000 g centrifuged supernatant) were analysed by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

Following confirmation for expression of NHis-SDF1, the reaction was setup in Eppendorf tubes using selected SHuffle lysate, of which expression was most favourable in producing soluble protein. In sterile 1.5 mL Eppendorf tubes, batches of 50 µL of CFPS reaction (30 % SHuffle lysate, 300 K/Glu, 30 $^{o}$C) were carried out in order to confirm optimum Mg/Glu concentration (3.2 – 11.2 mM) for the expression of NHis-SDF1. The concentration range of Mg/Glu screen was extended here in order to confirm its influence on NHis-SDF1 expression since previous 3.2 – 9.6 mM Mg/Glu range indicated increase in expression of NHis-SDF1 with increasing concentration of Mg/Glu. Samples of total reaction and soluble portion (30 000 g centrifuged supernatant) were analysed by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

Following identification of reaction conditions for NHis-SDF1 (30 % lysate, 8.0 mM Mg/Glu, 300 mM K/Glu), 50 µL of CFPS reactions used for SDF1-His, SDF1-LH and CHis-SDF expression was carried out to confirm expression and screen for Mg/Glu concentration (3.2 – 11.2 mM). Two lots of reactions were set up and incubated at 20 $^{o}$C and 30 $^{o}$C for 16 h in a Thermomix at 500 rpm. Samples of total reaction were analysed by SDS-PAGE, anti-histidine western blot and Coomassie blue gel staining.

### *Deuterated SDF1-α expression*

SANS experiments require deuterated protein. With the success in the production of unlabelled SDF1-α, it was considered as model for optimising the CFPS expression of deuterated proteins. Instead of unlabelled amino acids (uaa), deuterated amino acids ("CELL FREE" AMINO ACID MIX (20 AA) (U-D, 98 %), Cambridge Isotope Laboratories, Inc) were used instead. A stock solution of deuterated amino acids (daa) was made 100 mg/mL using DEPC-water. A screen for the optimum amounts of deuterated amino acids (1 – 30 mg/mL) to be used was carried out. Using a final total concentration of 5 mg/mL of amino acids, mixes of daa:uaa (95:5, 90:10, 85:15, 80:20) were incorporated into the CFPS reaction mix. The expression of labelled and deuterated NHis-SDF1 SDF1-α was confirmed by western blot analysis.

### *Purification using magnet beads*

Small-scale purification of both labelled and unlabelled NHis-SDF1 was performed using the MagneHis$^{TM}$ Protein purification system from Promega. An equal volume of binding buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole at pH 7.5) was added to 500 µL of soluble fraction of SDF1 containing reaction. 20 µl of beads were added to the reaction, pipetted to mix and incubated at room temperature for 2 min. The supernatant was extracted and the beads were washed three times with 1 mL of wash buffer (50 mM HEPES, 500 mM NaCl, 25 mM imidazole at pH 7.5). NHis-SDF1 was then eluted from the beads following 2 minutes incubation in 200 µL of elution buffer (50 mM HEPES, 500 mM NaCl, 500 mM imidazole at pH 7.5). The purification of SDF1 was confirmed by Coomassie-stained SDS-PAGE. Using 3.5 kDa molecular weight cut-off Gebaflex-Midi dialysis tubes (Generon), the eluted samples were then dialysed extensively against dialysis buffer (50 mM HEPES, 500 mM NaCl at pH 7.5) at 4 $^{o}$C.

### *Thrombin cleavage*

Both labelled and unlabelled NHis-SDF1 were dialysed extensively against 50 mM HEPES pH 7.5 at 4 $^{o}$C. The dialysed SDF1 solution was incubated overnight with thrombin at 37 $^{o}$C for complete cleavage at a mass ratio of 5:1. The sample was then loaded on a benzamidine Sepharose 4 FF column (GE Healthcare) to remove the thrombin, and then through a HisTrap FF prepacked Ni Sepharose 1 mL columns (GE Healthcare) to remove the cleaved histidine tag. The presence of cleaved N-His-SDF1 was confirmed by were analysed by SDS-PAGE and Coomassie blue gel staining.

### *Large-scale expression and purification*

CFPS reaction for soluble NHis-SDF1 was scaled up to 10 mL of reaction. Reactions were carried out as 10 mL aliquots in 50 mL Falcon tubes at 30 $^{o}$C for 16 hr with agitation at 150 rpm. Purification was carried out using HisTrap FF prepacked Ni Sepharose 1 mL columns (GE Healthcare) on an ÄKTAprime plus protein purification system (GE Healthcare). The 10 mL of reaction supernatant was diluted to 50 mL with binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole) and filtered (0.45 µm). Binding to the column was carried out at a flow-rate of 0.5 mL/min. Subsequent washes with binding buffer and wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole) was carried out until no further protein was being washed from the column. Gradient elution with elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole) was conducted and eluted protein was collected as 1 mL fractions. The presence of NHis-SDF1 in fractions was confirmed by SDS-PAGE and Coomassie blue gel staining.

The fractions containing NHis-SDF1 were pooled and exchanged into SEC flow buffer (50 mM HEPES pH 7.5, 500 mM NaCl) using desalting column (BioRad 10DG desalting column). The sample was then concentrated to 1 mL using 3.5 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore). For SEC, The concentrated sample was loaded onto pre-equilibrated with Superdex$^{TM}$ 75 10/300 GL column (GE HealthCare) using 50 mM HEPES pH 7.5, 500 mM NaCl flow buffer. The presence of NHis-SDF1 in 1 mL collected fractions was confirmed by Coomassie-stained SDS-PAGE. The purified NHis-SDF1 was concentrated to 10 mg/mL in an approximate volume of 100 µL, and sent for high-throughput crystallisation trials (HTXlab, PSB/EMBL, Grenoble) – see chapter five.

CFPS reactions for soluble NHis-SDF1 and CHis-SDF1 was scaled up to 50 mL of reaction. Reactions were filtered (0.45 µm) and carried out in sterile, capped 250 mL flask at 30 $^{o}$C for 16 hr with low agitation (50 rpm). The completed reactions were diluted 5 X with binding buffer and purification using HisTrap FF prepacked Ni Sepharose 5 mL columns was carried out. Eluted protein fractions was visualised on Coomassie blue stained SDS-PAGE gel. The fractions containing NHis-SDF1 and CHis-SDF1 were pooled and exchanged into SEC flow buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM β-mercaptoethanol, 3 % glycerol) using desalting column (BioRad 10DG desalting column). The sample was then concentrated to 1 mL using a 3.5 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore). For SEC, the concentrated sample was loaded onto pre-equilibrated Superdex$^{TM}$ 75 10/300 GL column (GE HealthCare). The presence of soluble NHis-SDF1 and CHis-SDF1 in 1 mL collected fractions was confirmed by Coomassie-stained SDS-PAGE. The purified NHis-SDF1 and CHis-SDF1 was concentrated to 5 mg/mL in an approximate volume of 100 µL, and sent for high-throughput crystallisation trials (HTXlab, PSB/EMBL, Grenoble) – see chapter five.

### *2dCD4-wt expression*

In batches of 50 µL CFPS reactions (30 % lysate, 8 mM Mg/Glu, 300 mM K/Glu) in sterile 1.5 mL Eppendorf tubes, 2dCD4-wt-NHis and 2dCD4-wt-CHis variants were expressed using SHuffle lysate at 20 $^{o}$C and 30 $^{o}$C in a Thermomix at 500 rpm. Samples of total reaction and soluble portion (30 000 g centrifuged supernatant) were used for SDS-PAGE gel and anti-histidine western blot analysis. Once expression was confirmed, 50 µL CFPS reactions using *2dCD4-wt-Chis* with Mg/Glu screen (4.8 – 9.6 mM) were then carried out in order to optimise Mg/Glu working conditions. Samples of total reaction and soluble portion (30 000 g centrifuged supernatant) were used for SDS-PAGE gel and anti-histidine western blot analysis.

### *Large-scale expression and purification of deuterated 2dCD4-wt*

Using 2 mg/mL of deuterated amino acid, CFPS reaction (30 % SHuffle lysate, 4.8 mM Mg/Glu, 300 mM K/Glu, 20 $^{o}$C) for soluble deuterated 2dCD4-wt-CHis was scaled up to 100 mL of reaction. Reactions were filtered (0.45 µm) and carried out as 50 mL aliquots in 50 mL Falcon tubes at 20 $^{o}$C for 16 hr with low

agitation at 50 rpm. Purification was carried out using HisTrap FF prepacked Ni Sepharose 5 mL columns (GE Healthcare). The reaction supernatant was diluted to 500 mL with binding buffer (50 mM Tris.Cl pH 7.5, 300 mM NaCl, 5 mM imidazole) and filtered (0.45 µm). Binding to the column was carried out at a flow-rate of 0.5 mL/min. Subsequent washes with binding buffer, wash buffer 1 (50 mM Tris.Cl pH 7.5, 1 M NaCl, 20 mM imidazole) and wash buffer 2 (50 mM Tris.Cl pH 7.5, 300 mM NaCl, 5 % sucrose, 40 mM imidazole) was carried out until no further protein was being washed from the column. Gradient elution (0 % - 100 %) with elution buffer buffer (50 mM Tris.Cl pH 7.5, 300 mM NaCl, 300 mM imidazole) was performed and eluted protein was collected as 1 mL fractions. The presence of 2dCD4-wt-CHis in fractions was confirmed by SDS-PAGE and Coomassie blue gel staining..

The most pure fractions were pooled and concentrated to 3 mL using 10 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore). The samples were exchanged into SEC flow buffer (50 mM Tris-Cl pH 7.5, 300 mM NaCl, 5 % sucrose) using desalting column (BioRad 10DG desalting column). The most pure samples were separated from less pure samples and kept on ice.

The less pure 2dCD4-wt-CHis samples were subject to an additional round of purification. These were first concentrated to 3 mL using 10 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore). The samples were exchanged into SEC flow buffer (50 mM Tris-Cl pH 7.5, 300 mM NaCl, 5 % sucrose) using desalting column (BioRad 10DG desalting column). The sample was then further concentrated to 1 mL before being loaded onto a Bio-Rad NGC™ medium pressure chromatography system for SEC using an equilibrated Superdex™ 200 10/300 GL column (GE HealthCare). This separate SEC run was carried out in order to limit contaminants and purify 2dCD4-wt-CHis. The fractions of SEC were collected and the presence of 2dCD4-wt-CHis in fractions was confirmed by Coomassie-stained SDS-PAGE analysis. Pure fractions of 2dCD4-wt-CHis were pooled with previous pure fractions of 2dCD4-wt-CHis. Samples were stored on ice at 4 $^{o}$C until further use.

### 3.2.5. Forming the GP120 – 2dCD4 Complex

#### *GP120 expression and purification*

A mammalian cell system (FreeStyle 293-F) was used for the expression of GP120. These stable cell-lines, provided by the HIV Research and Pathogenesis Laboratory (University of Witwatersrand, South Africa), were created following transfection of GP120-plasmid vector (pCI-Neo vector), and an intense screening and selection process.

Vials of the cells, stored at -80 $^{o}$C, were thawed and added to 10 ml of FreeStyle 293 Expression media (Life Technologies) supplemented with 500 µg/ml of G418 (potency 700 – 730 µg/mg) and seeded in a small (25 cm$^2$) culture flask (Easy Nunc, Nunclon D). The flask(s) were incubated at 5 % $CO_2$/37 $^{o}$C until reaching 80-90 % confluence. Media changes and collection were made every 3-5 days. The semi-adherent cells, having a tendency to attach to the bottom of the flask and form a monolayer, were flushed off by gentle pipetting. The cells were transferred from the flask to 50 mL tubes and then collected by means of centrifugation (800 g, room temperature, 2-10 min). The supernatant was discarded. The cells were then transferred to larger flasks (75 cm$^2$, Easy Nunc, Nunclon D) for further expansion – three such flasks are required for transfer into roller bottles (about 15-20 million cells). When confluence was reached, cells from three flasks were pooled together into 100 mL of media (supplemented with 500 µg/mL G418) for transfer into a roller bottle (Corning 1750 cm$^2$). Rotation was maintained at low rate (0.3-0.5 rpm) for the first 24 h, and then increased to 0.8 - 1.0 rpm. Cells were allowed to settle in roller bottles for about 5 days. Once cell confluency had been reached the media was changed daily. Media was collected and replaced, with G418 at 250 µg/mL. Collected supernatants were centrifuged, filtered (0.8 µm) and stored at -20 $^{o}$C.

GP120 purification was performed at 4 $^{o}$C in a cold room. In batches of 3 L, the GP120 supernatant was allowed to thaw overnight. 2 mL of lectin resin (lectin from *Galanthus nivalis*-agarose, Sigma) were was pre-equilibrated extensively with 1 X PBS. The GP120 supernatant was passed through the column and flow through was collected. The resin was then washed sequentially with 200 mL of 500 mM NaCl in 1 X PBS, and 200 mL of 1 X PBS. GP120 was eluted with 12.5 mL of 1 M methyl α-D-mannopyranoside (MMP) in 1 X PBS. 1 mL fractions of eluate were collected. GP120 elution was repeated and a second batch of elution fractions was collected. The presence of purified GP120 in fractions was confirmed by Coomassie blue-stained SDS-PAGE. Purified fractions were pooled GP120 was extensively dialysed

against SEC buffer (50 mM Tris-Cl pH 7.5, 300 mM NaCl, 5 % sucrose). Samples were stored on ice at 4 $^{o}$C until further use.

A sample of purified GP120 was concentrated using 100 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore) to 1 mL and loaded onto a Bio-Rad NGC™ medium pressure chromatography system for analytical SEC using an equilibrated Superdex™ 200 10/300 GL column (GE HealthCare).

### *Purification of the GP120 – 2dCD4 complex by SEC*

Dilute, freshly purified GP120 and 2dCD4-wt were combined, with 2dCD4-wt being in excess, at least twice the molar concentration of GP120. The GP120-CD4 mix was incubated on ice for 30 minutes and was concentrated using 100 kDa MWCO Amicon® Ultra-4 centrifugal filter units (Merck Millipore) to between 1 – 3 mg/mL of complex. Volumes of 900 µL of complex were loaded onto a Bio-Rad NGC™ medium pressure chromatography system for SEC using an equilibrated Superdex™ 200 10/300 GL column (GE HealthCare). The presence of complex in 500 µL elutions was verified were analysed by SDS-PAGE and Coomassie blue gel staining.

## 3.3. RESULTS

### 3.3.1. Liposome and proteoliposome-based CFPS

#### *Proteoliposome CXCR4 expression in various lysates*

CXCR4-wt and CXCR4-lys were expressed in the form of proteoliposomes in the various lysates and expression was detected using anti-histidine western blot analysis (Figure 3.7). The expression of the proteins was heavily dependent upon the concentration of $Mg^{2+}$ as low amounts impeded expression. Compared to Nico21, the reference strain for lysate preparation, expression of both full-length CXCR4-wt and CXCR4-lys, at respectively 39 kDa and 52 kDa, was particularly successful in the SHuffle and Rosetta lysate. It should be noted that due to the intrinsic hydrophobic properties of membrane proteins and the high concentration of lipids and DNA present in the samples, protein smearing during gel migration is evident on the stained gels. In contrast, expression in C43 and Rosetta Gami based-lysates was very limited, despite having the ability to express BAK protein (chapter two). From previous experience, the use of Nico21 as lysate typically results in approximately 50 – 70 µg of CXCR4-wt expression per mL of CFPS reaction. SHuffle and Rosetta showed overall increases in expression for both variants of CXCR4 of approximately twice the amount compared to reference Nico21 expression (100 – 150 µg/mL of reaction). Whilst the occurrence of monomeric CXCR4 on the western blot was nearly the same as reference Nico21 expression, increased expression was notably observable due to the increased amounts of dimeric and truncated species present. The presence of dimeric CXCR4, particularly in the case of CXCR4-wt, at 75 kDa was observed. The highest levels of CXCR4-wt dimer were observed with Rosetta lysate, further reinforcing Rosetta as the lysate with the strongest expression. Around twice the amount of dimer was produced in the Rosetta lysate compared to the reference Nico21 lysate. The occurrence of the GP120 dimer could potentially be indicative of correct folding of the product protein. Rosetta lysate derived from H-Enfors (HL) was as efficient as the Rosetta lysate produced using standard methods.

However, truncated protein, particularly between 15 – 25 kDa, was observed in all cases where expression happened. Truncated protein was particularly abundant when using SHuffle lysate when either CXCR4-wt or CXCR4-lys was expressed. Much less truncated protein was observed with the use of Rosetta lysate. Whilst the truncation profile of CXCR4-wt in Rosetta and RH was similar, that of CXCR4-

lys in RH was not. With two pronounced species at approximately 15 and 20 kDa, the production of truncated CXCR4-lys when using RH was almost abundant as in SHuffle lysate. The occurrence of truncated protein is therefore due to a combined effect of the lysate used and of the protein construct being expressed.

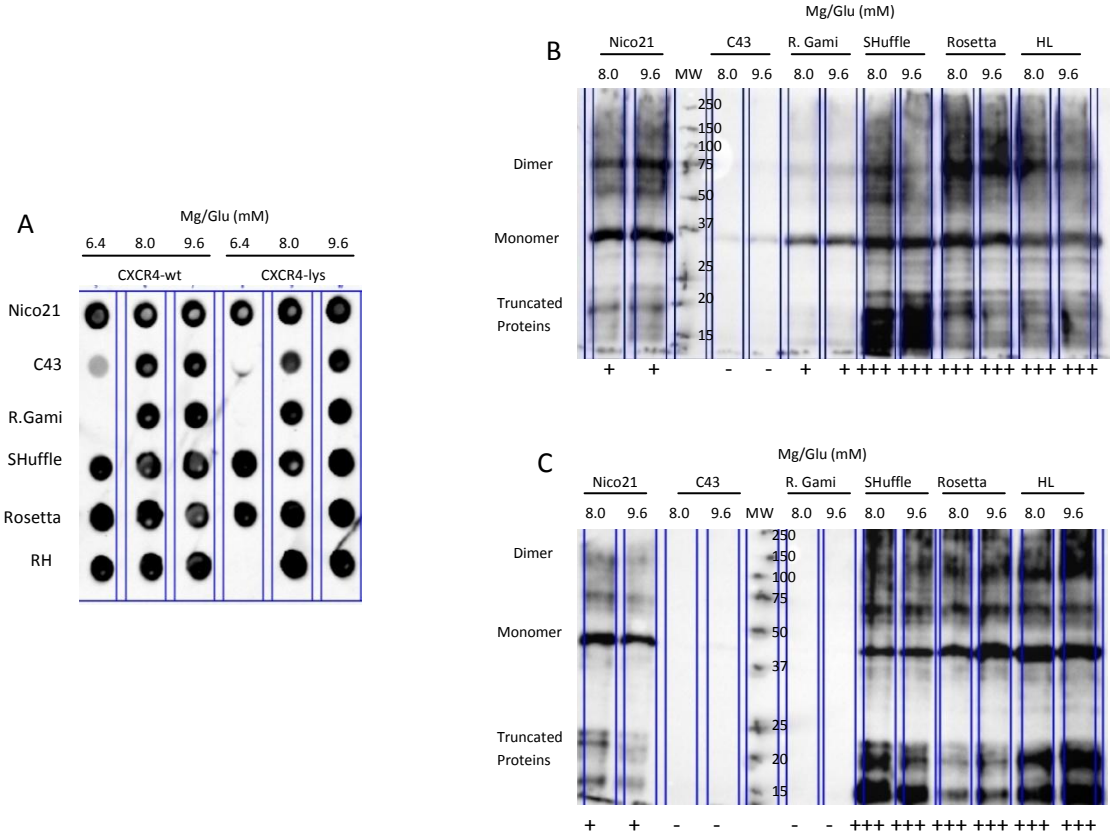



***Figure 3.7:*** *Anti-histidine dot blot and western blot analysis of CFPS reactions using non-standard lysates to express histidine-tagged CXCR4-wt and CXCR4-wt protein in liposome, resulting as monomers, oligomers and truncated proteins. Potassium glutamate (300 mM) and lysate (30 %) amounts are constant; magnesium glutamate was varied (6.4 – 9.6 mM) to validate and identify optimum conditions for CFPS (A). From the dot blot results, selected samples were used for western blot in order to confirm correct expression CXCR4-wt and CXCR4-wt, and provide rough estimates of the yield of proteins.+ signs indicate relative amounts of CXCR4 generated; - signs indicate poor to no expression. HL: Hydrogenated lysate (Rosetta lysate derived from HCDC in H-Enfors). MW: Molecular weight.*

### *Time-course assay for proteoliposome CXCR4 expression*

The time-course assay was designed to identify the most appropriate length of time for the CFPS reactions with regards to greatest expression of soluble protein whilst minimising expression of truncated protein species. The highest level of CXCR4-wt expression was found to be after 16 h incubation at 30 $^{o}$C (Figure 3.8; it should be noted that total sample of 16 h incubation was diluted by 5 X with buffer). Expression of CXCR4-wt begins as soon as all of the reagents are combined and incubated at 30 $^{o}$C, with detection of full-length protein within the first 30 minutes. Maintaining the reaction longer than 16 h did not increase the yields of the proteins (not shown here) – this is likely to be due to depletion of resources and accumulation of by-products detrimental to CFPS. The amount of truncated species is dependent on the amount of full-length protein being expressed: higher amounts of full-length CXCR4-wt results in higher amounts of observed truncated proteins. Truncations were particularly observed at about 25 kDa and lower throughout the reaction course. Another prominent truncated protein at approximately 30 kDa was observed as from 4 h. These truncated forms remain present even after 16 h, and are therefore unlikely to be due to incomplete cell-free reaction processing. Since the expression of CXCR4 is already relatively low for the purposes of structural biology work, achieving less truncation by compromising the yields of full-length expression of the protein was not considered as advantageous. Without further analysis, for instance using mass-spectrometry, it can only be speculated that these fragments could correspond to rare codon sites or internal Shine-Delgarno resembling sequences.

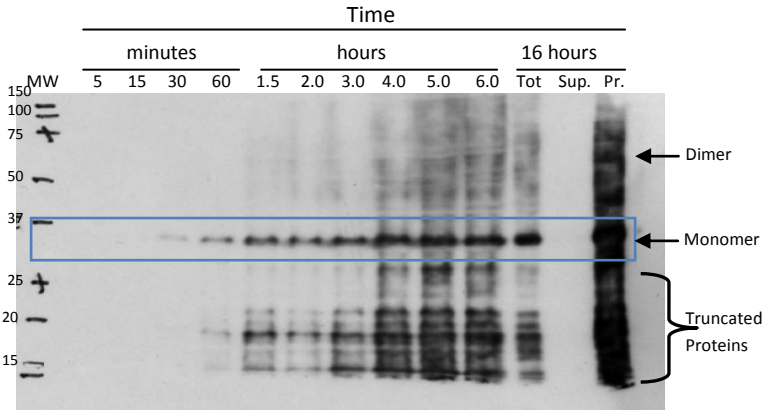



***Figure 3.8:*** *Anti-histidine western blot analysis of the time-course assay for the expression of full-length CXCR4-wt (indicated inside horizontal blue box) in proteoliposomes. Samples of 10 µL were taken from 200 µL CFPS reaction between 0 – 16 h of reaction for analysis by anti-histidine* western *blot. CXCR4-wt is detected as from 30 minutes of reaction, and increases to maximum amounts after overnight incubation. Total (Tot) and supernatant (Sup) from 16 h samples were diluted are 5 X (due to viscosity), and proteoliposomes (Pr) was resuspended in 200 µL of buffer. MW: Molecular weight.*

### *Temperature and solubility assay of proteoliposome CXCR4*

At 30 $^{o}$C, the standard temperature for CFPS, higher full-length CXCR4-wt expression was achieved after 16 h incubation as compared to incubation at 20 $^{o}$C (Figure 3.9). However, higher amounts of truncated protein and aggregates were also produced. As far as the indications provided by the western blot profile, the inclusion of liposome or the use of previously centrifuged lysate did not reduce the amounts of truncated protein and aggregates. However, the inclusion of liposome into the CFPS reaction certainly provides the environment for the production of folded and functional protein (Soranzo *et al.*, 2016, Soranzo *et al.*, 2017, Liguori *et al.*, 2008, Liguori *et al.*, 2015). The solubility of the protein was, however, improved when the CFPS reaction was carried out at 20 $^{o}$C, and with the use of centrifuged lysate and in the absence of liposomes. In this case it is suspected that the presence of residual bacterial lipids in the lysate could be acting as solubilising agents. Whilst centrifuging the lysate initially readily clears the lysate of larger insoluble particles, these lighter lipid particles remain. CFPS at lower temperature and in the absence of liposome, into which the lipid particles could have integrated, favours the maintenance

of small amounts CXCR4-wt in soluble form. It is interesting to note that soluble cytokine receptors reported in blood serum are suggested to play roles in the inflammatory and immune response (Levine, S.J *et al.* 2004). These were suggested to be formed by several mechanisms, which include proteolytic cleavage of receptor ectodomains, alternative splicing of mRNA transcripts, transcription of distinct genes that encode soluble cytokine-binding proteins and, the release of full-length receptors within the context of exosome-like vesicles. In the case of the CFPS reaction, the influence of the solublised lipid particles could contribute to the occurrence of such exosome-like vesicles, which in turn, can explain the presence of the soluble forms of CXCR4.

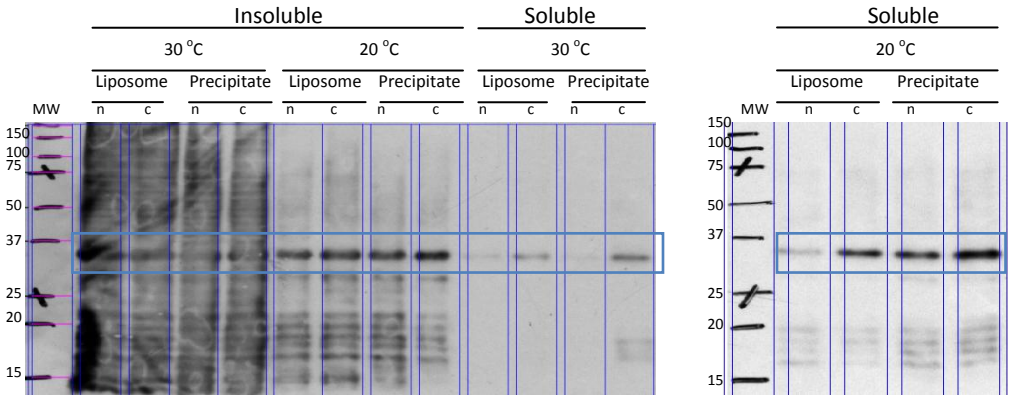



***Figure 3.9:*** *Anti-histidine western blot of 16 h CFPS expression of 50 µL of full-length CXCR4-wt (indicated inside horizontal blue box) in proteoliposomes or as precipitate, at 20 °C and 30 °C, using centrifuged (c) or non-centrifuged lysate (n). CFPS reaction was diluted 5 X in buffer and centrifuged to separate soluble and insoluble fractions. Yields of CXCR4-wt are higher when using 30°C compared to 20°C. Lower temperature prevents higher molecular weight oligomers and aggregates, and improves solubility of the protein. The use of centrifuged lysate improves solubility of CXCR4-wt. MW: Molecular weight.*

### *Proteoliposome washes and purification*

Incubation and washes with NaCl and ammonium sulphate were carried out to separate CXCR4-wt and protein impurities using the salting-out technique (Figure 3.10 and 3.11) (Duong-Ly and Gabelli, 2014). However, with the presence of buoyant proteoliposomes, a floating lipid layer could also be obtained with increased salt concentration, in addition to insoluble precipitate and soluble supernatant fractions. Indeed, at high concentration (5 M) NaCl and 50 % (w/v) ammonium sulphate, most of the CXCR4-wt and impurities were located in the lipid layer. At low salt concentration, CXCR4-wt in proteoliposomes, along with impurities were collected in the white, insoluble precipitate fraction. It should be noted that with the use of 1 M NaCl, the lipid fraction contained protein impurities, but no CXCR4-wt.

Proteoliposomes embedded with CXCR4-wt are likely to be denser than proteoliposomes without CXCR4-wt. Therefore, by initially treating CXCR4-wt proteoliposomes with 1 M NaCl, impurities adhering in the lipid layer can be removed. Subsequently, treating the CXCR4-wt proteoliposomes with 5 M NaCl could be used to separate CXCR4-wt proteoliposomes from the proteoliposomes without any CXCR4-wt. Salt concentrations using 3 M NaCl) assisted in purification since CXCR4-wt in proteoliposomes were localised in the lipid layer and a portion of impurities remained either in soluble and insoluble fraction. The use of 40 % (w/v) ammonium sulphate produced much more clear-cut results as all of CXCR4-wt was localised in lipid layer.

The hydrophobic nature of the liposome provides an adequate environment for the localisation of well-folded CXCR4-wt, but also for contaminating proteins. Incubation of the proteoliposomes with sodium carbonate or urea can serve to remove peripheral proteins which adhere to the surface of the proteoliposomes, and provides an indication on the integration of *de-novo* synthesised membrane protein in the liposome bilayer since the treatment does not destroy the liposome (Long *et al.*, 2012, Fujiki *et al.*, 1982, Sakai *et al.*, 2009). Sequential incubation and washes with 1 M sodium carbonate allowed the elimination of some contaminants, while preserving the majority of CXCR4 in the lipid fraction (Figure 3.12).

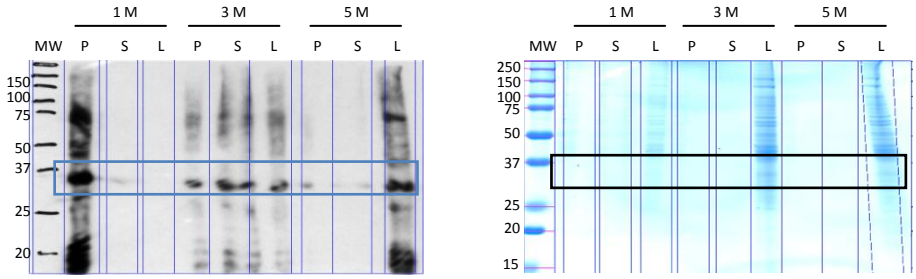



***Figure 3.10:*** *Anti-histidine western blot (left) and Coomassie-blue stained gel (right) analysis of CXCR4-wt (boxed) proteoliposome incubation with NaCl (1, 3 and 5 M). MW: Molecular weight; P: Precipitate (insoluble fraction); S: Supernatant (soluble fraction); L: Liposome (lipid fraction).*

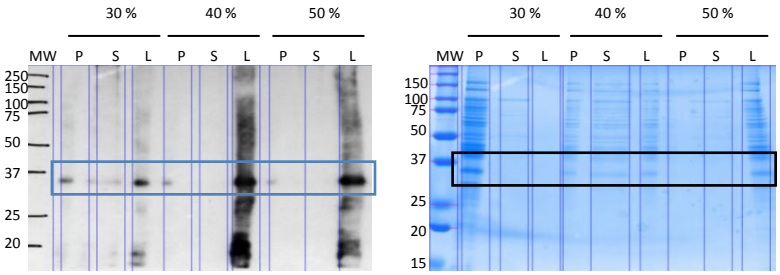



***Figure 3.11:*** *Anti-histidine western blot (left) and Coomassie-blue stained gel (right) analyses of CXCR4-wt (boxed) proteoliposomes incubation with ammonium sulphate (30, 40 and 50 % (w/v)). MW: Molecular weight; P: Precipitate (insoluble fraction); S: Supernatant (soluble fraction); L: Liposome (lipid fraction).*

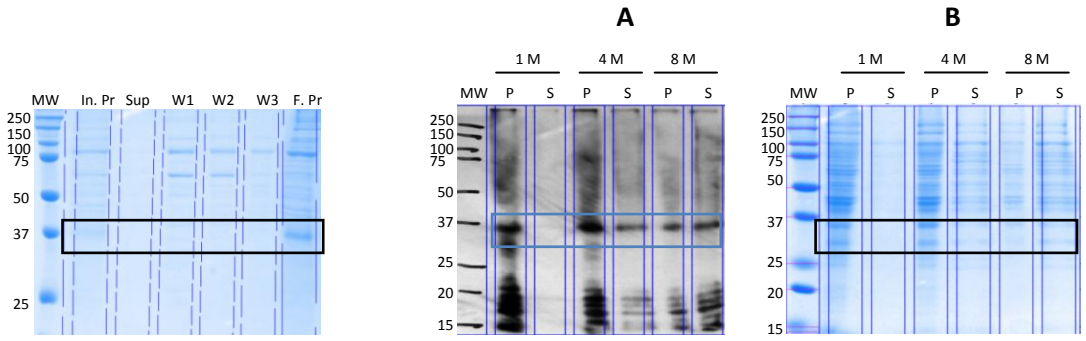



***Figure 3.12:*** *Left: Anti-histidine western blot of CXCR4-wt (boxed) proteoliposome incubation with urea (1, 4 and 8 M). MW: Molecular weight; P: Precipitate (insoluble fraction); S: Supernatant (soluble fraction). Right: Anti-histidine western blot (A) and Coomassie-blue stained gel (B) of sequential 1 M $Na_2CO_3$ incubation and washes. In. Pr: initial proteoliposome pellet, in 50 mM HEPES (dilute); Sup: supernatant of 50 mM HEPES wash; W1: Supernatant of first 1 M $Na_2CO_3$ incubation and wash; W2: Supernatant of second 1 M $Na_2CO_3$ incubation and wash; W3: Supernatant of third 1 M $Na_2CO_3$ incubation and wash; F. Pr: Final proteoliposome pellet, resuspended in 50 mM HEPES.*

The use of increasingly higher concentrations of urea, a very strong protein denaturant which, however, tends to preserve lipidic bilayer intergrity, resulted in dissociation of a large amount of undesirable protein into the soluble fraction (Figure 3.12). At 8 M urea concentration, almost an equal amount of CXCR4 was localised in the soluble fraction as in the insoluble-lipid fraction. Several explanations can be suggested. Firstly, the proteoliposome fraction is a subset of the insoluble fraction, which contains insoluble, non-lipid associated proteins. These proteins, including unfolded, non-integrated CXCR4, are solubilised by the strong denaturating effect of urea. Secondly, with its strong denaturant effect, urea can cause much more effective dissociation of peripheral proteins. This includes poorly integrated CXCR4 membrane protein which adheres to the liposome. Thirdly, the strong denaturating effect of urea is likely to have a destabilising impact on the proteoliposome and is able to interfere with hydrophobic and polar forces in the lipid bilayer (Patra *et al.*, 1998). This leads to smaller, water-soluble proteoliposomes, which nevertheless, still have the integrated membrane protein preserved. Fourthly, the strong osmotic effect during proteoliposome resuspension leads to rupturing. The ruptured proteoliposomes consequently reform spontaneously, giving smaller propteoliposomes, which are then soluble.

Sucrose-gradient and ultracentrifugation purification is the established purification method for proteoliposomes at Synthelis (Soranzo *et al.*, 2016, Liguori *et al.*, 2008, Liguori *et al.*, 2015). However, large amounts of impurities still remain in the final "purified" proteoliposomes fraction (Figure 3.13, lane 1), making it unsuitable as a final stage for proteins destined for structural biology work such as crystallography or SANS. In conjunction with sucrose-gradient and ultracentrifugation purification, pre-incubation and washes of the proteoliposomes with sodium carbonate or urea only slightly improved CXCR4 purity, as compared to 50 mM HEPES buffer wash (Figure 3.13, lane 2, 3, 4). As compared to the CXCR4 reaction which was loaded directly onto the sucrose gradient, with protein losses at each washing step, slightly less CXCR4 was recovered.

The use of pre-centrifuged lysate for cell-free reactions, in conjunction with washes, sucrose-gradient and ultracentrifugation purification, showed some success in attaining a higher purity of CXCR4 in proteoliposomes (Figure 30). Already with the use of pre-centrifuged lysate, it should be noted that a clearer western blot analysis with less smearing is obtained (Figure 3.14, lanes 1 and 2). This is likely to be due to further removal of residual insoluble particles, which impede protein migration during SDS-PAGE. As observed previously, small amounts CXCR4 are present in the soluble fraction with the use of centrifuged lysate. Subsequent proteoliposome washes with 50 mM HEPES, sucrose-gradient and ultracentrifugation therefore results in loss of proteins in the soluble fraction, and low recovery of CXCR4-associated proteoliposomes (Figure 3.14, lanes 11 and 12). However, directly loading cell-free reactions that made use of pre-centrifuged lysate onto the sucrose-gradient and ultracentrifugation purification led to high CXCR4-proteoliposome purity.

Whilst ensuring the recovery of proteoliposomes in the lipidic fraction, the sucrose-gradient and ultracentrifugation purification technique is nevertheless a tedious and lengthy process. For small-scale protein productions, low-speed centrifuging washes (Figure 3.14, lanes 5 and 6) can provide sufficient purity which nearly equals that of sucrose-gradient and ultracentrifugation purification. The use of pre-centrifuged lysate once again shows a clearer western blot profile.

NaCl/EDTA treatment is a commonly used technique for dissociating ribosome sub-units and is exploited in membrane purification protocols to additionally inhibit protease activity (Reid and Nicchitta, 2012, Connolly and Gilmore, 1986). As far as observations from western blots and Coomassie gels in Figure 3.14 are concerned, no significant difference in the final purity of proteoliposomes was achieved, as compared to controls with where no NaCl/EDTA treatment was carried (Figure 3.14, lanes 7, 8, 9 and 10).

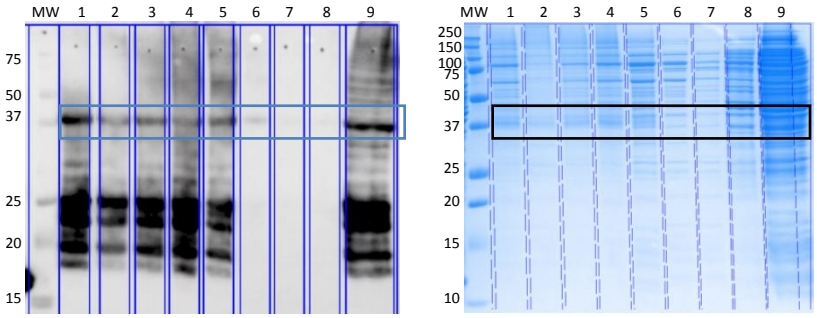



***Figure 3.13:*** *CXCR4 (boxed) in liposome, with lysate pre-centrifugation, liposome washes and ultracentrifugation.* ***Left:*** *Anti-histidine western blot;* ***Right:*** *Coomassie-blue stained gel* ***MW:*** *molecular weight;* ***1:*** *Non-centrifuged lysate, direct loading on sucrose-gradient ultracentrifuge;* ***2:*** *Non-centrifuged lysate, 8 M urea incubation, sucrose-gradient ultracentrifuge;* ***3:*** *Non-centrifuged lysate, 1 M $Na_2CO_3$ incubation, sucrose-gradient ultracentrifuge;* ***4:*** *Non-centrifuged lysate, 50 mM HEPES incubation, sucrose-gradient ultracentrifuge;* ***5:*** *8 M urea incubation, supernatant (soluble fraction, 5 X dilution);* ***6:*** *1 M $Na_2CO_3$ incubation, supernatant (soluble fraction, 5 X dilution);* ***7:*** *50 mM HEPES incubation, supernatant (soluble fraction, 5 X dilution);* ***8:*** *total reaction supernatant, non-centrifuged lysate;* ***9:*** *total reaction, non-centrifuged lysate*

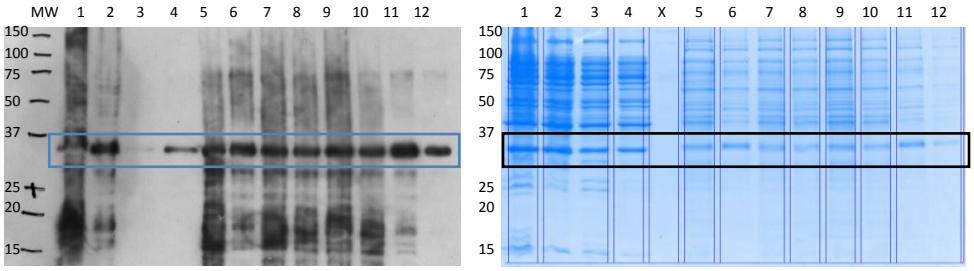



***Figure 3.14:*** *CXCR4 (boxed) in liposome, with lysate pre-centrifugation, liposome washes and ultracentrifugation.* ***Left:*** *Anti-histidine western blot;* ***Right:*** *Coomassie-blue stained gel.* ***MW:*** *molecular weight;* ***1:*** *Total reaction, non-centrifuged lysate;* ***2:*** *Total reaction, centrifuged lysate;* ***3:*** *Soluble supernatant of reaction, non-centrifuged lysate;* ***4:*** *Soluble supernatant of reaction, centrifuged lysate;* ***5:*** *Non-centrifuged lysate, 50 mM HEPES centrifuge wash;* ***6:*** *Centrifuged lysate, 50 mM HEPES centrifuge wash;* ***7:*** *Total reaction, non-centrifuged lysate, direct loading on sucrose-gradient ultracentrifuge;* ***8:*** *Non-centrifuged lysate, 50 mM HEPES incubation control, sucrose-gradient ultracentrifuge;* ***9:*** *Non-centrifuged lysate, 1 M NaCl incubation, sucrose-gradient ultracentrifuge;* ***10:*** *Non-centrifuged lysate, 1 M NaCl/3 mM EDTA incubation, sucrose-gradient ultracentrifuge;* ***11:*** *Centrifuged lysate, direct loading on sucrose-gradient ultracentrifuge;* ***12:*** *Centrifuged lysate, 50 mM HEPES wash, sucrose-gradient ultracentrifuge. X is an empty lane.*

### *Full-length CD4 expression as proteoliposomes*

Both full-length N-terminal and C-terminal histidine tagged CD4 could be successfully expressed using *E. coli* based CFPS in proteoliposomes (Figure 3.15). As observed with CXCR4, the presence of N-terminal histidine tagged CD4 truncated species is observed, but not in the case of C-terminal histidine tagged CD4. This observation allows the conclusion that the truncated species derived from CFPS are due to non-completed translation rather than proteolytic degradation. Furthermore, the expression of C-terminal histidine tagged CD4 is important as it confirms that the protein is fully translated.

Expression of either species was favoured at 20 $^{o}$C as compared to the standard 30 $^{o}$C. However, CFPS at lower temperatures (down to 18 $^{o}$C, data not shown) resulted in severe inhibition of protein expression. In varying the lysate percentage in the CFPS reaction, optimal expression was obtained at 30 % lysate. Thus, scaling the reaction to 1 mL, with subsequent sucrose-gradient and ultracentrifugation purification, enabled an estimated final recovery of 30 µg of CD4 per mL of CFPS reaction.

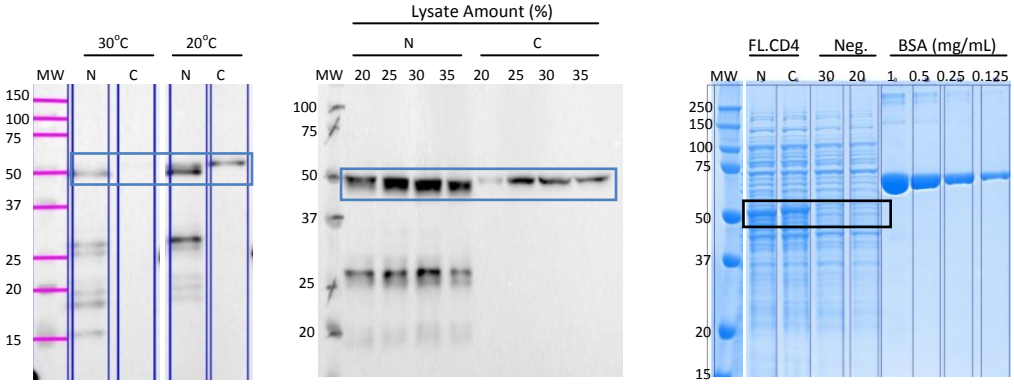



***Figure 3.15:*** *CPFS expression of full-length CD4-FL-NHis and CD4-FL-CHis in proteoliposomes (boxed).* ***Left:*** *Anti-histidine western blot of CD4, expressed at 20 $^{o}$C and 30$^{o}$C.* ***Middle:*** *Anti-histidine western blot of CD4 expressed using different lysate amounts.* ***Right:*** *Coomassie-blue stained gel of CD4, and CFPS expression control (no DNA included during reaction), both recuperated after sucrose-gradient and ultracentrifugation. MW: Molecular weight.*

### 3.3.2. Detergent-based CFPS

#### *Detergent expression of CXCR4 variants*

In a previous publication, crystal structure of CXCR4 was shown to be achieved with the lysozyme mutant of CXCR4, which was solubilised from insect cells, and then purified and crystallised using detergents (Wu *et al.*, 2010, Qin *et al.*, 2015). Therefore, with the goal of obtaining a crystal structure of CXCR4 that was produced using CFPS, it was imperative to produce solubilised and highly purified CFPS CXCR4. To do so, detergents were included in the CFPS reaction in order to establish conditions which enabled the high expression and efficient recovery of solubilised CXCR4.

#### *Preliminary detergent expression of CXCR4 variants in either SHuffle or Rosetta lysate*

CFPS of CXCR4-wt and CXCR4-lys in the presence of various detergents was firstly confirmed in either SHuffle or Rosetta lysate (Figures 3.16). For all of the detergents tested between 2 – 50 X of their corresponding critical micelle concentration (CMC), expression of CXCR4-wt and CXCR4-lys was possible. However, the use of digitonin seemed to slightly hinder expression when SHuffle lysate was used for CXCR4-lys, whereas the DDM-CHS mix severely hindered expression in both cases for either protein.

For the purpose of detergent-solubilised CFPS of CXCR4, Rosetta lysate was selected. Whilst no major difference in the expression of CXCR4 when using either SHuffle or Rosetta lysates, the use of SHuffle was prioritised for the expression of either SDF1-α (section 3.4.3) or 2dCD4 (3.4.4)

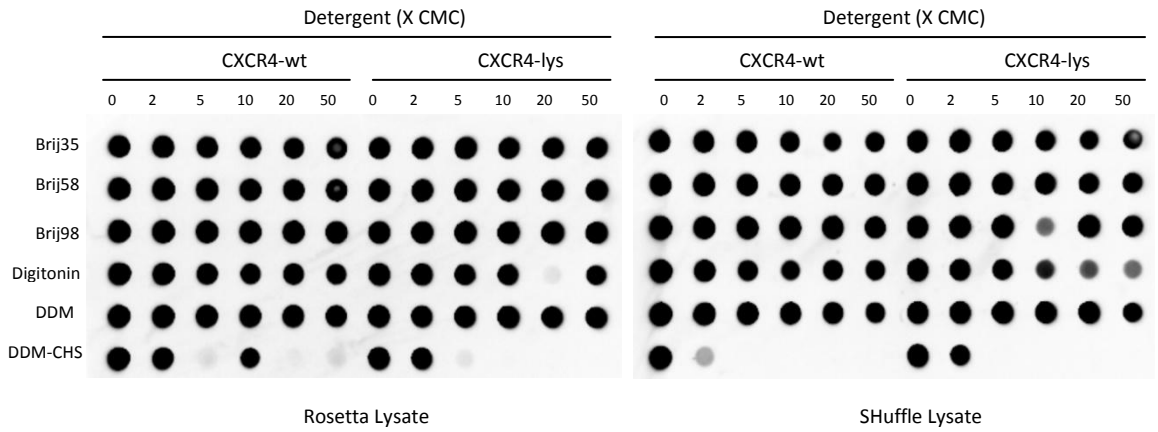



***Figure 3.16:*** *Anti-histidine dot blot for the detection of the expression of CXCR4-wt and CXCR4-lys in either Rosetta or SHuffle lysate and in the presence of various detergents (2 – 50 X CMC). All of the individual detergents are permissive for expression. DM-CHS mix however severely inhibits the expression of CXCR4.*

### *Confirming soluble detergent expression of CXCR4*

Whilst expression of CXCR4 in the presence detergents was readily achievable, obtaining soluble membrane protein was not straightforward. In an attempt to improve the solubility of CXCR4 and eliminate insoluble residual particles, pre-centrifuged Rosetta lysate was used for expression of CXCR4 in detergent at 3 X CMC (Figure 3.17). The use of pre-centrifuged lysate however did not show enhancement in the solubility of CXCR4, but did help to remove oligomers and aggregates of CXCR4. A relatively high solubility of CXCR4-wt was observed, followed by CXCR4-lys and poor solubility of CXCR4-h/s. Brij35 and Brij58 were found to be the most successful detergents in producing and maintaining soluble CXCR4, particularly for CXCR4-wt and CXCR4-lys. Brij58 was particularly efficient for generating high amounts and maintaining soluble CXCR4-wt. Although permitting expression of CXCR4, the use of DDM did not generate any solubilised CXCR4. Brij98 could only permit soluble CXCR4-wt and CXCR4-lys, but not CXCR4-h/s. Despite repeating the CFPS reaction and still obtaining low amounts of soluble CXCR4-h/s, detection by anti-strep western blot of procured important proof on the expression of full-length CXCR4, which can be seen as either monomer or dimer (Figure 3.18).

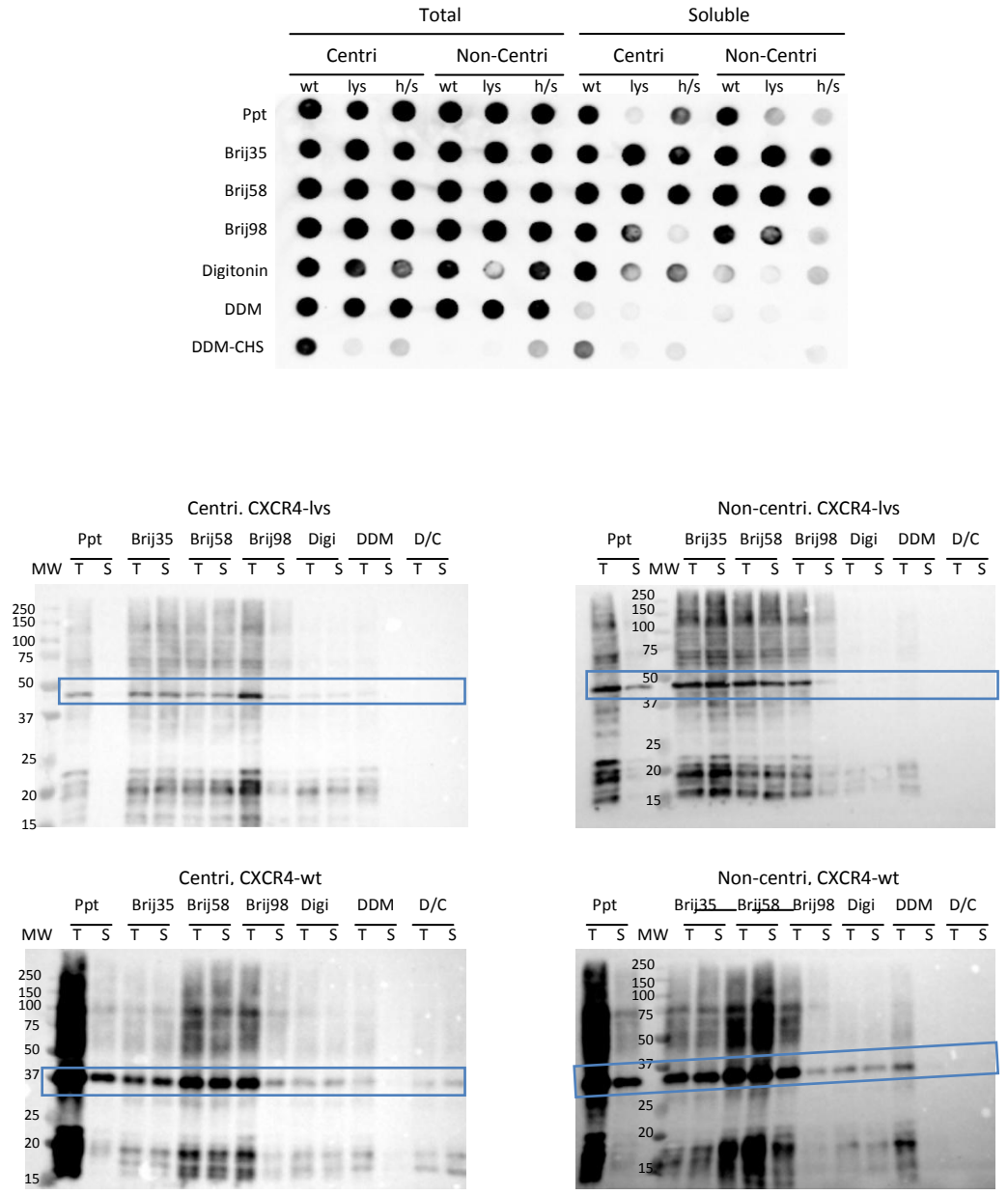



***Figure 3.17:*** *Anti-histidine dot blot and western blots for the detection of expressed CXCR4-wt, CXCR4-lys and CXCR4-h/s (boxed) using non-centrifuged or pre-centrifuged Rosetta lysate, and in the presence of various detergents (3 X CMC). Brij58 was especially successful for the high expression of soluble CXCR4-wt. The use of pre-centrifuged lysate did not improve yields or solubility of full-length CXCR4, but prevented the occurrence of oligomers and aggregates. T: Total reaction fraction; S: soluble reaction fraction. MW: Molecular weight.*

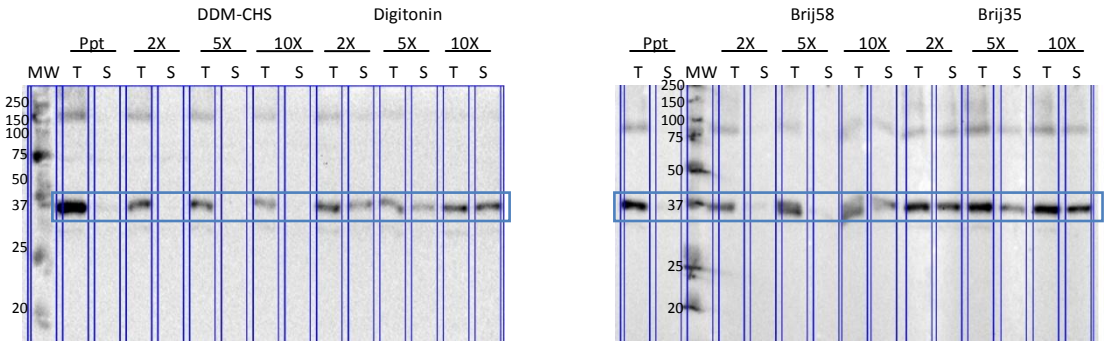



***Figure 3.18:*** *Anti-strep western blots for the detection of the expression of solubilised CXCR4-h/s in the presence of various detergents between 2 – 10 X CMC. Despite low levels of expression of soluble CXCR4-h/s, the results provide clear indication the occurrence of full-length, un-truncated CXCR4. The use of Brij35 showed greatest solubilising effect for CXCR4-h/s. MW: Molecular weight.*

### *Optimising soluble detergent expression of CXCR4-lys*

Aiming towards crystallisation of CXCR4, only CXCR4-lys was considered for further work since CXCR4-wt was reported to fail during crystallisation trials in a previous publication (Wu *et al.*, 2010). To further enhance CFPS of CXCR4-lys in detergents, the $Mg^{2+}$ concentration was optimised (Figure 3.19). A higher threshold concentration of Mg/Glu is required for initiating CXCR4-lys expression compared to CPFE without any detergent: 8.0 mM Mg/Glu was selected as optimum concentration for CXCR4-lys expression in Brij35 at 5 X CMC. In both cases however, excessive amounts of Mg/Glu compromised the expression of CXCR4-lys. Compared to the precipitate format, lower amounts CXCR4-lys are produced, indicating a detrimental effect of Brij35 (Figure 3.19).

To further confirm the expression of soluble CXCR4-lys using selected optimised 8.0 mM Mg/Glu condition, 50 µL batches CFPS reactions in the presence of various detergents (2 – 10 X CMC) were repeated in 1.5 mL Eppendorf tubes, with higher shaking rates during incubation (500 rpm in a Thermomix) (Figure 3.20). It is clear that compared to precipitate or proteoliposome format, the inclusion of detergents in the CFPS reaction lowers the expression of CXCR4-lys. Nonetheless, Brij35 at 5 X CMC, was found to be the most successful in producing high amounts of solubilised CXCR4-lys and which could occur even at low CMC. Whilst allowing the expression of solubilised CXCR4-lys, Brij58 however reduced the yield of soluble CXCR4-lys compared to the use of Brij35. Digitonin enabled soluble

expression of CXCR4-lys, but reduced its yields. DDM and DDM-CHS mix were found to be most detrimental for CFPS.

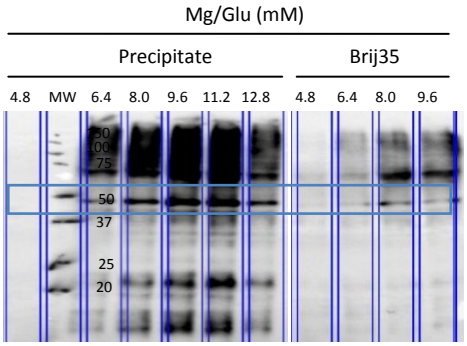



***Figure 3.19:*** *Anti-histidine western blot for the expression of full-length CXCR4-lys (boxed), as without or with Brij35, with range of Mg/Glu (4.8 – 12.8 mM). Expression of CXCR4-lys, without detergent, as precipitate requires less Mg/Glu as compared to when Brij 35 at 5 X CMC is present. Optimum Mg/Glu for the expression of CXCR4-lys as precipitate was determined to be between 9.6 – 11.2 mM, whereas for Brij35, 8.0 mM was considered. MW: Molecular weight.*

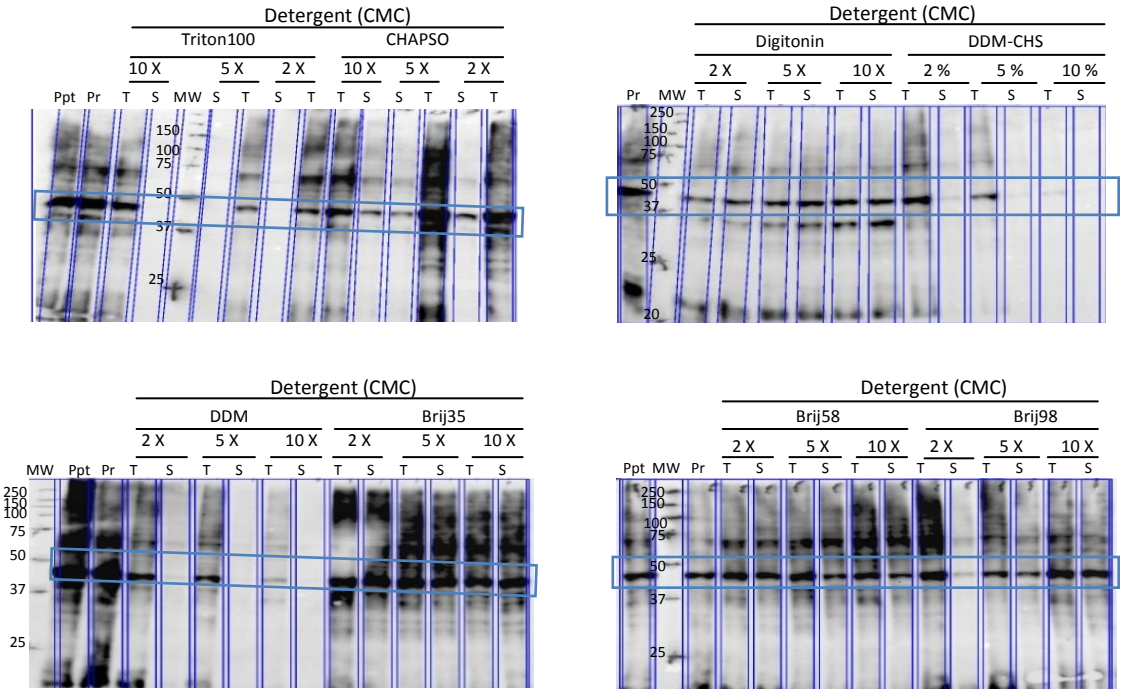



***Figure 3.20:*** *Anti-histidine western blot of soluble expression of CXCR4-lys (boxed) in detergents (2 – 10 X CMC), as 50 µL CFPS batch reactions. Mg/Glu in CFPS was maintained at 8.0 mM. Total (T) and soluble (S) fractions were analysed by anti-histidine* western *blot to confirm presence of CXCR4-lys. Brij35 at 5 X CMC was found to be condition which allowed for most expression of solubilised CXCR4-lys. MW: Molecular weight.*

### *Purification of detergent solubilised CXCR4-lys*

From the previous tests, Brij35 at 5 X CMC, which indicated to was selected for detergent solubilised expression of CXCR4-lys. The initial attempt to purify Brij35 solubilised CXCR4-lys from a small-scale (1 mL of CFPS reaction) using the MagneHis$^{TM}$ mangetic bead protein purification system was not successful (Figure 3.21). CXCR4-lys did not bind to the beads and remained the in the flow through, and only a trace amount was eluted. Overnight incubation with beads did not improve the binding.

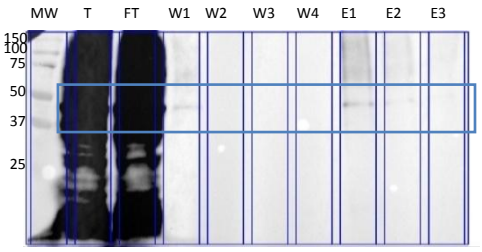



***Figure 3.21:*** *Anti-histidine western blot of the small-scale expression (1 mL) and purification of Brij35 (5 X CMC) solubilised full-length CXCR4-lys (boxed) using MagneHis$^{TM}$ Protein purification system. Molecular weight (MW), total (T), flow-through (FT), 1 mL washes (W1, W2, W3, W4) and 100 µL elutions (E1, E2, E3).*

Nevertheless, large-scale production and purification of CFPS detergent-solubilised CXCR4-lys was attempted. Using slow flow rates (0.5 mL/min) of the supernatant from large scale expression of Brij35 solubilised CXCR4-lys onto Ni-NTA affinity resin on an AKTA Prime system, however, permitted the capture of and eventual elution of CXCR4-lys (Figure 3.22A). The flow through and washes still contained unbound CXCR4-lys which indicates its relatively low affinity to the resin. From 50 mL of CFPS reaction an estimated total 100 □g of purified full-length CXCR4-lys could be obtained. The elution fractions also contained truncated N-terminal CXCR4-lys species.

Following Ni-NTA affinity purification, it was possible to retain and concentrate full-length and truncated CXCR4-lys in elution fractions using 100 kDa MWCO filter concentrating device (Figure 3.22B, input).

With Brij35 having micellar molecular weight of 49 kDa, this indicates the occurrence of solubilised CXCR4 in Brij35 detergent micelles, having total molecular weight above 100 kDa.

As described in a previous publication, DDM-CHS buffer was used for solubilisation, purification and crystallisation of the lysozyme mutant of CXCR4 (Wu *et al.*, 2010, Qin *et al.*, 2015). Therefore, the DDM-CHS buffer was selected for buffer exchange using SEC. Application of the concentrated CXCR4-lys through SEC, however led to heavy CXCR4-lys dilution. This resulted into approximately 50 % of protein loss (Figure 3.22B). Furthermore, SEC could not distinctly separate CXCR4-lys truncated and full-length species. This suggests the occurrence of a heterogeneous mix of CXCR4-Brij35 micelles of various sizes.

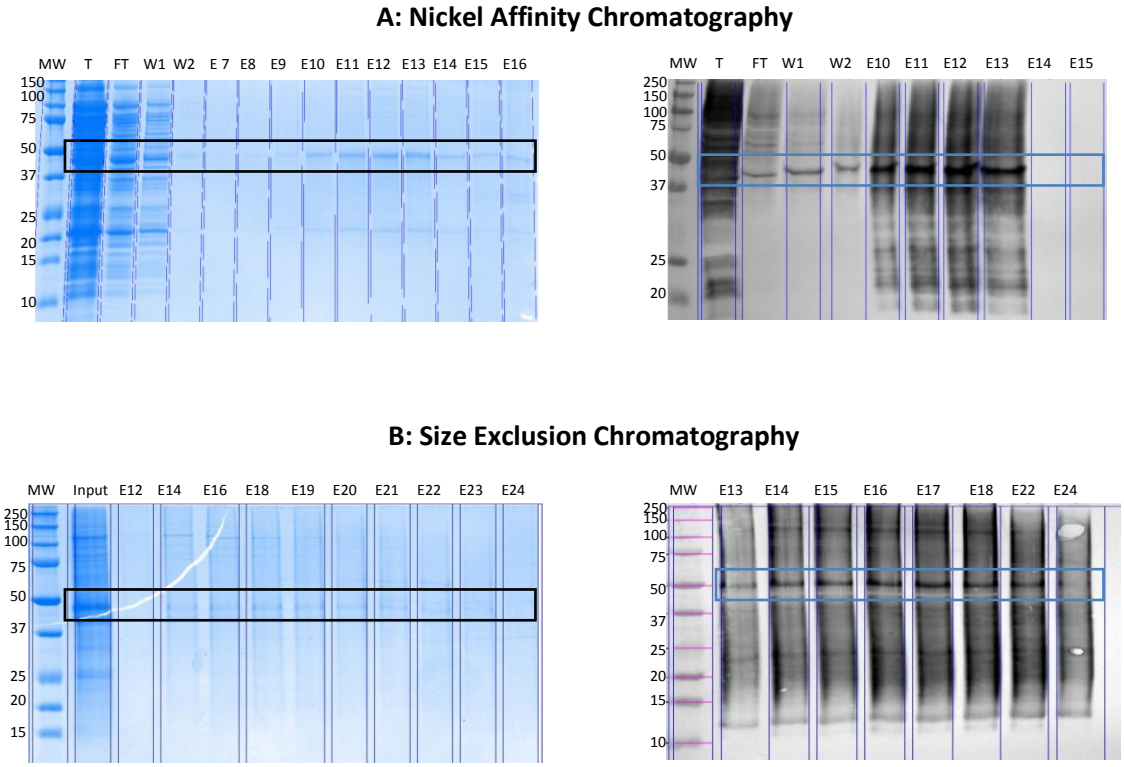



***Figure 3.22:*** *Coomassie-blue stained gels (left) and anti-histidine western blots (right) of the purification of 50 mL of batch CFPS reaction of Brij35 solubilised full-length CXCR4-lys (boxed) by Ni-NTA affinity (A), and subsequent size exclusion chromatography (SEC) using Superdex$^{TM}$ 200 10/300 GL column (B). Purified CXCR4-lys was eluted from Ni-NTA purification, with estimated total yields of 100 µg for full-length species. Concentrating the eluates and running on SEC showed poor separation and heavy dilution. Molecular weight (MW), total reaction (T), flow-through (FT), washes (W) and elutions (E).*

To avoid subsequent SEC and protein losses, an alternative method where Brij35 solubilised CXCR4-lys was exchanged directly into DDM-CHS buffer during Ni-NTA was attempted (Figure 3.23). Similar yields of 100 ug of purified full-length CXCR4-lys from 50 mL CFPS reaction were obtained. However, the presence of truncated CXCR4-lys and other non-specific micelle-associated contaminants still remained in the eluates. Increasing the amount of imidazole in the wash buffers in order to improve purity of the sample during purification led to loss of full-length CXCR4-lys, and this method was therefore not considered further.

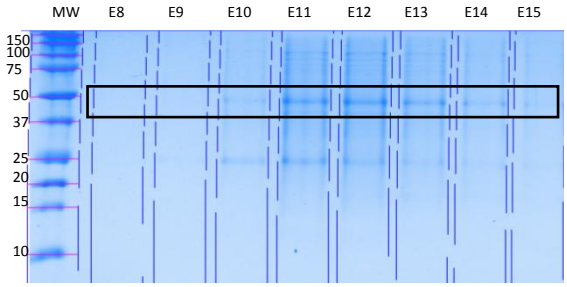



***Figure 3.23:*** *Coomassie-blue stained gel of the purification of 50 mL of batch CFPS reaction of Brij35 solubilised full-length CXCR4-lys (boxed) by Ni-NTA affinity and buffer exchange into DDM-CHS buffer. Purified full-length CXCR4-lys could be recovered with estimated total yields of 100 µg. Non-specific protein contaminants and truncated CXCR4-lys were also present in eluates. Molecular weight (MW), elutions (E).*

### 3.3.3. Soluble expression of SDF1-α

#### *Soluble expression of SDF1-α in various lysates*

CFPS of NHis-SDF1 in either BL21 C43 or SHuffle lysate was confirmed using the preliminary working conditions of 30 % lysate, 8.0 mM Mg/Glu, 300 mM K/Glu, 30 $^{o}$C (Figure 3.24A). However, the yields of NHis-SDF1 in both the total reaction and in the soluble fraction were much higher when using SHuffle lysate expression. With further investigation, it was shown that NHis-SDF1 could be expressed in all of the various tested lysates as from 6.4 mM of Mg/Glu (Figure 3.24B, C; Nico21, SHuffle, R. Gami, Rosetta lysates). The highest expression of soluble SDF1 was achieved using Shuffle (250 µg/mL of reaction),

which was visible at the correct molecular weight of 10 kDa on Coomassie-stained SDS-PAGE gel. Similar expression yields could be observed in Rosetta and Rosetta lysate derived from hydrogenated Enfors medium (RH). However, the fraction of soluble NHis-SDF1 in these was much less than SHuffle – less than 50 %.

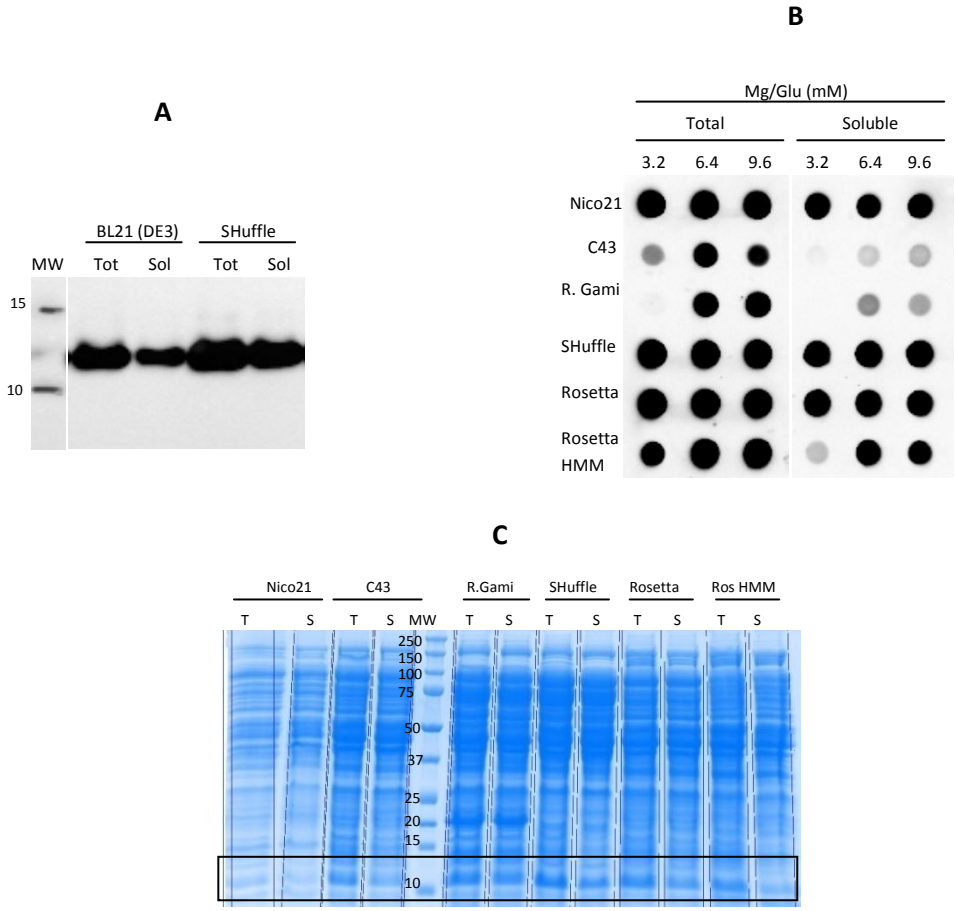



***Figure 3.24:*** *Preliminary soluble NHis-SDF1 expression in Nico21 and SHuffle lysate. Analyses by anti-histidine western blot (A), and anti-histidine dot blot (B) and Coomassie-blue stained gel (C) for NHis-SDF1 expression (boxed) in various lysates MW: Molecular weight; T: total reaction; S: soluble fraction. Higher expression and soluble NHis-SDF1 was favoured using SHuffle lysate. Potassium glutamate (300 mM) and lysate (30 %) amounts are constant; magnesium glutamate (3.2 – 9.6 mM) concentration was carried out to validate and identify optimum conditions for CFPS. From the dot blot, samples were used for SDS-PAGE and Coomassie-blue staining in order to confirm correct expression NHis-SDF1, and provide estimates of the yield of the protein.*

### *Optimising and confirming soluble expression of SDF1-α in SHuffle lysate*

The expression of NHis-SDF1 was repeated in 50 µL batches of CFPS reaction, using selected 30 % SHuffle lysate at 30 $^{o}$C. The optimum working concentration of Mg/Glu for the expression of soluble NHis-SDF1 was further confirmed and determined to be 6.4 mM (Figure 3.25). At 6.4 mM Mg/Glu, nearly all of the expressed NHis-SDF1 was found to be soluble. Mg/Glu concentrations at either below or above the optimum, led to decrease in the expression and solubility of NHis-SDF1. $Mg^{2+}$ ions, serving as a co-factor for RNA polymerase during transcription, enhance the rate of transcription and thus the rate protein synthesis. The increase in the rate of protein synthesis can however be detrimental for protein folding, which can explain the loss of solubility of NHis-SDF1.

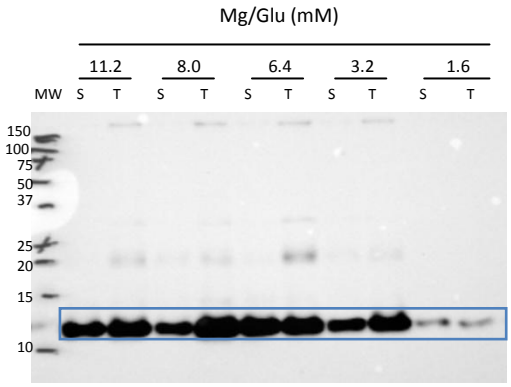



***Figure 3.25:*** *Anti-histidine western blot analysis of total (T) and soluble (S) NHis-SDF1 (boxed) expressed in 50 µL batches of CFPS reaction, with Mg/Glu concentration variations (1.6 – 11.2 mM). MW: Molecular weight.*

Upon establishment of the optimum Mg/Glu concentration, the expression of NHis-SDF1 served as a expression control for confirming the efficacy for soluble expression of SDF1-His and SDF1-LH (Figure 3.26). Despite being encoded in non-CFPS-optimised pET20-b plasmid vector, SDF1-His and SDF1-LH could be expressed abundantly. The expression of all of the SDF1 variants was clearly superior at 30 $^{o}$C compared to 20 $^{o}$C, with solubility maintained. With optimised expression, the results show that approximately 300 µg/mL of CFPS reaction of soluble NHis-SDF1 was produced. Almost 400 µg/mL, and 200 µg/mL of CFPS reaction of soluble SDF1-LH and SDF1-His respectively could be produced. The overall

expression level was, as expected, determined by the concentration of Mg/Glu used. The optimum concentration for soluble expression of both SDF1-LH and SDF1-His were found to be between 6.4 and 8.0 mM. Concentration of Mg/Glu below or above the optimum concentration proved once again to be detrimental for soluble expression of either protein.

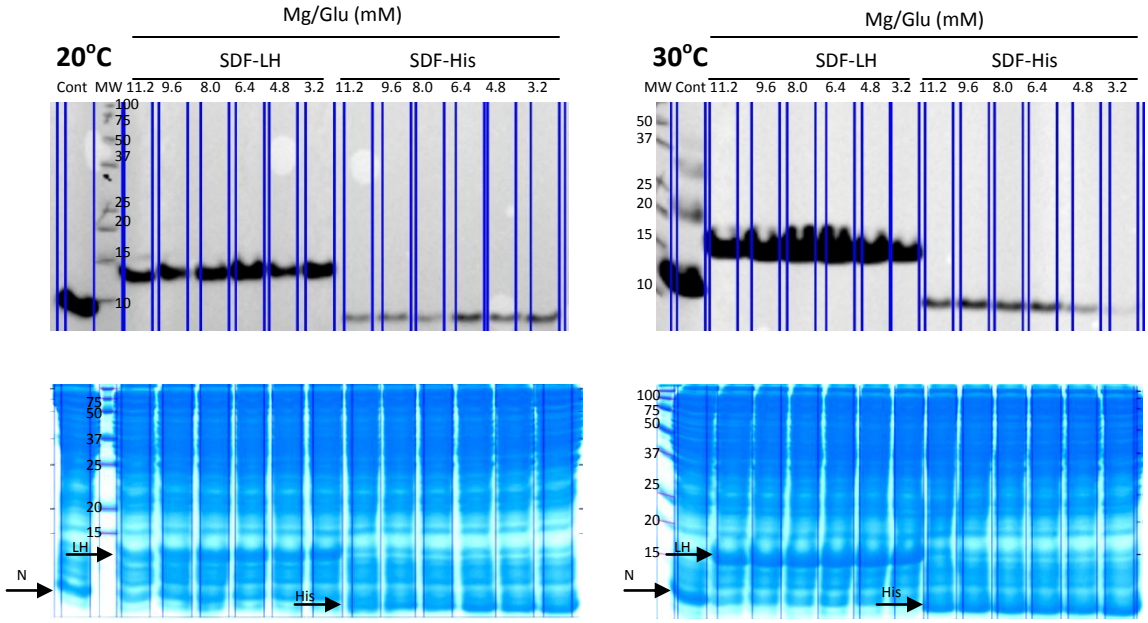



***Figure 3.26:*** *Anti-histidine western blot analyses (top) and Coomassie-blue stained gels (bottom) showing the soluble expression of NHis-SDF1 (N), SDF1-His (His) and SDF1-LH (LH) at either 20°C or 30°C, with variation in Mg/Glu concentration (3.2 – 11.2 mM). Reactions were carried out as 50 µL batches in 1.5 mL Eppendorf tubes, with 30 % SHuffle lysate and 300 mM K/Glu at 500 rpm in a Thermomix. MW: Molecular weight.*

### *Soluble expression of deuterated SDF1-α*

With the aim to nurture CFPS as a methodology to prepare deuterated proteins for use in small-angle neutron scattering (SANS) studies, the highly expressing soluble NHis-SDF1 was used as a model system to verify the viability of CFPS to produce deuterated proteins (Figure 3.27). In 50 µL of CFPS reaction, a study using deuterated amino acids (daa) instead of hydrogenated amino acids (haa), confirmed the expression of soluble deuterated NHis-SDF1 (d-NHis-SDF1). For d-NHis-SDF1, the optimum concentration of daa was found to be between 5 – 10 mg/mL of reaction (Figure 3.27A). Concentrations of daa below or above the optimum resulted in drop in overall d-NHis-SDF1 expression. At low concentrations of daa, amino acid quantity is limiting, and very rapidly depletion of one or more amino acid restricts the CFPS reaction. At high concentration, amino acids with hydrophobic side-chains (e.g methionine, phenylalanine, proline) are likely to become insoluble, and are therefore no longer available for CFPS reaction.

For subsequent applications in the field of SANS with match-out labelling, proteins need to have deuteration levels of approximately 75 % (Dunne *et al.*, 2017b). It is therefore important to verify that CFPS can be modulated to produce protein with suitably adjusted deuteration levels. The incorporation of mixed ratios of daa:haa still permitted the expression of d-NHis-SDF1 (Figure 3.27B). The deuteration level of d-NHis-SDF1was measured by mass-spectrometry, and is described in the next chapter.

Expression of soluble d-NHis-SDF1 under scaled-up conditions of 1 mL was confirmed and it was verified that with the use of 5 mg/mL of daa, CFPS was not compromised compared to standard CFPS where haa is used (Figure 3.27C). The results show that expression of soluble d-NHis-SDF1 was around 300 µg/mL of CFPS reaction, similar to the quantities expressed under unlabelled conditions.

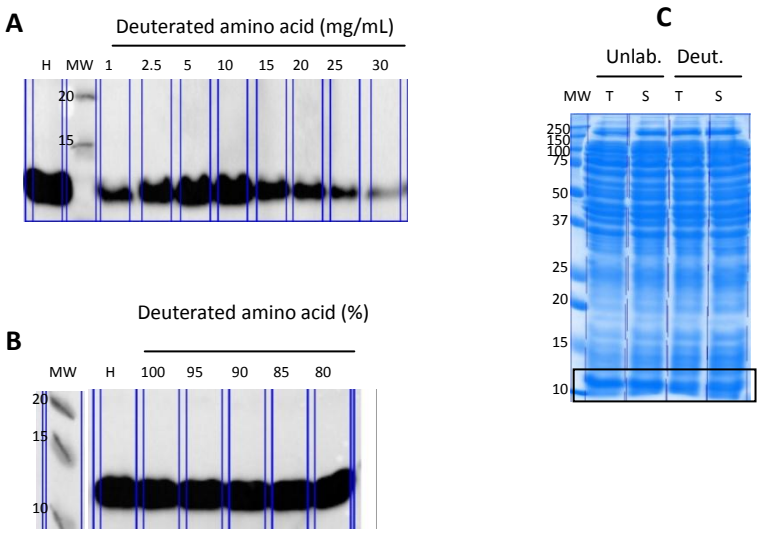



***Figure 3.27:*** *Anti-histidine western blots (A,B) and Coomassie-blue stained gel (C) of soluble deuterated NHis-SDF1 (boxed). Screening of deuterated amino acid (daa) incorporated in CFPS reaction confirmed the expression of soluble d-NHis-SDF1, with most protein produced between 5 – 10 mg/mL of daa (A). The use of mixes of daa and hydrogenated amino acid (haa) for the expression of NHis-SDF1 with varied daa content (B). The use of daa in CPFE enabled the production of about 300 µg/mL of soluble NHis-SDF1 (boxed), which is equal to expression yields when standard haa is used (C, shown as un-purified total (T) and soluble (S) fractions). MW: Molecular weight.*

### *Purification of SDF1-α*

Initial small-scale purification using the MagneHis™ protein purification system to purify either unlabelled or deuterated NHis-SDF1, derived from 1 mL of CFPS reaction, was very effective (Figure 2.28A). In contrast to CXCR4 purification, practically no protein remained in the flow through nor was present in washes and almost all of the protein was recovered in elutions. From 1 mL, a maximum total of 300 µg of purified soluble unlabelled or deuterated NHis-SDF1 could be obtained. Additionally, the MagneHis™ protein purification system could be used to purify SDF1-His and SDF1-LH resulting from CFPS reactions (Figure 2.28B). The MagneHis™ protein purification system therefore represents a very efficient, quick and easy method for purification of soluble histidine tagged cell-free expressed protein at a small scale (2 mL reaction volumes). NHis-SDF1 was successfully cleaved using thrombin digest with

almost 100 % purity of cleaved NHis-SDF1 being obtained (Figure 2.29). After buffer exchange by dialysis, all of these purified proteins were then analysed by mass spectroscopy and in functional chemotaxis assays. These will be described in the following chapter.

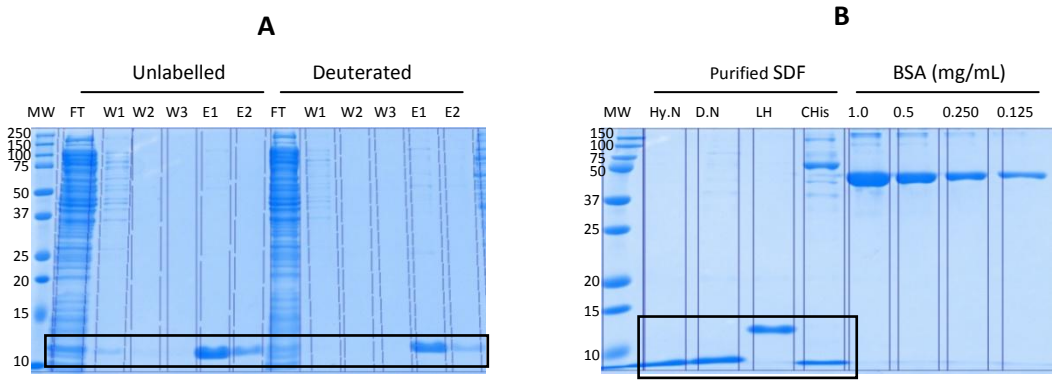



***Figure 3.28:*** *Coomassie-blue stained gel showing the results of purification of unlabelled and deuterated NHis-SDF1 (A), and SDF1-His and SDF1-LH (B) using the MagneHis™ protein purification system. The yields of purified protein were 300 µg/mL of CFPS reaction for unlabelled and deuterated NHis-SDF1, and about 250 µg/mL of CFPS reaction for both SDF1-His and SDF1-LH. MW: Molecular weight.*

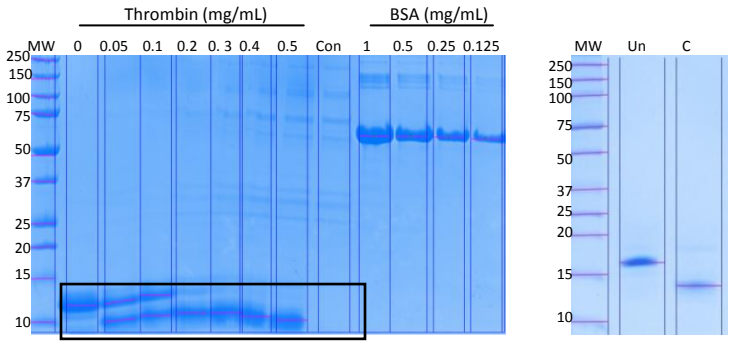



***Figure 3.29:*** *Coomassie-blue stained gel of purified cleaved NHis-SDF1 used for subsequent chemotaxis assay (see next Chapter). Left: Titration assay with increasing thrombin concentration with overnight incubation at 37 °C. Right: Control un-cleaved and cleaved NHis-SDF1 that has been flowed through a benzamidine sepharose 4 FF column to remove thrombin. MW: Molecular Weight; Un: Un-cleaved NHis-SDF1; C: Cleaved NHis-SDF1.*

Following the confirmation that soluble NHis-SDF1 could be produced and purified at a small-scale, large-scale expression of NHis-SDF1 in 10 mL CFPS reaction batches was carried out (Figure 3.30A). NHis-SDF1 was then purified from the reaction mix by nickel affinity using the ATKA Prime system resulting in 3 – 4 mg of eluted protein with purity of above 90 % (Figure 3.30B, C). Further purification to obtain a purity suitable for structural biology work and also removing imidazole was performed using SEC. Highly purified, monodisperse, monomeric SDF1 (Figure 3.30D, E) was obtained with total yields between 2 – 2.5 mg (Figure 3.30E).

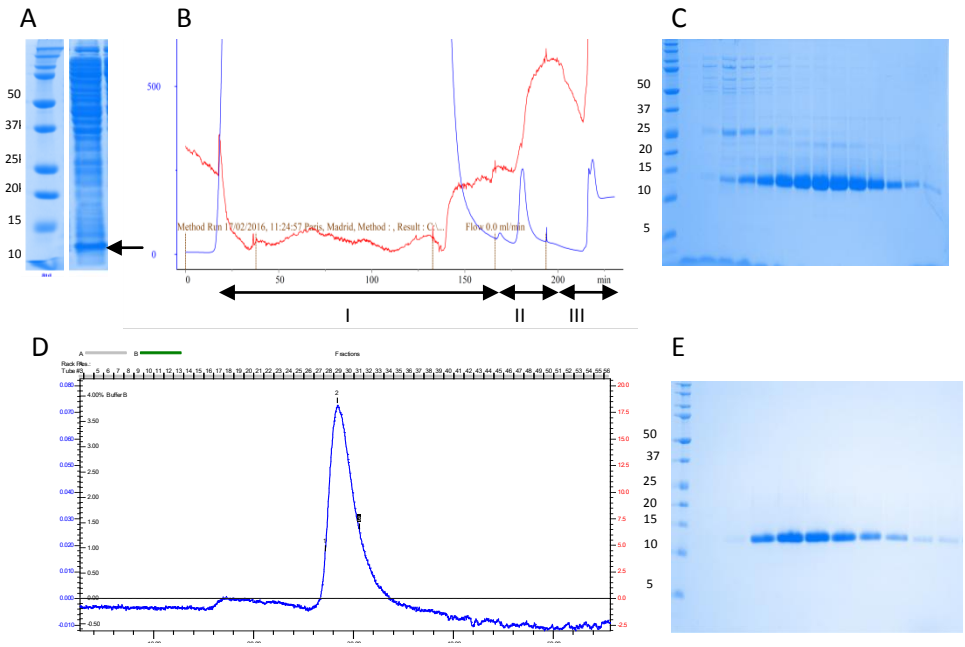



***Figure 3.30:*** *Large-scale purification on NHis-SDF1. NHis-SDF1 was present in the supernatant of the centrifuged CFPS reaction (A). Purification of NHis-SDF1 was carried on the AKTA Prime instrumentation (B) – blue: 280 nm UV profile; red: conductivity. The supernatant, diluted to 50 mL of binding buffer, was allowed to slowly flow through the nickel affinity column (B, I). After binding, 40 mM imidazole wash was carried (B, II), followed by gradient elution with 500 mM Imidazole buffer (B, III). Coomassie-blue stained gel of 1 mL elution was carried out to confirm the presence and purity of NHis-SDF1 (C). SEC was then conducted to further purify mono-disperse NHis-SDF1 (D and E).*

### *Optimising purification of SDF1-α for crystallisation*

As described in chapter seven, obtaining a crystal of NHis-SDF1 proved to be difficult. Furthermore, as described in chapter four, NHis-SDF1 showed poor response in the chemotaxis of CXCR4+ cells. The N-terminal of SDF1-α is critically involved in the binding to CXCR4. Therefore, the steric hindrance brought about positioning of the histidine tag at the N-terminal is the cause of poor chemotaxis response. As an alternative therefore, soluble CHis-SDF1 was also expressed for the purposes of crystallisation. CHis-SDF1 is advantageous compared NHis-SDF1 since the histidine tag is located at the C-terminal, and therefore does not compromise the functional activity of SDF1-α (see chapter four). As such, whilst it is possible to cleave the histidine tag using TEV cleavage, purified CHis-SDF1 can, however, be used directly in functional assays.

Using the same conditions for the expression of NHis-SDF1, both NHis-SDF1 and CHis-SDF1 were expressed in directly as 50 mL batches of CFPS reaction. The proteins were purified by nickel affinity using an ÅTKA Prime system, concentrated and subjected to SEC (Figure 3.31). Unfortunately however, due to the lower quality of lysate (due to machinery fault during French-press lysis step) used in this case, much lower expression of SDF1-α was achieved compared to previously with NHis-SDF1.

Instead of the SEC buffer as used previously for the NHis-SDF1 purification (50 mM HEPES pH 7.5, 500 mM NaCl), NHis-SDF1 and CHis-SDF1 were exchanged into a buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM β-mercaptoethanol, 3 % glycerol) similar to that described by Ryu *et al.*, where refolded *E. coli* expressed SDF1 was crystallised (Ryu *et al.*, 2007). Single peaks in the SEC profiles of both NHis-SDF1 and CHis-SDF1 indicated monodispersity. With CHis-SDF1 having a lower molecular weight than NHis-SDF1, the molecule emerged later in the SEC elution. Albeit with lower amounts of recovered proteins compared to previously with NHis-SDF1, these were nevertheless concentrated (5 mg/mL in 100 µL volume) sent for crystallisation trials at the EMBL/PSB high-throughput crystallisation platform (HTX-lab, Grenoble) – see chapter seven.

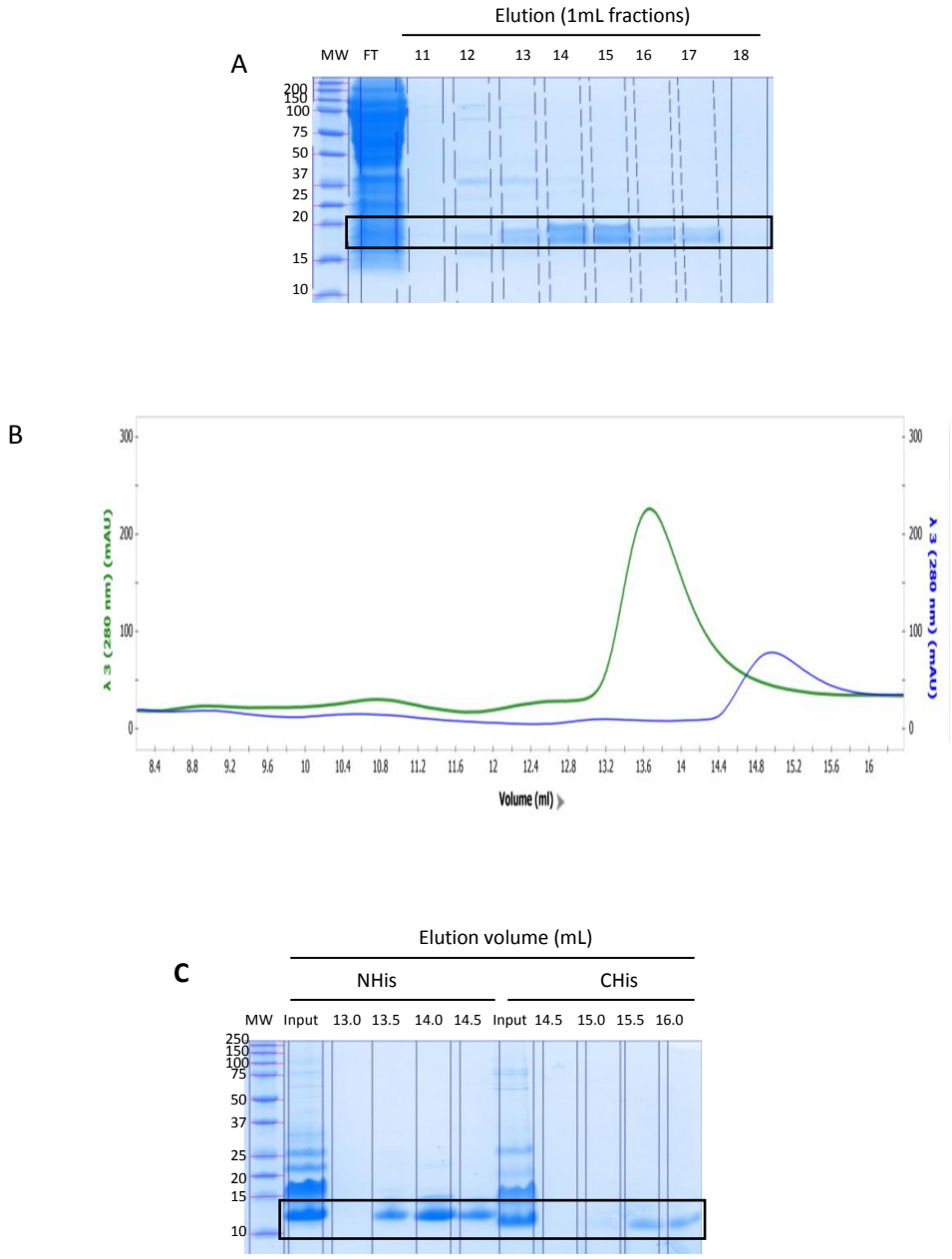



***Figure 3.31:*** *Purification of CHis-SDF1, with NHis-SDF1 serving as the reference control. From 50 mL of batch reaction, CHis-SDF1 was purified by Ni-NTA affinity (A), and subject to previously calibrated SEC (B), with buffer exchange into recommended crystallisation buffer. The NHis-SDF1 (green) and CHis-SDF1 (blue) SEC profiles showed monodisperse and monomeric species, with smaller CHis-SDF1 emerging later in elution. Purified NHis-SDF1 and CHis-SDF1 were visualised on Coomassie-blue stained gel (C). MW: Molecular weight.*

### 3.3.4. Expression of soluble 2dCD4

#### *Confirming CFPS expression of 2dCD4-wt*

From prior experience at Synthelis (unpublished data), the use of a histidine tag for purification purposes on the N-terminal of a target protein is generally more permissible for CFPS compared to a C-terminal tag. As such, constructs of both 2dCD4-wt-NHis and 2dCD4-wt-Chis were initially considered in CFPS trials. Both 2dCD4-wt-NHis and 2dCD4-wt-Chis were successfully expressed using SHuffle lysate (30 % lysate, 8.0 mM Mg/Glu, 300 mM K/Glu) (Figure 3.32). Expression of both proteins was possible at both 20 $^{o}$C and 30 $^{o}$C. However, the presence of truncated species of 2dCD4-wt-NHis, which can be observed on the anti-histidine western blot (Figure 3.32), would co-purify during Ni-NTA purification. Furthermore, the presence of a histidine tag at the C-terminal is less likely to interfere with the functionality of CD4 since the N-terminal is closely associated with binding to the partner protein, GP120, during the HIV infection process. Since both proteins could be expressed successfully, work therefore continued with the 2dCD4-wt-CHis construct.

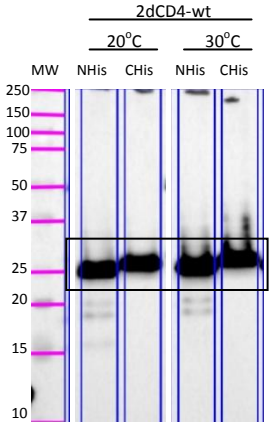



***Figure 3.32:*** *Anti-histidine western blot analysis of CFPS of 2dCD4-wt-NHis and 2dCD4-wt-Chis at 20$^{o}$C and 30$^{o}$C (shown in the box) with standard reaction conditions of 30 % SHuffle lysate, 8.0 mM Mg/Glu, 300 mM K/Glu. MW: Molecular weight.*

### *CFPS expression and purification of deuterated 2dCD4-wt-Chis*

In the specific work of this thesis, the formation of the GP120-CD4 complex for SANS is of concern. Deuterated soluble 2dCD4-wt-Chis was produced in 50 mL CFPS reaction batches and purified from the reaction mix by nickel affinity using the ATKA Prime system resulting in 3 – 4 mg of partially purified protein (Figure 3.33A). Following SEC, highly purified 2dCD4-wt-Chis was obtained with total yield of between 2.5 – 3 mg (Figure 3.33B, C, fractions 15.5 – 18 mL).

### *Expression and purification of GP120*

For the purposes of forming the complex with 2dCD4, GP120 was expressed in mammalian cells. Highly purified GP120 was successfully eluted from HEK293 cell culture supernatants, and subject to SEC (Figure 3.34). From batches of 3 L of supernatant, the total yield of purified GP120 was about 4 mg. Whilst not being visible on the Coomassie-blue stained SDS-PAGE gel due to the denaturation of proteins, the SEC profile indicated the presence of dimeric and monomeric species of GP120. Of the total amount of GP120, it was estimated that the ratio of dimeric to monomeric GP120 was about 2:1. Following a first SEC, the eluates containing the two species were separated and subjected to another round of SEC. This second SEC confirmed that the GP120 species are not in a dynamic equilibrium, but remain as stable monomeric and dimeric forms.

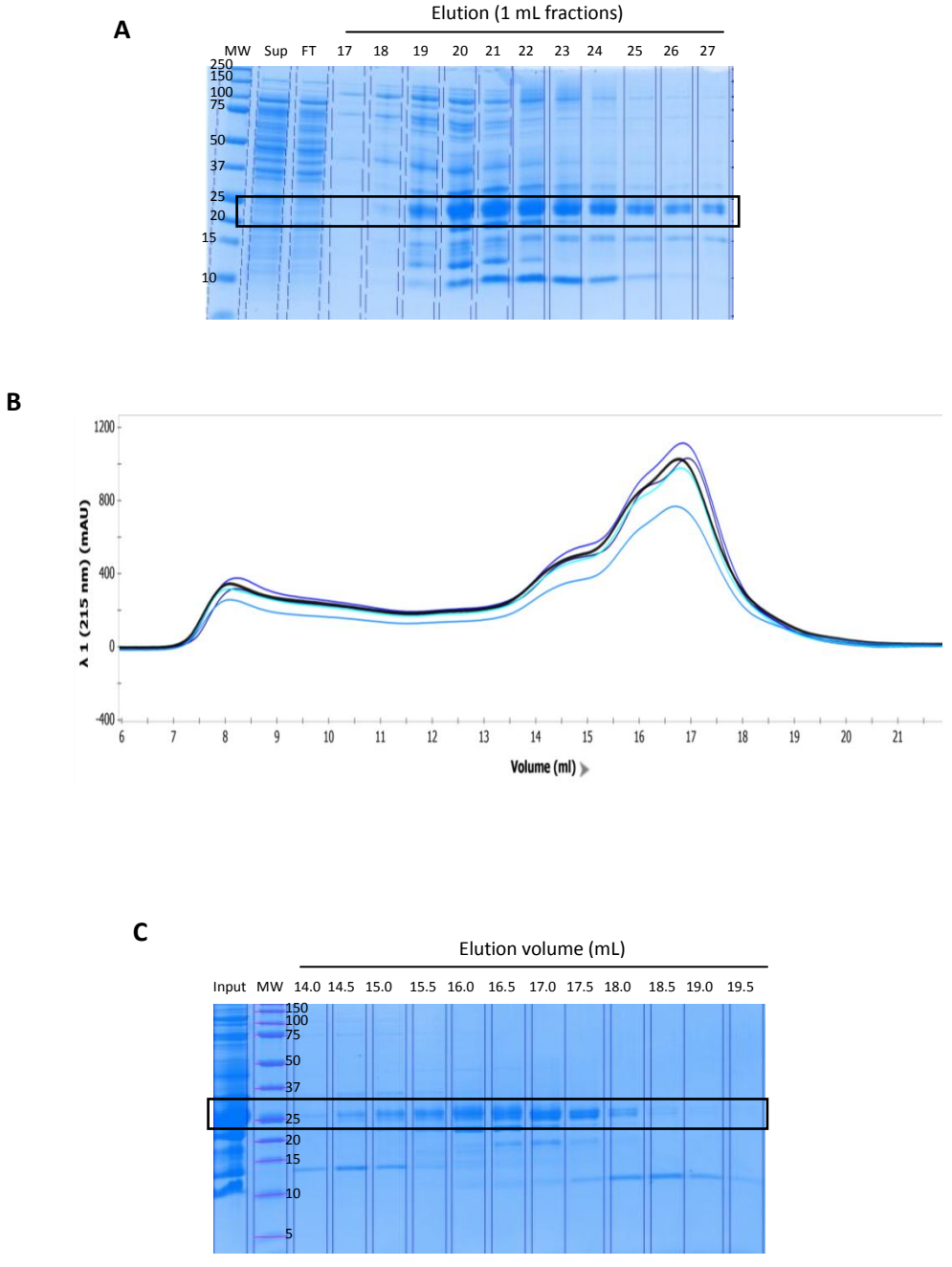



***Figure 3.33:*** *Analysis by Coomassie-blue stained gel of purification steps for deuterated CFPS 2dCD4-wt. (A) Coomassie-stained gel analysis of 50 mL of deuterated 2dCD4-wt reaction purified by Ni affinity (box). (B) SEC purification profile (multiple runs) of deuterated 2dCD4-wt. (C) Coomassie-stained gel analysis of the fractions of the SEC purified deuterated 2dCD4 (box). MW: Molecular weight.*

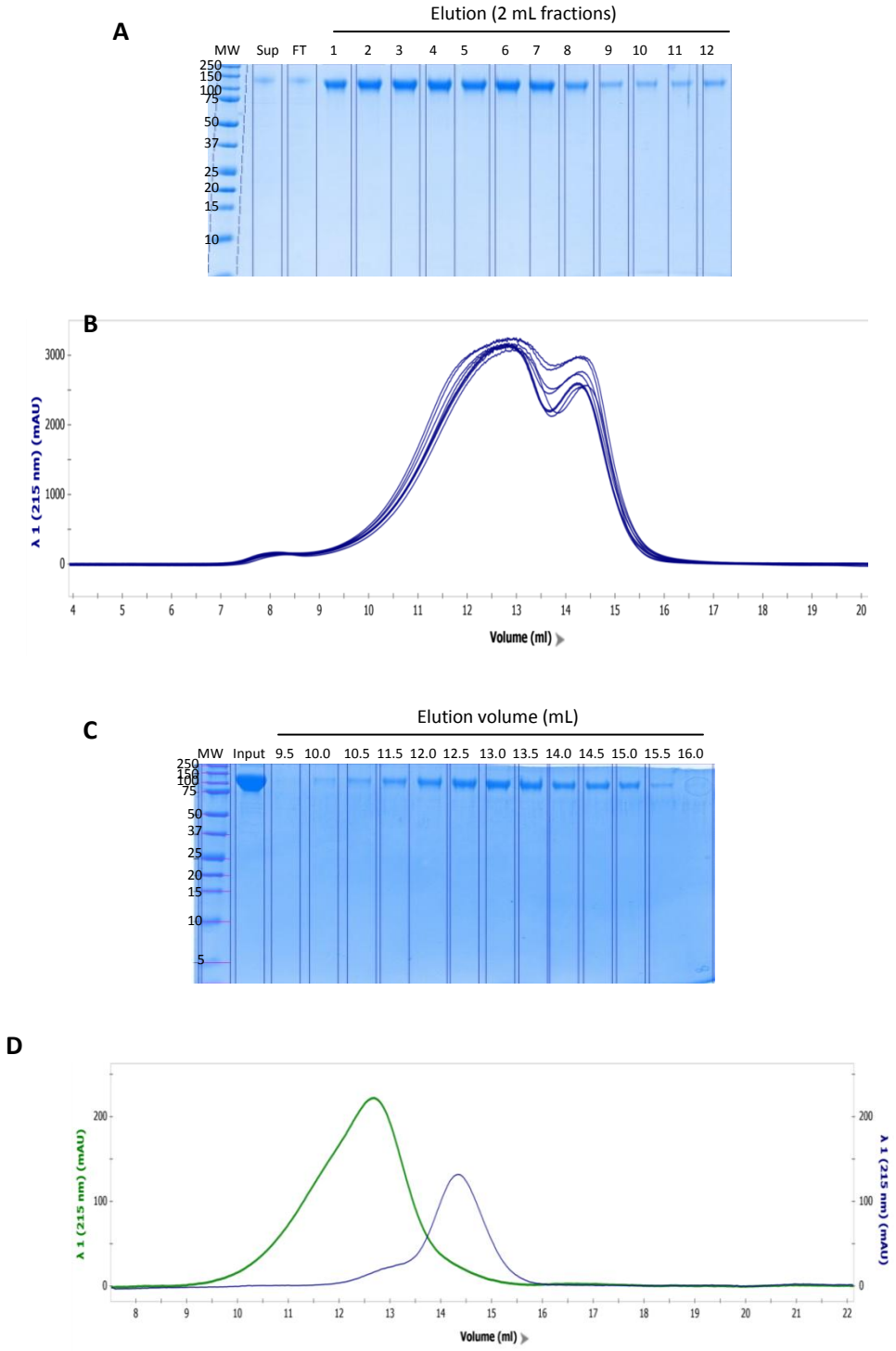



***Figure 3.34**: Analysis by Coomassie-blue stained gel of purification steps of mammalian cell expressed GP120. (A) Coomassie-stained gel analysis of lectin purification of GP120. (B) SEC profile of GP120 purification. (C) Coomassie-stained gel analysis of fractions from SEC. (D) SEC profile showing the separation of dimeric (green) and monomeric (blue) forms of GP120. MW: Molecular weight.*

### *Formation and purification of deuterated 2dCD4-GP120 complex*

Since access to neutron instruments was limited and scheduling was very tight, it was imperative to obtain a GP120:2dCD4 complex with as high a concentration as feasible within the limited amount of time. Since purified dimeric GP120 was available in the largest quantity, this species was selected for complexing with 2dCD4-wt-Chis. Freshly purified dimer GP120 and 2dCD4-wt-Chis were mixed with each other and incubated in order to generate the complex. The sample was then concentrated and subject to SEC in order in order to separate complex from unbound 2dCD4.

The SEC profile (Figure 3.35, multiple runs) showed the presence of two distinct peaks: the first one being the GP120-2dCD4-wt-Chis complex (Figure 3.35, Peak I, fractions 10 – 14.5 mL), and the later peak that of unbound 2dCD4-wt-Chis (Figure 3.35, Peak II, fractions 15.5 – 18 mL). Indeed, as seen on Coomassie-blue stained SDS-PAGE gel, there was co-elution of 2dCD4-wt-Chis with GP120 which indicated successful formation of the complex. The complex was then concentrated to approximately 10 mg/mL in a final volume of 300 µL. The sample was used for analysis by SANS analyses which are described in chapter five. It should however be mentioned that a limitation of this SEC purification technique is the lack of resolution in order to successfully separate unbound GP120 dimers and complex. Therefore, the sample used for SANS was likely to contain some unbound GP120 dimer.

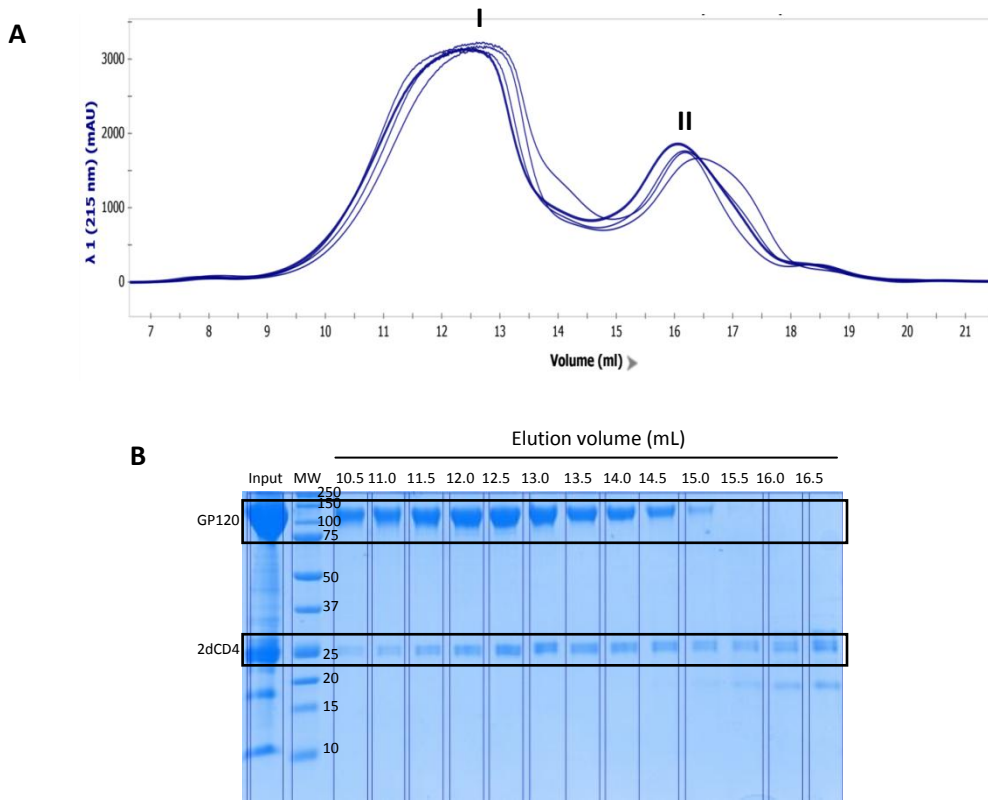



***Figure 3.35:*** *Analysis by Coomassie-blue stained gel of purification steps of deuterated 2dCD4-GP120 complex.SEC elution profile (A, multiple runs) showing peaks of GP120-CD4 complex (I) and un-complexed CD4 (II); and Coomassie-blue gel of deuterated 2dCD4-GP120 complex (B). The co-elution of 2dCD4 in with GP120 indicates successful complex formation. MW: Molecular weight*

### 3.3.5. Summary of CPFS expression and purification

Optimisation of CFPS towards scaling-up protein production to cope with the structural needs was a key consideration throughout this work. The table below summarises the maximum expression and purification yields of the proteins (full-length, excluding truncated species) which were considered in this chapter. Whilst some of these figures remain approximations, they provide an indication of the expected expression and yields, and act as a reference upon which improvements to the protocols can be made. In this context, CXCR4 expression needs to be increased and efficient scaling up protocols established, for both proteoliposome and detergent-based CFPS. Particularly for detergent-based CFPS, the expression yields for CXCR4-lys recorded here for scaled up reaction did not result in a linear increase. This can be attributed to multiple factors such as inadequate aeration, shaking and rapid accumulation of inhibitory inorganic magnesium pyrophosphates by-products. Nevertheless, despite low expression for 2dCD4 and CXCR4, the purification of the former was however very efficient, which allowed for high recovery. CFPS was found to be an efficient system for the expression of deuterated form of 2dCD4, which was considered for SANS as described in chapter five. SDF1-α expression and purification was particularly successful and allowed the protein to be taken ahead for X-ray crystallisation, as described in chapter seven.

| Full-length (FL) Protein | Small-scale CPFE expression (µg/mL of CFPS reaction) | Small-scale Soluble CPFE expression (%) | Recovery of FL protein after purification (%) | Protein Purity after purification (%) | Maximum CFPS reaction volume achieved (mL) | Maximum protein recovered from maximum CPFE volume (µg) |
|---|---|---|---|---|---|---|
| **CXCR4-wt (liposome)** | 70 (70) | NA | 65 | 30 – 50 | 10 | 400 – 450 |
| **CXCR4-lys (liposome)** | 70 (35) | NA | NT | NT | NT | NT |
| **CXCR4-wt (detergent)** | 70 (70) | 100 | NT | NT | NT | NT |
| **CXCR4-lys (detergent)** | 50 (50) | 100 | 90 – 100 | 90 – 95 | 50 | 90 – 100 |
| **NHis-SDF1** | 200 | 100 | 100 | 100 | 50 | 3000 |
| **CHis-SDF1** | 200 | 100 | 100 | 100 | 50 | 3000 |
| **CD4-FL-NHis** | 20 | NA | 65 | 10 – 20 | 2 | 30 |
| **CD4-FL-CHis** | 20 | NA | 65 | 10 – 20 | 2 | 30 |
| **2dCD4-wt-NHis** | 30 | 100 | 100 | 1000 | 2 | 60 |
| **2dCD4-wt-CHis** | 30 | 100 | 100 | 100 | 100 | 3000 |
| **2dCD4-wt-CHis (deut)** | 30 | 100 | 100 | 100 | 100 | 3000 |

***Table 2.2:*** *List of proteins expressed, approximate yields, purification recovery and maximum scaled up production levels achieved in this study. In brackets is the amount of dimer of protein. NA, not applicable; NT, not tested.*

### 3.3.6. Creating standard operating procedures

Quality control (QC) is critical in the biotechnology industry in order to ensure quality of the product and services provided. QC encompasses multiple facets, which include sampling, testing, optimising, implementation, validation and re-evaluation of standard working methods. Of essence in QC, are standard operating procedures (SOPs), which serve to break-down complex operations into simpler tasks, and which can be easily and reliably followed. The detailed and clear instructions of SOPs thus ensure accuracy, reproducibility and uniform performance in the workplace, while respecting health and safety regulations, which is of notable concern in the biotechnology industry. The work presented here, can therefore serve for the establishment of such generalised working protocols. These protocols, extracted as a result of the work of the last two chapters on the proteins considered here, are given in the appendices as flow-charts going from initial CFPS feasibility through to protein production and purification.

## 3.4. DISCUSSION

### 3.4.1. Better expression of CXCR4 variants in proteoliposome using Rosetta lysate

The use of different lysates from various *E. coli* sources has been shown to influence the efficiency and yields of CFPS (Krinsky *et al.*, 2016). Here, starting off with the inclusion of liposomes in the reaction, Rosetta lysate was found to be superior to the other tested lysates, including the previous standard Nico21, as it improved yields of full-length CXCR4-wt and CXCR4-lys, whilst reducing the amount of truncated protein (Figure 3.7). Consistent with previous studies, Rosetta strain-based lysate was found to lead to increased effectiveness of CFPS reaction (Kigawa *et al.*, 2004). The Rosetta strain is particularly adapted for the over-expression of rare tRNA, which would favour complete protein translation during *in vivo* expression. However, in the case of CFPS, the reaction is supplied with an excess of tRNAs. This would imply that some other intrinsic properties of Rosetta, related or not to the over-expression tRNAs, could be promoting the improved CFPS. With the inclusion of protease inhibitors in the reaction mix, the immediate appearance of truncated species within first few hours of expression (Figure 3.8) and the absence of truncated C-terminal strep-tagged CXCR4 species (Figure 3.18), seems to exclude post-translation proteolysis. Lowering the expression temperature, which is known to promote protein folding, did not specifically reduce truncated species, but lowered overall protein expression.

### 3.4.2. CXCR4-lys is expressed in CFPS

As described in a previous study, to permit the crystallisation of CXCR4, the protein has been heavily modified to include an intracellular T4-lysozyme fusion inserted between the fifth and sixth transmembrane helices, as well three stabilising point mutations (Wu *et al.*, 2010). Further described in this previous study, the expression of the CXCR4 was carried out in baculovirus-infected *Spodoptera frugiperda* (Sf9) insect cells and was purified from membranes using DDM-CHS detergent based buffers. Here however, CFPS was used for the expression of both wild-type and lysozyme mutant of CXCR4 (Figure 3.7, 3.17, 3.18). The low yields of the purified protein however, coupled with the contamination with truncated species, did not enable us to use of CXCR4 for crystallisation. Indeed, highly homogenous CXCR4-lys concentrated to 60 – 70 mg/ml and with above 95 % purity, was eventually required for crystallisation (Wu *et al.*, 2010). Therefore, improving the yields of full-length CXCR4 expression in *E. coli*

based CFPS remains a critical step. With the openness of the system, there is certainly potential in improving the yields of full-length CXCR4. For instance, addition of folding chaperones groES-groEL was reported to enhance the folding of CXCR4 and analogue CCR5 (Chi *et al.*, 2016, Chi *et al.*, 2015).

### 3.4.3. Understanding protein species in CFPS: truncation and dimers

Throughout, the expression of CXCR4 in CFPS always led to the formation of truncated species (Figures 3.7 – 3.18). The formation of such truncated species is a wasteful process in the CFPS reaction as it consumes resources which would rather be more effectively utilised in the production of full-length protein. Ultimately, the formation of truncated species reduces the yields of the full-length protein and their presence is detrimental to the purification process. Understanding reason for the presence of truncated species is therefore important in order to avoid their formation.

The sequence of the encoding DNA plays a critical role in determining complete, or incomplete, protein expression in CFPS. Prokaryotic protein translation is a complex series of well-orchestrated events, which is initiated by the binding of mRNA-containing ribosome binding site (RBS) to a ribosome upstream of the start codon, and the recruitment of multiple initiation factors (Laursen *et al.*, 2005, Kramer *et al.*, 2009). Native *E. coli* shows strong bias against the presence of such RBS, or RBS-like sites, in the protein encoding segments (Plotkin and Kudla, 2011). When occurring within the protein coding mRNA sequence, disruption in protein elongation ensues, resulting in truncated species (Whitaker *et al.*, 2015). The RBS in prokaryotes is characterised by the Shine-Delgarno (SD) sequence (AGGAGGT) which, if removed from the coding sequence of mRNA, has been shown to reduce the truncated species of yeast histone acetyltransferase in *E. coli* CPFE (Jennings *et al.*, 2016). Indeed, the presence of internal SD-sites or SD-like sites in the coding DNA, leads to ribosome stalling, slowing down of translation rate and, occasionally, interruption of protein translation (Li *et al.*, 2012). To increase efficiency of the CFPS studied here, a systematic study in the alteration of the coding sequences of CXCR4-wt and CXCR4-lys to avoid codons which interrupt protein translation is to be considered for future work.

### 3.4.4. Confirming *membrane protein insertion into liposomes to form proteoliposomes*

Recombinant membrane protein expression remains challenging due to its intrinsic hydrophobicity and the need to be supported structurally by the lipid bilayer. Backed by previous studies which have confirmed the use of *E. coli*-based CFPS systems for functional protein expression into proteoliposomes for structural biology applications (Soranzo *et al.*, 2016, Soranzo *et al.*, 2017, Deniaud *et al.*, 2010), the expression of CXCR4 in proteoliposomes was therefore explored. The inclusion of liposomes in the CFPS reaction provides an artificial lipid environment for the *in-situ* incorporation of functional membrane protein (Nozawa *et al.*, 2011). Indeed, from our work here, the possibility to recover CXCR4 in the lipid layer after sucrose-gradient ultracentrifugation (Figure 3.13, 3.14) (Liguori and Lenormand, 2009, Nozawa *et al.*, 2011), carbonate and urea washes (Figure 3.12, 3.13) (Long *et al.*, 2012, Fujiki *et al.*, 1982, Sakai *et al.*, 2009) indicates membrane insertion. The structural verification of functional cell-free expressed proteoliposome p7 viral protein in its pore-forming capacity by neutron reflectivity, confirmed the correct insertion of the membrane protein in the lipid bilayer (Soranzo *et al.*, 2017). More recently, the functionality of cell-free expressed CXCR4 in its capacity to bind to synthetic ligand T22 was demonstrated by surface plasmon resonance (SPR) and atomic force microscopy (AFM) (Chadli *et al.*, 2017). Furthermore, the inclusion of liposome in the reaction mix has been reported to enhance the yields of KcsA potassium channel in CFPS (van Dalen *et al.*, 2002). Additionally, a proteoliposome based CFPS system was used for production of full-length CD4 (Figure 3.15), which as depicted by its crystal structure (PDB, 1WIO), is an elongated, rod-like structure that is comprised of four distinct domains (Wu *et al.*, 1997). Full-length CD4, as expressed in such a case, lacked the transmembrane and intracellular domains, and was expressed as a soluble protein using CHO mammalian cells (Kwong *et al.*, 1990). Here, however, full-length CD4, including the transmembrane and intracellular domains, was expressed into a lipid bilayer using *E. coli* based CFPS. CPFE therefore shows potential as an alternative for lipid-embedded membrane protein expression. The functionality and hence, conformational integrity of full-length CD4 embedded in its proteoliposomes is described in chapter four.

### 3.4.5. The challenge of membrane protein purification

Whilst allowing for insertion into the liposome and immediate recovery of functional membrane proteins, the inclusion of liposome in CFPS also provides a hydrophobic environment for the non-specific attachment of lysate derived proteins (Nozawa *et al.*, 2011, Soranzo *et al.*, 2016). Complete elimination of these contaminants, whilst preserving the integrated *de novo* synthesised protein of interest, and without destroying the proteoliposome, remains a significant challenge. For use in structural studies via crystallisation and X-ray crystallography, and through small angle scattering, highly homogenous and purified protein is required. The inclusion of detergents instead of liposomes in CFPS for soluble expression of CXCR4 was therefore explored as it could enable further protein purification (Figure 3.16 – 3.23). The use of CPFS for the production of insoluble membrane proteins (i.e. without the inclusion of hydrophobic matrix in the reaction), followed by detergent solubilisation was not considered here since it was deemed as equivalent to protein refolding, which was a central tenet to be avoided by using CFPS. Further, as exemplified by Fos-choline detergent solubilised cell-free expressed human VDAC having far-UV CD spectra that was closely followed that of SDS denatured VDAC, re-solubilisation of precipitate does not guarantee folding integrity (Nguyen *et al.*, 2010). Additionally, whilst enabling purification and crystallisation, crystals grown from detergent-solubilised VDAC expressed by CFPS showed relatively poor X-ray diffraction (Deniaud *et al.*, 2010). On the other hand, VDAC expressed in the presence of mild Brij35 detergent had far-UV CD spectra which showed a high degree of β-barrel conformation, indicating folding integrity (Nguyen *et al.*, 2010).

### 3.4.6. Solubilised CXCR4 expression occurs with the inclusion of detergents in CFPS reaction mix

Here, a variety of detergents, including non-ionic detergent alkyl-glycosides (DDM), polyoxyethylene-alkyl-ethers (Brij derivatives), polyethylene-glycol derivatives (Triton-X100) and steroid derivatives (CHAPSO, Digitonin) were tested at a range of CMCs for CPFS expression of CXCR4-wt and CXCR4-lys (Figures 3.16 – 3.20). It was found that Brij detergents were the most efficient in producing solubilised CXCR4. Consistent with a previous study, where Brij35 was also used for CFPS expression of solubilised CXCR4 (Chi *et al.*, 2016).

Additionally, the Brij family of detergents was also particularly effective for cell-free expression of solubilised bacterial α-helical multidrug transporter EmrE, β-barrel nucleoside transporter Tsx, or porcine GPRC vasopressin receptor V2R (Klammt *et al.*, 2005). In a study comprising of CFPS of 13 GPRCs, Brij35, followed by Brij58, were found to be the two best detergents in enabling greater than 80 % soluble expression for all of the tested proteins (Corin *et al.*, 2011). In another study, Brij35 and digitonin were found to be most effective in permitting soluble CFPS of thioredoxin-fused GPRCs (Ishihara *et al.*, 2005). On the other hand, whilst being the recommended detergent for CXCR4 solubilisation (Wu *et al.*, 2010), DDM proved to be detrimental for CPFS expression here. Indeed, according to previous work, DDM was described to limit or completely inhibit *E. coli*-based CFPS of GPRCs, but not to influence expression of non-GPRC membrane proteins (Ishihara *et al.*, 2005, Klammt *et al.*, 2004a).

The intrinsic solubilisation properties of a detergent are governed by its structure and concentration, which influences its critical micellar concentration (CMC). The CMC is the minimum concentration of the detergent at which it forms micelles. Typically, isolation and solubilisation of a membrane protein from its membrane involves using a detergent above its CMC, which firstly leads to destabilisation and interaction of detergents with lipid components. This is followed by the formation of detergent-shielded lipid-protein fragments and the eventual occurrence of soluble detergent-bound membrane proteins in micelles (le Maire *et al.*, 2000). The CMC and micelles of detergent are however highly dynamic structures which are affected by multiple factors such salt concentration, pH or temperature (Sammalkorpi *et al.*, 2007). The highly heterogeneous CFPS reaction mix therefore makes it very difficult to predict the soluble expression of membrane proteins without experimental testing. Furthermore, the detergent used needs to be compatible with CPFE. DDM, for instance here, compromises overall CPFS of CXCR4. Additionally, even small modifications in the sequence of CXCR4 can severely alter its solubility in detergents, as it was the case of CXCR4-h/s in this work.

### 3.4.7. *High expression levels of SDF1-α using SHuffle lysate*

CFPS was particularly successful here for high-levels of soluble SDF1-α expression (Figures 3.24 – 3.26). SDF1-α when typically produced *in vivo* in *E. coli* host cells produces insoluble protein which needs to undergo a complex refolding process (Picciocchi *et al.*, 2014, Murphy *et al.*, 2007). The SDF1 was produced here as a soluble protein could be easily and rapidly purified (Figure 3.28), of which the

activity was confirmed (see chapter four). The expression of soluble protein could be explained by the correct formation of disulphide bonds during protein translation during CFPS. The strongly reducing intracellular environment i*n vivo* in *E. coli* prevents the disulphide bond formation (Baneyx and Mujacic, 2004). With the openness of CFPS, the *E. coli* based system can be controlled to permit the correct protein folding and disulphide bond formation (Goerke and Swartz, 2008, Yang *et al.*, 2004, Kolb *et al.*, 2000). Namely, with the inclusion of isomerase DbsC in CFPS, the expression of proteins with correctly paired disulphide bonds has been demonstrated in previous studies (Zawada *et al.*, 2011a, Yang *et al.*, 2004, Kim and Swartz, 2004). It was seen in this study, that the use of the SHuffle *E. coli* strain which constitutively expresses disulphide isomerase DbsC, is important for the expression of soluble and correctly folded SDF1-α. Indeed, the use of Nico21 as opposed to SHuffle lysate showed lowered expression of soluble SDF1 (Figure 3.24).

### 3.4.8. *CFPS of labelled SDF1-α*

Additionally, with the absence of cellular boundaries, the openness of the system provides a relatively easy way to produce isotopic-labelled proteins, especially for the field of NMR analysis (Kigawa *et al.*, 1999, Matsuda *et al.*, 2007, Staunton *et al.*, 2006). By incorporating deuterated amino acids in the CFPS reaction, deuterated SDF1-α was expressed (Figure 3.27). In chapter four, it will be shown that both the unlabelled and deuterated SDF1-α, are biologically active. The production of deuterated protein *in vivo* in *E. coli* requires the adaptation of the transformed *E. coli* to grow and express the protein of interest in deuterium-based Enfors medium (Haertlein *et al.*, 2016, Dunne *et al.*, 2017a). However, this task is labour-intensive, time-consuming and costly – especially when it concerns challenging protein targets. The use of CFPS for the production of deuterated proteins can prove advantageous over the *in vivo* system. Firstly, direct incorporation of deuterated amino acids in the reaction allows for deuterated protein production. Secondly, the reaction is not limited by cellular viability. This can be problematic when growing cells in $D_2O$ based media where the cells must be adapted to a deuterated environment. These two advantages, when coupled together, demonstrate CFPS as a potentially rapid method for the expression of deuterated protein. However, it should be noted that deuterated protein expression by CPFS can be rather costly compared to the classical method of deuterated protein expression *in vivo*. The overall costs, efficiency of production and time limitations are therefore factors to be considered in determining the appropriateness of each system.

### 3.4.9. *Establishing large-scale CFPS for structural studies with 2dCD4*

As reported by Zawada *et al.*, CFPS scaled up to 100 L for the successful production of a human cytokine produced about 700 mg/L within 10 h (Zawada *et al.*, 2011a). The production of cytokine SDF1 in *E. coli* based cell-free lysate here was optimised to produce about 0.25 mg of purified protein per millilitre of reaction. As described later (chapter five), the resulting CFPS SDF1-α was successfully crystallised and used for X-ray crystal diffraction. With the decreasing costs of reagents due to increase supply, and optimisation of the reaction systems for increased yields of high-quality, folded, soluble proteins, the cell-free system certainly has potential as an alternative system for the expression of proteins for structural biology in both the academic and industrial settings (Carlson *et al.*, 2012b, Zawada *et al.*, 2011b).

Based on the success of SDF1-α expression, the exploitation of CFPS for soluble proteins was repeated with two-domain soluble CD4 (2dCD4) expression. 2dCD4 is one of the most important and practical molecules used in the field of HIV research. Indeed, 2dCD4 represents the established epitopes for primary viral GP120 binding (Kwong *et al.*, 1998), and has been used in an innumerable number of structural, immunogenic and clinical studies, based around therapeutic neutralising antibody research (Pegu *et al.*, 2014, Kong *et al.*, 2015, Garces *et al.*, 2014, Scharf *et al.*, 2014, Ozorowski *et al.*, 2017). The solubility, stability and strong immunogenicity of 2dCD4, as compared to its full-length counterpart, have permitted and facilitated its widespread use. The majority of 2dCD4 mentioned in the literature has been produced as a soluble protein secreted using mammalian cell expression systems. *E. coli* expression of 2dCD4 results in the formation of insoluble inclusion bodies, which subsequently requires an extensive refolding procedure, in order to obtain soluble and functional proteins (Cerutti *et al.*, 2014a, Cerutti *et al.*, 2010). Despite being *E. coli* based, CFPS in this work was successful in expressing soluble, and as it will be later described (chapter 4), active protein. Consistent with *E. coli* expression of 2dCD4 (Cerutti *et al.*, 2014a, Cerutti *et al.*, 2010), soluble 2dCD4 production in CFPS was more favourable at lowered temperature. This is indicative of the need for slower translation and folding rates (Rosano and Ceccarelli, 2014, Vera *et al.*, 2007).

Similarly to SDF1-α, 2dCD4-wt was successfully produced in CFPS using deuterated amino acids. Using 2 mg/mL of deuterated amino acid approximately 2.5 mg of highly purified deuterated 2dCD4-wt was obtained from 50 mL of CFPS reaction (comparable to the yields when hydrogenated amino acid is used), and was used for complexing with the GP120 dimer. As reported by others, the over-expression of GP120 in mammalian cells results in the production of monomeric and disulphide-linked dimeric forms of GP120, which could be relatively easily separated from each other by using SEC (Finzi *et al.*, 2010, Guttman *et al.*, 2012). Here, due the limitations on the amounts of deuterated 2dCD4-wt and time constraints with approaching neutron beam time, CD4 was complexed with the more abundantly occurring GP120 dimer. The results of the subsequent SANS experiments will be given in chapter six.

## 3.5. CONCLUSIONS

Aimed towards enabling industrial and structural biology applications on a routine basis with consistent results and reliability, *E. coli* based CFPS was optimised for expression of various model proteins, and using different expression modes. A comparison of lysates derived from the various *E. coli* strains enabled the identification of conditions which improved the quantity and quality of cell-free expressed proteins.

In the case of CXCR4 expression, Rosetta lysate permitted a reduction in truncated proteins and enhanced the yields of the membrane protein. Whilst there still remains work to be done with regards to the overcoming the limited expression of CXCR4, the work has nevertheless laid the foundations for optimising and scaling-up CFPS for CXCR4 large-scale production. Indeed, up to 100 mL of batch CFPS reaction was subsequently setup for the successful soluble expression of normally insoluble-occurring proteins. To improve the expression of CXCR4 in CFPS, a systematic study into the use of CXCR4 coding sequences that avoid codons and secondary mRNA secondary structures which limit translation, is to be considered. To obtain higher amounts of CXCR4 currently, one could consider refolding the precipitated CXCR4 using detergents – although this procedure does not guarantee the structural and functional integrity of the protein. To avoid the issue of protein contamination with proteoliposomes, the re-insertion of detergent solubilised proteins into empty liposomes is to be investigated. To enable the study proteoliposome in its current un-purified state, electron microscopy, which was initiated (chapter four), could be a solution. Nevertheless, the problematic of impure proteoliposome samples was the

driving force towards the development of novel SANS methods for the structural analysis of CFPS proteins despite the presence of contaminating proteins.

In the case of SDF1-α and 2dCD4, it was found that the SHuffle lysate could enhance soluble protein expression and yields by favouring disulphide bond formation. This represents a critical advantage of CFPS over *in vivo* expression where both of these occur as insoluble, aggregated and non-functional proteins (Picciocchi *et al.*, 2014, Murphy *et al.*, 2007). CFPS thus avoids the laborious refolding procedure – a potentially wasteful step that does not necessarily guarantee high yields of soluble and functional proteins. Additionally, successfully being used for the production of deuterated SDF1-α and 2dCD4, CFPS can therefore be an effective system for the almost immediate expression of deuterated proteins. The speed and ease with which both labelled and unlabelled protein production that can be achieved when using CFPS, therefore sets the system as powerful and accessible tool for routine protein expression.

Following the initial small-scale optimisation, for purposes of structural biology and the requirement of large amount of proteins, CFPS was suitably used for the large scale production of both unlabelled and labelled SDF1-α and 2dCD4 (100 mL batches). These humble beginnings nevertheless represent the initial steps for increasing CFPS to industrial scales. Throughout the process of optimisation, standard operating protocols (SOPS) were written. In the context of quality control in industry, these can assist to standardise protocols, ensure reproducibility and ease in troubleshooting.

# CHAPTER FOUR

## BIOPHYSICAL AND FUNCTIONAL CHARACTERISATION OF CFPS PROTEINS

## ABSTRACT

Following the CFPS expression and purification of proteins, the proteoliposomes and proteins were then biophysically and functionally characterised using an array of methods. This is an important step in validation of the success of the CFPS process in producing viable proteins, but also in being able to provide insights into the CFPS process itself to permit longer-term routine tools for process and product quality control. In the context of the first objective, functional assessments using ELISA and SPR provided evidence for the structural integrity of CFPS-generated CD4 and CXCR4 by showing that they could bind either their ligand or a conformational antibody. Furthermore, chemotaxis assays using CFPS-generated SDF1-α variants, both labelled and unlabelled, served to compare and confirm the functionality of the proteins by showing that their chemotactic properties were maintained. Finally, mass-spectrometry analysis of SDF1-α established that CFPS, as used here, allowed the formation of correctly paired disulphide bonds which are important for the structural integrity of the protein.

However, in the context of the latter objectives the characterisation of proteoliposomes remains challenging due to their heterogeneity and complexity, which can also limit their eventual purification. Therefore, with the view to improving and standardising the homogeneity of CFPS products, the size and size distribution of liposomes/proteoliposomes was studied. Firstly, the capability of a NanoSight™ particle tracker in monitoring the changes in the size distribution of liposomes/CXCR4-proteoliposomes was assessed as a potential routine quality control system. It was shown that the small, uniformly-sized liposomes (<100 nm diameter) introduced at the start of the CFPS reaction undergo an increase in size. This is likely due to the insertion of CXCR4 into the liposomes during the formation of proteoliposomes. Secondly, to obtain a visual depiction, proteoliposomes were assessed using negative transmission electron microscopy (TEM). However, due to the fragility of proteoliposomes, these were destroyed during TEM making interpretation difficult. Nevertheless, the TEM studies served as an initial step towards more the advanced cryo-electron microscopy techniques, which are less destructive to the samples.

These techniques, particularly in the context of the future routine application of the Synthelis CFPS technology, reinforce the confidence in the CFPS products and in moving towards standardising quality control methodology using simple techniques.

## 4.1. INTRODUCTION

The validation of the CFPS protein products prepared as part of this thesis is a key step in demonstrating that the proteins are viable and equal to the native protein or protein produced using traditional methods. Equally important is the ability to apply such quality control procedures to CFPS procedures in a routine way, as well as providing biophysical insights to the CFPS process itself in order to improve and enhance its efficiency. This chapter therefore describes the application of an array of biophysical and functional tests applied to the CFPS products and the process itself.

### 4.1.1. The need for easy-to-implement quality control methods for biophysical and functional characterisation of CFPS proteins

In the previous chapter CFPS was optimised for the expression of various proteins: CXCR4, CD4 and SDF1-α. However, these CFPS products need to be shown to be structurally and functionally sound. Membrane proteins in proteoliposomes expressed by CFPS methods have been shown to be functionally active (Soranzo *et al.*, 2016, Soranzo *et al.*, 2017, Deniaud *et al.*, 2010, Cortes *et al.*, 2018, Liguori *et al.*, 2015, Chadli *et al.*, 2017). However, in the context of industry, particularly for small and medium enterprises, these validation methods should ideally be economical, accessible, enable high-throughput processing and, yet, remain accurate. The implementation of these methods in routine CFPS tasks would provide added value to the final protein product. In line with quality control concerns, the application of these methods would furthermore ensure reproducibility, and enable the identification of deviations in CFPS and the expressed proteins.

However, liposomes and proteoliposomes are highly complex and dynamic structures making their characterisation challenging. Their structures are influenced by multiple extrinsic and intrinsic factors (White *et al.*, 2000). For instance, high NaCl concentrations have been found to cause liposome aggregation, and alteration of the constitution of the phospholipid bilayer of the membrane can affect the orientation of the insertion of membrane proteins (Hickey and Buhr, 2011). However, most of the studies upon liposomes considered reconstitution only of purified proteins into proteoliposomes (White *et al.*, 2000). Here though, proteoliposomes derived from the results of CFPS are highly heterogeneous and carry a relatively large amount of baggage in the form of contaminating proteins, as well as the

target protein of interest. The difficulty of the task is further increased with the need for the large amounts of pure sample required for structural studies. The crystallisation of CXCR4-lys, for instance, requires at least 95 % pure sample and at exceptionally high concentrations of about 80 mg/mL (Wu *et al.*, 2010). For this purpose, insect-cell expressed CXCR4-lys had to be removed from its lipid environment and maintained by detergents instead. Unfortunately here, the low levels of CXCR4 expression using CFPS, and the large amount of contaminating proteins which cling to the proteoliposome make structural analyses of CFPS CXCR4-lys very difficult. Whilst the purification of proteoliposomes remains challenging, methods for assessing the biophysical characteristics of the proteoliposomes, despite low purity, are nevertheless critical.

### 4.1.2. Tools used for characterisation of proteoliposomes and soluble CFPS proteins

This chapter focuses upon the establishment of potentially routine methods for the biophysical characterisation and functional assessment the CFPS-generated membrane protein embedded within the proteoliposomes and the soluble CFPS proteins, and of the (proteo)liposomes themselves. Descriptions of the methods used are provided below.

#### *ELISA*

ELISAs (enzyme-linked immunosorbent assays) are a plate-based assay technique used for detecting and quantifying biological substances (Figure 4.1). ELISAs are typically carried out in chemically treated 96-well polystyrene plates which allow biological substances to passively adhere to the bottom of wells. The biological substance of concern, referred to as the antigen, is immobilised at the bottom on the solid surface of the well, forming a “coat”. Detection is then performed by assessing the activity of the conjugated enzyme. This is carried out by the incubation of a substrate which is converted to a quantifiable product. This typically involves a visible colour change which can be quantified using spectrophometry methods. The most commonly used enzyme labels are horseradish peroxidase (HRP) and alkaline phosphatase (AP), for which a large collection of substrates are commercially available.

In direct detection ELISA, the coated antigen is then complexed to a specific antibody that is linked to an enzyme. In indirect detection ELISA, the use of a primary antibody specific to the immobilised antigen is

firstly employed, followed by the use of a secondary antibody, which is targeted towards the primary antibody. The secondary antibody is linked to the enzyme which then allows detection.

Here, ELISAs were used to confirm that ligands and conformational-specific antibodies could bind to CFPS-generated CXCR4 and full-length CD4 embedded in proteoliposomes, with the results providing validation of the structural and functional integrity of those proteins. Furthermore, in the context of industry, the development of ELISA protocols could also provide a foundation to establish SOPs for routine ELISA as an affordable technique for validating successful expression of CFPS-generated proteins.

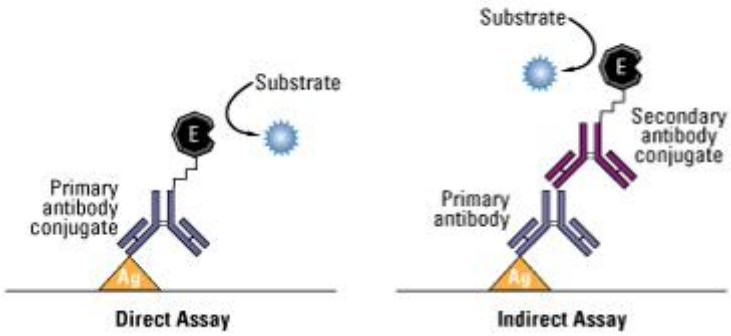



***Figure 4.1:*** *Schematic illustrating the basic principle of ELISAs. Left: direct ELISA setup, right: indirect ELISA setup (https://www.thermofisher.com/au/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html).*

### *Surface plasmon resonance*

Surface plasmon resonance (SPR) is a phenomenon that occurs when single wavelength polarised light strikes an electrically conducting surface (typically gold) at the interface between two media (Figure 4.2). The electrons in the conducting surface undergo resonant oscillation leading to the generation of electron charge density waves, called plasmons. With the formation of the plasmons, a quantum of energy is absorbed from the incident light. The reflected light, missing the specific energy quantum, is then detected. The SPR system makes use of the surface plasmon resonance angle as a measure for the missing quantum of energy. In its application in biology, the quantum of energy absorbed by the plasmons, and thus the surface plasmon resonance angle, is proportional to the molecular weight of the

biomolecules located at the surface of the conducting interface. These immobilised biomolecules located on the conducting surface are referred to as the ligands. Additionally, the levels of energy absorbed are highly sensitive allowing detection of small changes in the molecular weights of the ligands. By recording the changes in the surface plasmon resonance angle, the SPR technique is therefore able to detect the association and disassociation of molecules to the ligands. Typically, the molecules which can bind to the ligands, referred to as the analyte, are in a solution which is flown at a constant rate across the ligands.

Here, SPR was used to assess the binding of CXCR4-specific antibodies to CFPS CXCR4 embedded in proteoliposomes. Compared to ELISAs, where results provided are indicative of binding or non-binding, SPR is capable of giving detailed quantitative information on the binding and unbinding of ligands and analytes. Given the complexity of proteoliposomes however, further work in the preparation of samples and the SPR protocols are needed to enable the determination of quantitative results. In the context of industry, SPR could be used for determining the binding capabilities of CFPS products, and serving as more detailed quality control checks.

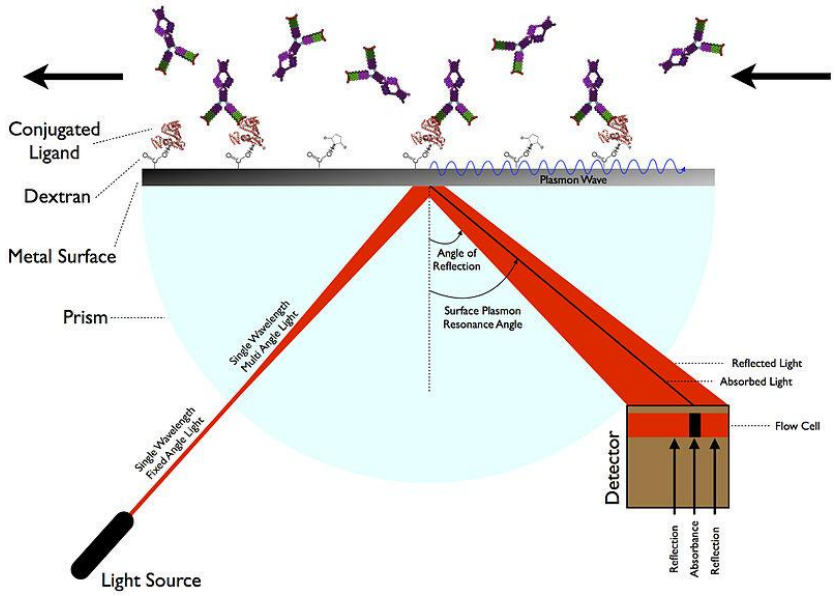



***Figure 4.2:*** *Schematic illustrating the principle of SPR (Sarri Sabban et al. 2011).*

### *Mass spectrometry*

Mass spectrometry (MS) is powerful biophysical method for accurately determining the molecular weights and composition of samples, which include proteins (Figure 4.3). In MS the particles present in the sample are ionised. The ionised particles, carrying a specific charge, are then accelerated towards a detector through a magnetic or electrical field. The field causes ionised particles to deviate and separate from each other during the acceleration phase. As a result, the ionised particles arrive at different times at the detector, which is capable of distinguishing the time-of-flight (TOF) of the ionised particles. Provided that all of the ionised particles underwent acceleration with an equal amount of energy, the TOF of each particle is proportional to their mass-to-charge ratio (m/z). Thus, ionised particles with higher m/z ratios have longer TOFs than those with low m/z ratios. With the use of calibrated instruments, MS permits the identification of the ionised particles by correlating the m/z of the detected signature profile to known and expected particles.

Here, the purified CFPS SDF1-α variants, as described in chapter three, were analysed with MS in order to confirm, firstly, their molecular weight as a quality control check. Secondly, MS can confirm the presence of paired disulphide bridges in CFPS SDF1-α. Finally, and with regards to the application of SANS using match-out labelling, MS is a critical tool to determine the level of deuterium incorporation in CFPS deuterium-labelled SDF1-α.

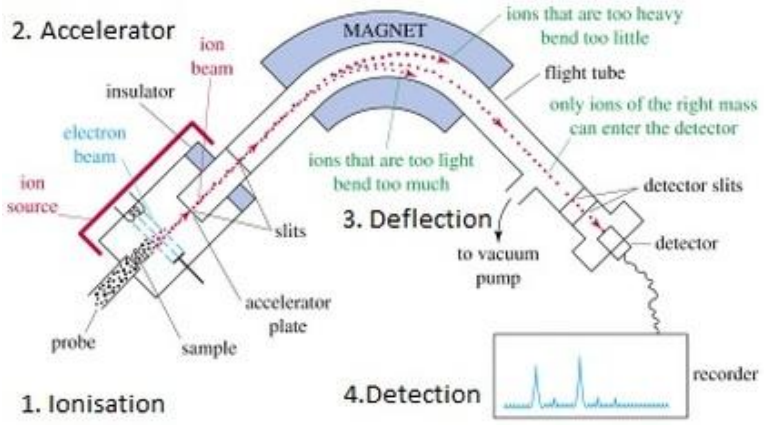



***Figure 4.3:*** *Schematic illustrating the principle of MS (http://chemicalinstrumentation.weebly.com/mass-spectrometry.html).*

### *Chemotaxis assay*

Chemotaxis assays are experimental tools used for assessing the chemotactic ability of cells towards a potential chemotactic agent. Chemotaxis assays can therefore provide evidence of functionality of the chemotactic agent. The chemotaxis assays used here involve the placement of a defined number of motile cells into an upper chamber and the test substance in the lower chambers of a 96-well Transwell system (Figure 4.4). The upper and lower chambers are separated by a filter of defined pore size. In the event of chemotaxis, the motile cells from the upper chamber actively migrate through the pores of the filter in order to enter the lower chamber containing the chemokine. The number of migrated cells is counted and compared, using negative controls to provide a baseline measurement.

Chemotaxis assays were used to confirm the functionality of CFPS SDF1-α variants with respect to their chemotactic properties.

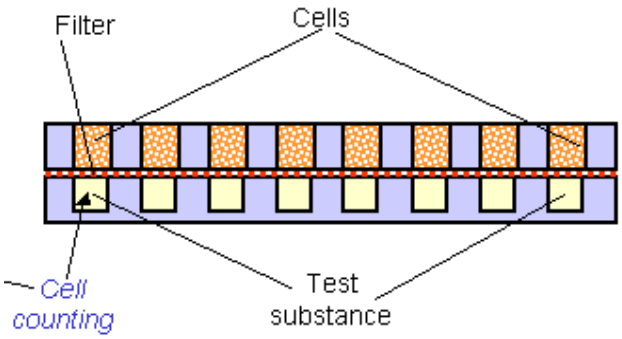



***Figure 4.4:*** *Schematic illustrating the principle of the chemotaxis assay using Transwell system (Kohidai, L., 2006, https://commons.wikimedia.org/wiki/File:Chtx-twochamber-techn.PNG).*

### *The NanoSight™ Nanoparticle Tracker*

The NanoSight™ (Malvern) uses Nanotracking Tracking Analysis (NTA) technology to visualise and analyse particles (of between 10 – 2000 nm diameter) in flowing liquids. NTA exploits light scattering and Brownian motion in order to obtain the size distribution of particles (Figure 4.5).

The sample chamber is illuminated by a laser. The particles in suspension in the flowing liquid which encounter the laser beam result in light scatter. The scattered light, and therefore the presence of the particles, are detected *in situ* by a video camera and visualised on a screen. Recording and tracking the movement of the particles, the NTA software uses the Stokes-Einstein equation and automatically calculates the hydrodynamic diameters of the particles. These are then recorded to produce size distribution charts at the end of the analysis. Incidentally, the concentration and aggregation levels of the samples are also recorded.

Here, with the view towards standardising and improving the homogeneity of samples, the NanoSight™ nanoparticle tracker was used to monitor the changes the sizes and distribution of sizes in the liposomes/proteoliposomes population during the CXCR4 CFPS reaction. Furthermore, with the need and challenge of characterising liposomes and proteoliposomes, the capability of the NanoSight™ nanoparticle tracker as an easy-to-use tool for routine application, was assessed

.

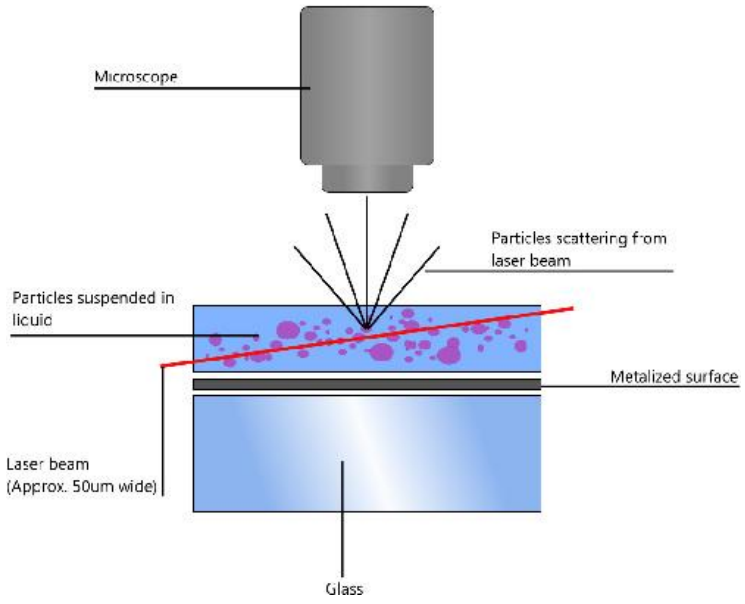



***Figure 4.5:*** *Schematic of the operation mode of NanoSight™ (Malvern) (https://www.malvernpanalytical.com/en/products/technology/nanoparticle-tracking-analysis).*

### *Transmission Electron Microscopy (TEM)*

Transmission electron microscopy (TEM) employs a beam of high-energy electrons that is transmitted through a specimen of interest in order to generate an image. The image generated is a result of the interaction of the electron beam with electrons found in the specimen. Electron-dense region of the specimen absorb energy from the incident electron beam, whereas low electron density region allow the beam to be transmitted. Biological specimens typically interact very weakly with the incident electron beam due to overall low electron density. These therefore require staining with a heavy metal ion solution in order "enrich" the specimen with electrons. The image is then detected on an adapted sensor (photographic film or camera) as a distribution of dark and light regions respectively corresponding to high-electron density and low-electron density regions. Operating on the same principle as optical microscopy, the use of the electron beams, which have a very short wavelength, enables high resolution (up to c. 0.2 nm) details to be observed in the recorded images.

TEM was considered here to obtain visual insights into the proteoliposomes, and as a first step towards more advanced cryo-EM techniques.

## 4.2. MATERIALS AND METHODS

### 4.2.1. ELISAs and surface plasmon resonance setups for antibody binding to CXCR4 and CD4 in proteoliposomes

CFPS reagents, including lysate, were shipped from Synthelis (France) to the University of Witwatersrand (South Africa) on dry-ice. Liposomes were reconstituted and CFPS reactions were carried out as described previously for the production of CXCR4-wt in proteoliposomes. An identical CFPS reaction, but excluding DNA, was prepared for the production of negative control proteoliposomes. CXCR4-wt, detected by anti-histidine western blot analysis, was used to confirm expression. The proteoliposomes were recovered by centrifugation, as described previously in chapter three, and resuspended into sterile 1 X PBS pH 7.4. Stocks of CXCR4-wt at 125 µg/mL in proteoliposome were prepared and used for subsequent experiments. Negative controls were resuspended in equal volumes to that used for resuspending CXCR4-wt proteoliposomes.

ELISA was used to confirm the binding of antibodies to CXCR4-wt in proteoliposomes and to provide estimates for the working concentrations of CXCR4-wt in proteoliposomes and antibodies to be used for SPR. CXCR4-proteoliposomes at 1 µg/ml and negative control proteoliposome at equivalent concentration in 1 X PBS pH 7.4 were added in duplicates into a Nunc MaxiSorp® flat-bottom plate 96-well plate. Also in duplicates, blank controls, where no proteoliposomes were coated, were also included. Following blocking, polyclonal anti-CXCR4 antibodies from the NIH Reagent Programme (www.aidsreagent.org) directed against either the N-terminal (anti-CXCR4 NT, cat. no. 11237), or the extracellular loop (anti-CXCR4 ECL, cat. No. 11236) of CXCR4, were added at 10 ug/mL. Secondary HRP-linked antibodies were added and detection with TMB-Ultra substrate (Thermo-Pierce, U.S.A) was carried out.

A BioRad ProteOn™ XPR36 instrument, general-use GLM sensor chip, equipment and consumables and the manufacturer's protocol were used to carry out SPR of antibodies against CFPS expressed CXCR4-wt in proteoliposomes. The ProteOn™ hydrochloric acid (HCl), ProteOn™ sodium hydroxide (NaOH), ProteOn™ sodium dodecyl sulphate (SDS), ProteOn™ immobilisation buffer kit, ProteOn™ ethanolamine-HCI and GLM chip, stored at 4 °C, were allowed to equilibrate at room temperature. 2 L of the two flow buffers, 1 X PBS pH 7.4 (PBS) and 1 X PBS pH 7.4 with 0.005 % Tween-20 (PBS-T), were prepared and sterile filtered (0.22 µm filter). The room temperature chip was inserted into the ProteOn™ XPR36 instrument and the standard air initialisation protocol was performed. Once air initialisation was completed, the standard chip conditioning protocol was carried out. This comprised of sequential washes with SDS, NaOH and HCl injections horizontally (30 µL/minute flow rate x 60 seconds contact time = 30 µL for each). The procedure was repeated for injections vertically. Once chip conditioning was completed, standard quality control flow buffer injection protocol, using PBS, was carried out for injections horizontally (4 X, at 100 µL/minute flow rate x 60 seconds contact time = 100 µL volume ; 0 seconds dissociation time). The procedure was repeated for injections vertically.

Stored at –20 $^{o}$C, the ProteOn™ amine coupling kit, containing EDAC (1-ethyl-3-(3-diethylaminopropyl)carbodiimide hydrochloride) and sulpho-NHS (N-hydroxysulphosuccinimide), was reconstituted according the manufacturer's protocol. EDAC (0.575 g) was dissolved in 75 ml of cold distilled water for a final concentration of 40 mM. 75 ml cold distilled water was added to ProteOn™ sulpho-NHS (0.163 g) to dissolve it to the final concentration of 10 mM. In aliquots of 1 mL, fresh activation buffer was prepared by mixing EDAC and sulpho-NHS in a ratio of 1:1 and used immediately. 150 µL of the activation solution was applied to a GLM chip at 30 µL/minute.

Anti-CXCR4 NT, anti-CXCR4 ECL, anti-histidine and 12G5 antibodies were immobilised on the activated SPR chip. Prepared at 10 ug/mL concentration in acetate buffer (pH 4.5), the antibodies were injected onto four of the six lanes present on the SPR chip horizontally (30 µL/minute flow rate x 300 seconds contact time = 150 µL). A fifth lane served as negative control with only the acetate buffer and without antibody flowed through. Once immobilisation was complete, the un-reacted NHS groups were deactivated by injecting 1 M ethanolamine-HCl vertically (30 µL/minute flow rate x 300 seconds contact time = 150 µL). PBS-T was flowed across the chip (30 µL/minute) for at least 30 minute, or until signal stabilisation.

The sample of CXCR4 in proteoliposomes was prepared by diluting the proteoliposomes to 1 ug/mL of CXCR4 in PBS-T. An equal volume of buffer was used to prepare the negative proteoliposome control. The CXCR4 sample and negative control proteoliposomes were flowed on to the chip (30 µL/minute flow rate x 300 seconds contact time = 150 µL) and the association (binding) phase signal was detected. PBS-T buffer was then flowed on to the chip and the dissociation (unbinding) phase signal was observed. To confirm CXCR4 binding to antibodies, the SPR curves were normalised and subtracted from the negative control signal.

### 4.2.2. ELISA setup for GP120 binding CD4 in proteoliposomes

In a Nunc MaxiSorp® flat-bottom 96 well-plate, 100 µl of lectin in PBS at 2 µg/ml was added per well. The lectin was incubated overnight at 4 $^{o}$C. The contents of the wells were aspirated and 200 µl of blocking solution (1 % BSA in PBS-T) was added to the wells. The plate was incubated for at least 1 hour at room temperature.

In the meantime, the GP120-CD4 complex was prepared. Stocks of full-length CD4 prepared using CFPS in proteoliposomes were prepared and estimated to be at a concentration of 100 µg/ml. Purified GP120 at a final concentration of 1 µg/ml in 1 X PBS was combined with full-length CD4 in proteoliposome, giving final CD4 concentrations of 1.0, 0.50 and 0.25 µg/ml.

The plate was washed five times using PBS-T. In duplicates, 100 µL of complex was added to each well. Controls, as described below, were included.

The plate was incubated for 1 hour at room temperature and then washed five times using PBS-T. 100 µl of the primary antibodies, 17b and ibalizumab (produced and provided by HIVPRL), were added to the wells at 0.03 µg/ml. 17b is an antibody directed against CD4-induced binding site of GP120 whereas ibalizumab is directed against the interface between domains one and two of CD4. Following 1 hour incubation at room temperature, the plate was washed five times with PBS-T. At 1:2000 dilution in PBS-T, 100 µL of secondary anti-human HRP-linked peroxidase (Amersham, GE healthcare) was added to the wells. Following 1 hour incubation at room temperature in the dark, the plate was washed five times with PBS-T. 100 µl of 1-Step$^{TM}$ Ultra TMB-ELISA Substrate Solution was added to the wells. The plate was incubated for at least 30 minutes in the dark at 37 $^{o}$C on a rotary plate at 300 rpm, or until adequate blue colouration appeared. The reaction was stopped by the addition of 50 µl of 0.1 M sulphuric acid. The plate was read using a plate reader (Bio-Rad Model 650 Microplate reader, CA, U.S.A; set at fast read, 450 nm, measure filter, 37 $^{o}$C).

### 4.2.3. Characterisation of purified SDF1

#### *Mass spectrometry of purified SDF1*

Mass spectrometry measurements were carried using a LC ESI-TOF MS (6210, Agilent Technologies) with an in-line HPLC binary pump system (Agilent Technologies). The system was calibrated in the mass-to-charge (m/z) 300-3,000 range with standard calibrants (ESI-L, low concentration tuning mix, Agilent Technologies) before measurements and mass spectra were recorded in the 300-3,200 m/z range.

20 µL of purified NHis-SDF1 (> 99 % purity), of which the CFPS expression and purification are described in chapter three, were diluted in acidic denaturing conditions to a final concentration of 4 µM with solution A (0.03 % trifluoroacetic acid [TFA] in water, Acros Organics). The temperature of the sample was adjusted to 10 $^{o}$C and the analysis was run by injecting 4 µL of each sample. They were first trapped and desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5 µm, 300 µm ID X 5 mm, Agilent Technologies) for 3 minutes at a flow rate of 50 µL/minute with 100 % solution A and then eluted with 70 % solution B (95 % acetonitrile, 5 % water, 0.03 % TFA) at a flow rate of 50 µL/minute for MS detection. The RP-C8 cartridge was then re-equilibrated for 4 minutes with 100 % solvent A at a flow rate of 50 µL/min.

Mass spectrometry acquisition was carried out in the positive ion mode with spectra in the profile mode. The instrument was operated with the following experimental settings: ESI source temperature was set at 300 $^{o}$C; nitrogen was used as drying gas (7 L/minute) and as nebuliser gas (10 psi); the capillary needle voltage was set at 4,000 V. The spectra acquisition rate was 1.03 spectra/second. All solvents used were HPLC grade (Chromasolv, Sigma [apart from TFA, as mentioned previously]). The spectra were acquired and the data processed with the MassHunter workstation software (v. B.02.00, Agilent Technologies) and with the GPMAW software (v. 7.00b2, Lighthouse Data, Denmark).

### *Chemotaxis assay of purified SDF1*

The migration of Jurkat T-cells in response to both labelled and unlabelled SDF1 was evaluated using a Transwell system (Corning, Sigma), as described previously (Connell *et al.*, 2016). Briefly, the cells were labelled with the intracellular fluorescent dye calcein-AM (BD Biosciences) at 1 µM in PBS, 0.1 % BSA for 30 minutes at 37 °C. Following labelling, the cells were washed twice with PBS, and $2 \times 10^5$ labelled cells in 100 µl of RPMI medium supplemented with 20 mM HEPES and 1 % human AB serum were added to the upper chamber of a 96-well, 5 µm pore diameter polycarbonate Transwell culture insert. The same medium (250 µl) containing 0 to 300 nM SDF1-α was placed in the lower chamber. After 3 hours at 37 °C in humidified air with 5 % $CO_2$, the fluorescence of cells that had migrated through the filter to the lower wells was measured using a fluorescence plate reader (Wallac Victor, PerkinElmer) with a 485 nm excitation filter and a 535 nm emission filter. The percentage of cells that migrated across the polycarbonate membrane was calculated as follows: {[(number of cells migrating to the lower chamber in response to chemokine) − (number of cells migrating spontaneously)]/number of cells added to the upper chamber at the start of the assay} × 100.

### 4.2.4. Characterisation of CXCR4 and CD4 in proteoliposomes and CD4-GP120 complex

#### *NanoSight™ particle tracker: assessment of liposomes to proteoliposomes during CXCR4 CFPS*

CFPS of CXCR4-wt in proteoliposomes was carried out as described in chapter four. Aliquots of the reaction were taken across the various stages of production and purification, as described in Table 4.1. The aliquots were stored on ice (for between 12 – 16 hours) following production, during transportation to the NanoSight platform and during measurement. To characterise the changes of liposomes and proteoliposomes during CFPS of CXCR4 a Malvern Panalytical NanoSight™ NS300 Instrument was used. The samples were diluted in high-grade distilled water at 1:600,000 for analysis. Three runs were made per sample and the resulting data were merged.

| Sample | Description of sample |
|---|---|
| **Liposomes** | Stock liposomes at 30 mg/mL added into CFPS reaction mix |
| **Lysate** | Stock of SHuffle lysate |
| **T = 0 s CFPS** | Freshly prepared reaction mix upon the addition of liposomes, excluding plasmid DNA required for CFPS reaction |
| **Negative Reaction** | CFPS mix with lipsosomes, excluding plasmid DNA required for CFPS reaction, incubated overnight at 30 $^{o}$C |
| **Negative Proteoliposomes** | Negative reaction proteoliposomes purified by sucrose-gradient and centrifugation washes, and resuspended into equal reaction volume of 50 mM HEPES pH 7.5 |
| **CXCR4-wt Reaction** | CFPS reaction with liposomes of CXCR4-wt,incubated overnight at 30 $^{o}$C |
| **CXCR4 Proteoliposomes** | CXCR4-wt reaction proteoliposomes recovered by sucrose-gradient and centrifugation washes, resuspended into equal reaction volume of 50 mM HEPES pH 7.5 |

**Table 4.1:** *Description of* CFPS *samples used for NanoSight™ analysis covering various stages of the CFPS procedure and necessary negative control samples.*

### *Transmission electron microscopy: visual assessment of CXCR4/CD4 proteoliposomes*

In order to obtain a visual description of the morphology of the proteoliposomes, transmission electron microscopy was used to assess CXCR4-wt and full-length CD4 embedded in proteoliposomes as produced by CFPS. CFPS CXCR4-wt or full-length CD4 in proteoliposomes were expressed as described in the previous chapter. The proteoliposomes were purified by sucrose-gradient and ultracentrifugation, centrifugation washes and resuspended into 50 mM HEPES pH 7.5 buffer at 1 X concentration. Additionally, a negative 30 $^{o}$C proteoliposome control, resuspended in an equal volume of 50 mM HEPES pH 7.5 buffer, was also assessed. The negative control was produced using standard liposome CFPS, but excluding DNA from the reaction mixture.

The proteoliposomes were diluted with 50 mM HEPES pH 7.5 buffer at dilutions of 1:100 and 1:10. For TEM, the negative stain-on-grid (SOG) technique was used: 8 µL of the diluted samples were added to an agar carbon grid for 2 minutes and the grid was stained with 50 µL of uranyl acetate (AcUr; $UO_2(CH_3COO)_2 \cdot 2H_2O$ at 2 % in distilled water (pH 4.2 – 4.5) or phosphotungstate acid (PTA; $H_3PW_{12}O_{40}$ at 1.5 % in distilled water) for 1 minute. The excess solution was soaked off by a filter paper and the grid was air-dried. Images were taken under low dose conditions (<10 $e^-/Å^2$) with defocus values between 1.2 and 2.5 µm on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using a Gatan Orius 1000 CCD camera.

## 4.3. RESULTS

### 4.3.1. ELISA and SPR of CXCR4 in proteoliposome

The aim of the ELISA and SPR experiments was to confirm the binding of the CFPS CXCR4 in proteoliposomes to anti-CXCR4 antibodies. The binding of conformational antibodies would provide evidence for the correct folding of the CXCR4 CFPS product.

ELISA demonstrated the binding of anti-CXCR4 antibodies to CFPS CXCR4 expressed in proteoliposomes (Figure 4.6). There was, as expected, a relatively large amount of contaminants present which led to a

significant amount of non-specific binding. The ELISA results allowed a rough estimation of the working concentrations of proteoliposomes and antibodies for the subsequent SPR analysis.

These results were confirmed using SPR analysis (Figure 4.7). When normalised to negative control, CXCR4 in proteoliposomes was confirmed to bind to its corresponding antibodies. Interestingly, the binding response was the same for all of the antibodies. Previous SPR studies, using detergent-solubilised mammalian cell-expressed CXCR4, have reported the loss of nearly half of the response when the 12G5 antibody was used instead of an antibody targeting the tag at the N-terminal (Kofuku *et al.*, 2009). As a conformational-dependent binding antibody (Baribaud *et al.*, 2001), the binding of 12G5 with equal response as the other conformational-independent antibodies used here, provides a strong indication that the CFPS produced CXCR4 embedded in proteoliposomes is correctly folded.

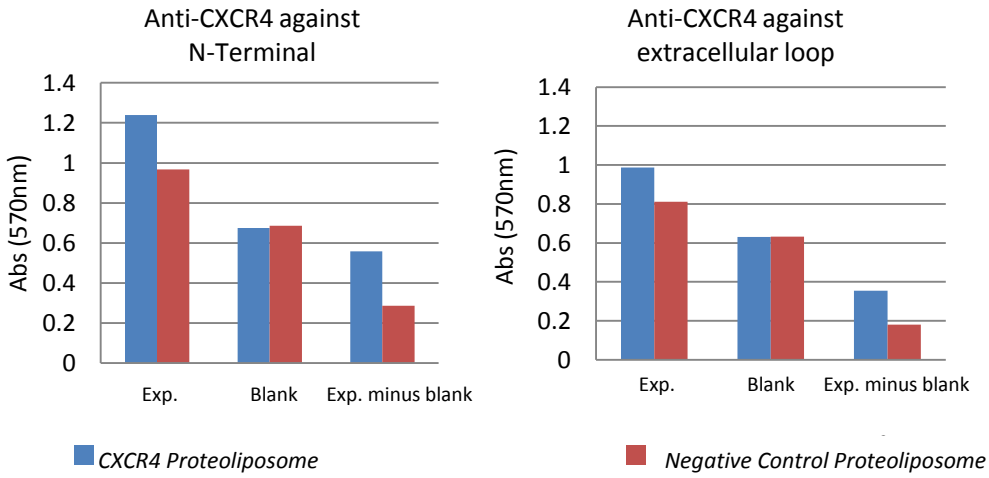



***Figure 4.6:*** *ELISA responses (average of duplicates) of two anti-CXCR4 antibodies against CFPS CXCR4 in proteoliposomes and negative control (Exp). Despite large amount of non-specific response, a higher response was observed for CXCR4-proteoliposome compared to negative proteoliposome controls. Also measured, was the average (duplicates) of ELISA response for blank controls, where no proteoliposome were added.*

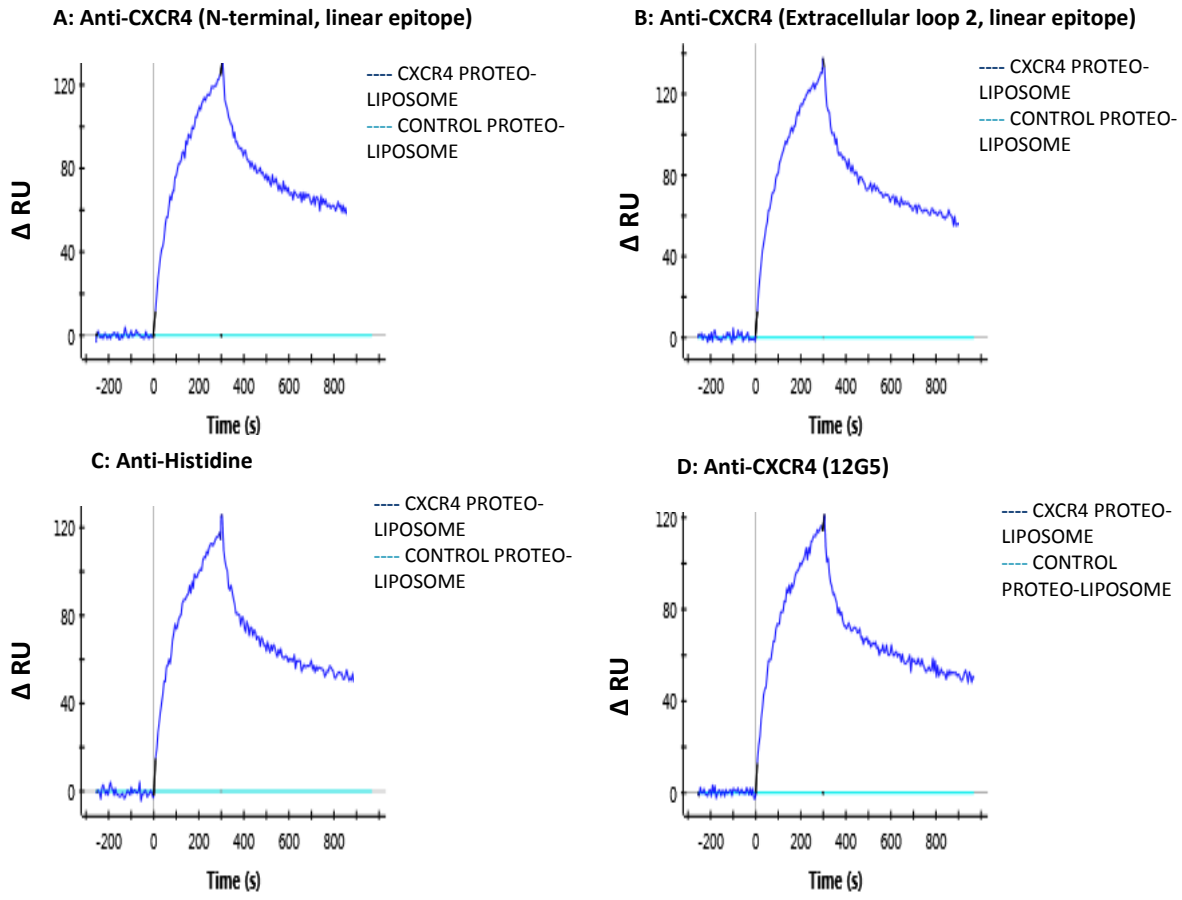



***Figure 4.7:*** *SPR binding assay (difference in response unit (ΔRU) against time (s)) of antibody binding to CXCR4 proteoliposomes (blue) relative and normalised to negative control (cyan). With occurrence of the association-dissociation signal, CXCR4-proteoliposome, as the analyte, was indicated to bind to the immobilised anti-CXCR4 antibodies ligands on the SPR chip. A, B and C show the association-dissociation signal of antibodies directed towards respectively the N-terminal, second extracellular loop and histidine tag of CXCR4. D confirms the binding of CXCR4 to conformational antibody 12G5, which indicates that CXCR4 is folded correctly.*

### 4.3.2. ELISA of full-length CD4 in proteoliposomes

The aim of these ELISA assays was to confirm the binding of the CFPS expressed CD4 in proteoliposomes to anti-CD4 antibodies. The binding of conformational antibodies would provide evidence for the correct folding of the CD4 protein.

Ibalizumab is a conformational antibody that is directed against the interface between the first and the second domain of CD4 (Song *et al.*, 2010, Freeman *et al.*, 2010, Burkly *et al.*, 1992). In this work, it was firstly confirmed to bind to the CFPS expressed full-length CD4, which carries the transmembrane domain, in proteoliposomes. It should be noted that ibalizumab does not inhibit the binding of CD4 to major histocompatibility complex (MHC) nor to GP120, but can hinder HIV infection (Burkly *et al.*, 1992, Moore *et al.*, 1992). The mode of action of ibalizumab is suspected to impose steric restrictions on the conformational changes required in either CD4 or GP120 and that ultimately lead to viral entry. Secondly, the binding of ibalizumab to CD4 provided a method to estimate the relative amount of active CD4 present in the sample. Indeed, estimating the amount of expressed protein in proteoliposomes whilst using Coomassie-blue stained PAGE gel is highly imprecise due the large amount of contaminating proteins. Whilst the yield was estimated to be about 100 µg/mL using a Coomassie-stained gel, the sensitivity of ELISA indicated that there was approximately 1.5 times the amount of CD4-FL-CHis as compared to CD4-FL-NHis in the samples used. However, it is very likely that ibalizumab is also binding to truncated species of CD4 which include the second domain and therefore not only quantifying completely expressed CD4.

To further assure the conformational integrity and functionality of CD4 prepared by CFPS, the binding of GP120 to CD4 was confirmed using the anti-GP120 17b antibody binding. Upon binding to CD4, GP120 undergoes conformational changes which lead to the exposure of CD4-binding site induced (CD4i) epitopes. 17b is a conformational antibody that binds to one of the CD4i epitopes, which is not available in native GP120. It should be mentioned that the ELISA setup using the lectin coat (Figure 4.8A) was used as it allows the binding of GP120 in a consistent orientation which exposes its 17b binding site.

Ibalizumab detection showed a higher response compared to that with the 17b antibody (Figure 4.8B). This can be expected as ibalizumab will bind to all CD4 molecules which have bound to GP120, or which have not formed a complex with GP120, whereas 17b will only bind to GP120 molecules which have bound to CD4.

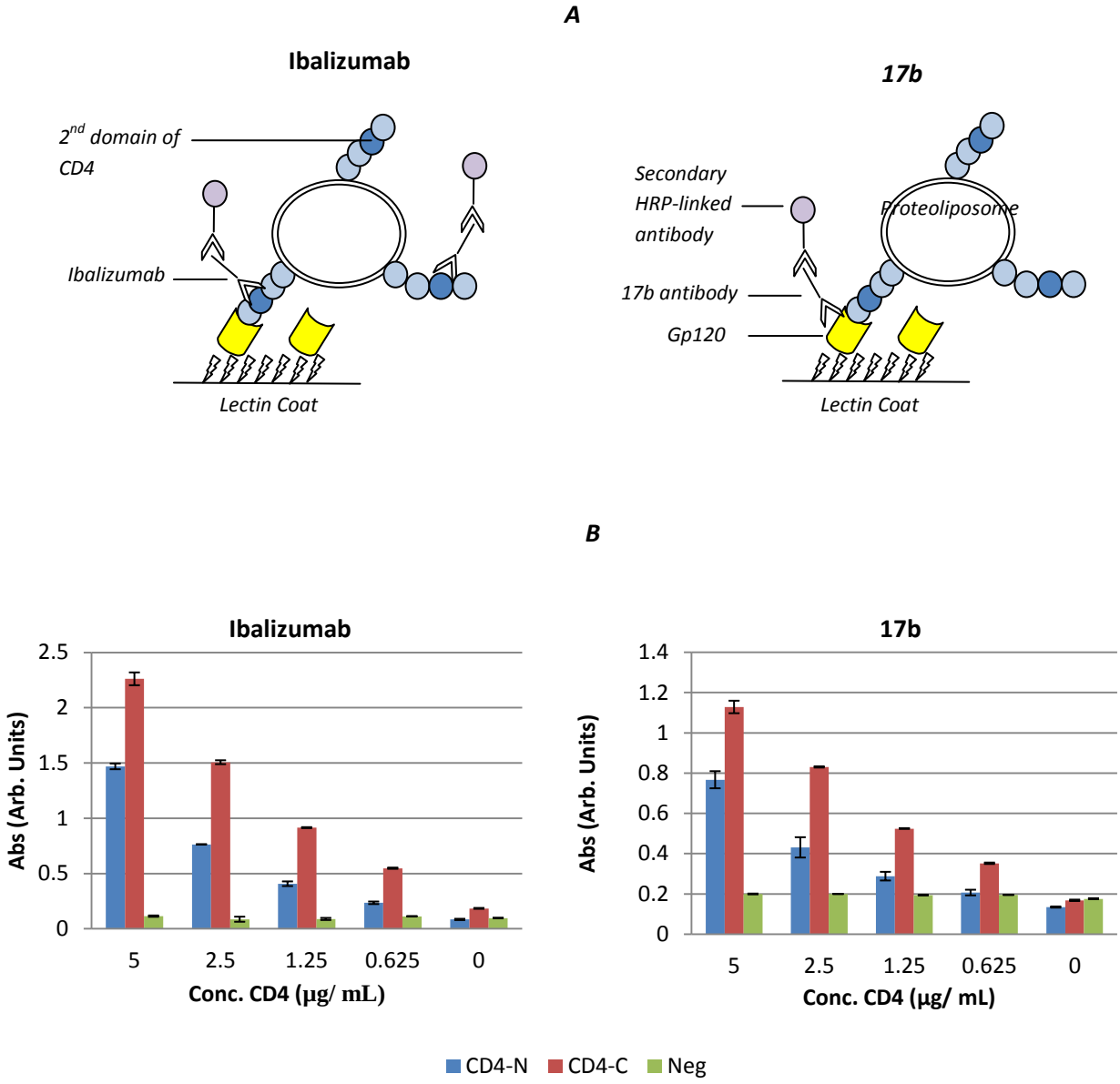


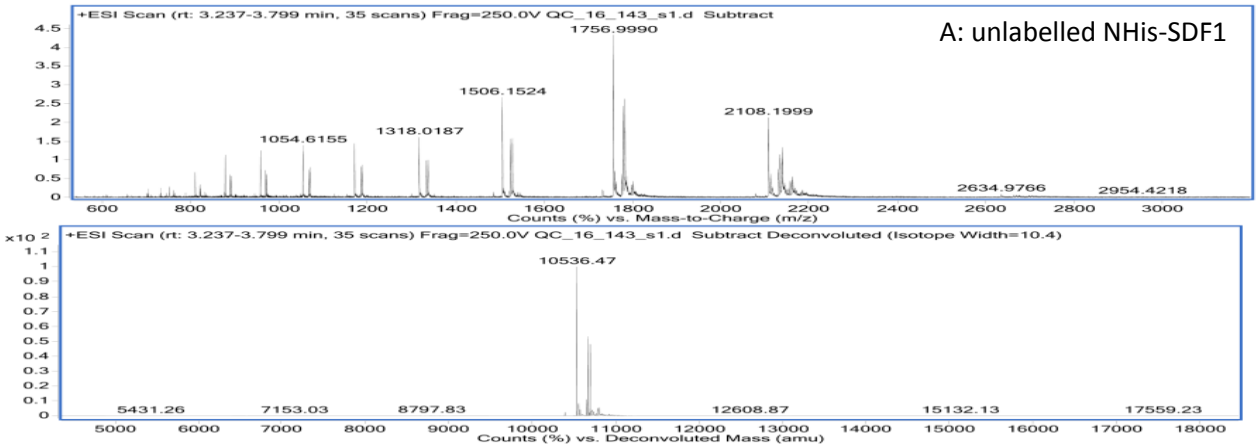

*Figure 4.8: Lectin coat ELISA setup for CFPS expressed CD4 in proteoliposome is illustrated (A). The lectin coat provides allows the capture of GP120 in an orientation that allows exposure its 17b binding site. ELISA results show dose-dependent responses which confirm the presence and binding of CD4 to GP120 (B).*

### 4.3.3. Mass spectrometry of unlabelled and deuterated SDF1-α

Mass-spectrometry was carried out on unlabelled and deuterated NHis-SDF1 (deuterated NHis-SDF1 produced at various deuterated levels – see section 3.27). The aim of the mass spectrometry analyses of SDF1-α was to firstly as a quality control check to verify that the protein was being translated to completion by confirming that purified CFPS SDF1-α had the correct molecular weights. Secondly, it was used confirm the presence of paired disulphide bridges in CFPS SDF1-α. Thirdly, since the labelled protein was to be used in SANS, mass spectrometry was used to determine the level of deuterium incorporation in the various deuterium labelled SDF1-α, and therefore confirm which amino acid ratio was suitable for producing deuterated SDF1-α for matching-out experiments in SANS (i.e. 75 % deuteration level).

According to ProtParam (ExPASy, www.expasy.org/tools/protparam.html) the theoretical molecular mass of the unlabelled NHis-SDF1 (lacking N-terminal methionine residue) is 10540 Da; reduced to 10536 if the two disulphide bonds are formed. With an experimentally determined molecular mass of 10536 Da, mass-spectrometry of unlabelled NHis-SDF1 provided the confirmation that the CFPS protein was translated to completion (Table 4.2 and 4.3) and the disulphide bonds formed.

Unlabelled NHis-SDF1, with the theoretical molecular weight of 10536 Da (disulphide bonds formed), has 744 hydrogens, of which 564 hydrogens are non-exchangeable. Following purification of labelled and unlabelled NHis-SDF1 mass spectrometry (carried out in $H_2O$ based buffer), provided data on the level of deuteration of the samples (Figure 4.9, Table 4.2). The theoretical molecular masses of unlabelled SDF1, 100 % deuterated and 75 % match-out deuterated in a $H_2O$ based buffer were calculated by adding the number of deuterium atoms that replaced the hydrogens in the protein to the molecular mass of the unlabelled protein (10536 Da, 11100 Da and 10959 Da respectively – see Table 4.2). The mass spectrometry data show that 75 % deuterated SDF1 is produced when the ratio between deuterated and unlabelled amino acids in the CFPS reaction is 95:5. This corresponds to NHis-SDF1 produced having a molecular mass of 10959 Da and 75 % deuteration level in non-exchangeable positions, which is able to be used for match-out SANS experiments (Figure 4.9B and Table 4.3). Due to the CFPS reaction being carried out in hydrogenated conditions (use of $H_2O$-based lysate, solvents and reagents), even when using 100 % deuterated amino acid, higher than 77 % level of deuteration could not be attained. However, this is not required for for SANS matching-out experiments.



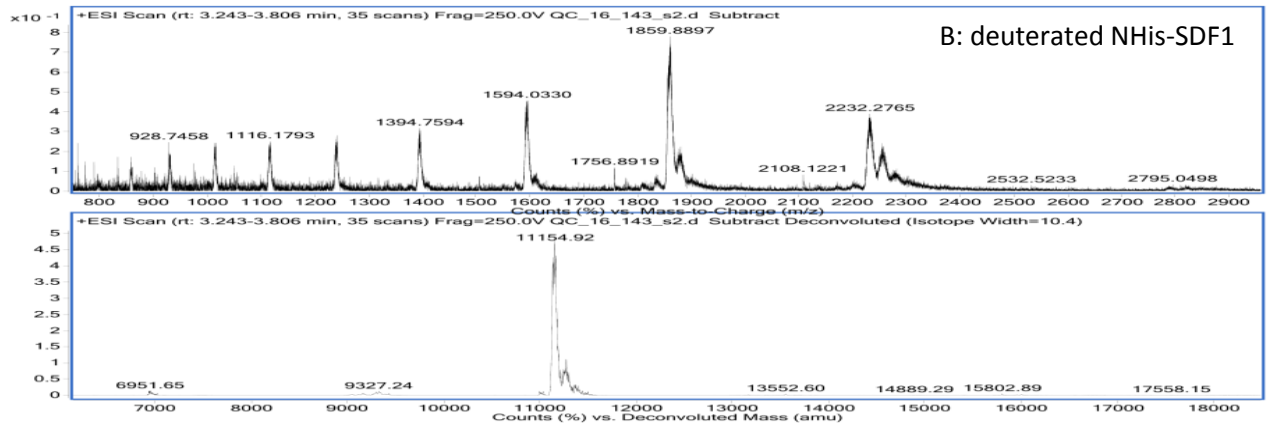




***Figure 4.9:*** *Mass spectrometry data of A: unlabelled NHis-SDF1, B: deuterated NHis-SDF1.*

| Sample | Deuteration level in non-exchangeable hydrogen positions (%) | Number of non-exchangeable hydrogen atoms | Number of non-exchangeable deuterium atoms | Expected molecular mass (Da) |
|---|---|---|---|---|
| **H-SDF** | 0 | 564 | 0 | 10536 |
| **D-SDF 100** | 100 | 0 | 564 | 11100 |
| **D-SDF 75** | 75 | 141 | 423 | 10959 |

***Table 4.3:*** *Theoretical number of hydrogen and deuterium atoms and expected corresponding molecular masses of unlabelled, 100 % and 75 % deuterated NHis-SDF1 samples where disulphide bonds are formed.*

| Sample | Observed Mass (Da) | Percentage of deuterated amino acid used in CFPS reaction (%) | Percentage deuteration for non-exchangeable hydrogens (%) |
|---|---|---|---|
| **H-SDF** | 10536 | 0 | 0 |
| **D-SDF 100** | 10969 | 100 | 77 |
| **D-SDF 95** | 10958 | 95 | 75 |
| **D-SDF 85** | 10875 | 85 | 60 |
| **D-SDF 80** | 10841 | 80 | 54 |

***Table 4.4:*** *Mass spectrometry data for unlabelled and deuterated NHis-SDF1. Figures in bold relate to ideal conditions for match-out labelling to be used in SANS experiments.*

### 4.3.4. Chemotaxis assay

The objective of the chemotaxis assay was to confirm the functionality of CFPS SDF1-α variants with respect to their chemotactic properties.

NHis-SDF1 was cleaved successfully using the thrombin digest with almost 100 % purity of SDF1-α being obtained (Figure 4.10). The cleaved NHis-SDF1 and the other variants of CFPS SDF1-α were subsequently used in chemotaxis assays. The major biological effect of SDF1-α is related to the ability of this chemokine to induce chemotactic responses. Using the Transwell system SDF1-α (0 – 300 nM) present in the lower chamber caused the migration of the cells from the upper chamber. The higher the chemotactic properties of the SDF1-α is, the lower is the dose (concentration) required for inducing migration of the cells. Indicative of its functionality, CHis-SDF1 (CFPS SDF1-α having the histidine tag at the C-terminal) was shown to have the highest chemotaxis response. In contrast, NHis-SDF1 showed less responsiveness. This was expected as it known that the exposure of the N-terminal of SDF1-α is required for functionality (Loetscher *et al.*, 1998, Veldkamp *et al.*, 2008b). Importantly, the chemotaxis assay showed that the deuterated CHis-SDF1 and deuterated NHis-SDF1 both had equivalent responses to their unlabelled counterparts.

At high chemo-attractant concentration, either due to the quick loss of the chemokine gradient by diffusion, a diminished cellular migration was observed (Picciocchi *et al.*, 2014). This was particularly observed for the highly potent variants of SDF1-α (SDF-Syn, SDF1-C-His, D- SDF1-C-His), which produces a typical bell-shaped dose-response curve. In case of the less potent SDF1-α variants (NHis-SDF1, D-NHis-SDF1), the bell-shaped dose-response curve is ,however, shifted to the right due to higher amounts of active SDF1-α required for producing equivalent response as that of potent SDF1-α variants.

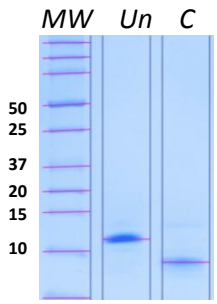



***Figure 4.10**: Coomassie-stained PAGE showing highly purified Cleaved NHis-SDF1 used for subsequent chemotaxis. MW: Molecular Weight; Un: Un-cleaved NHis-SDF1; C: Cleaved NHis-SDF1.*

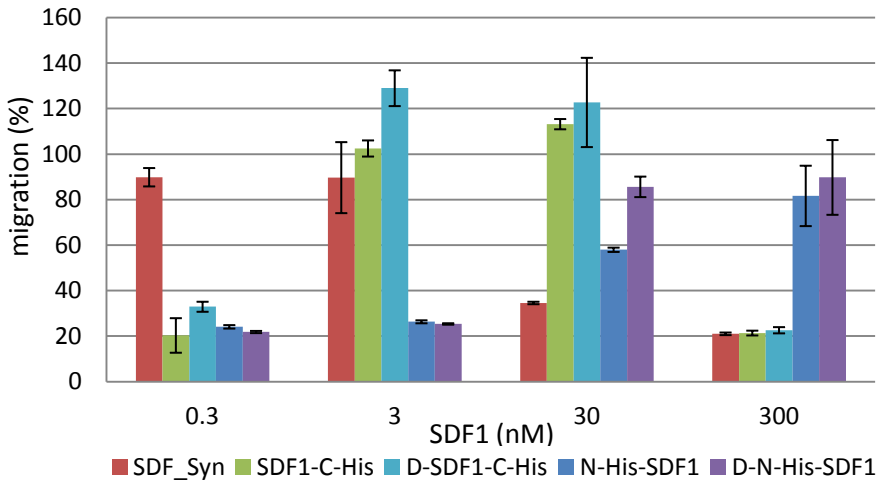



***Figure 4.11:** Chemotaxis assay results showing functionality of SDF1-α.The various forms of SDF1-α used in the chemotaxis assay showed dose dependent responses in terms of cellular migration as a percentage of control well without any SDF1 included. The higher the chemotactic potential of the SDF1-α, the less is required to bring about migration of cells. SDF-Syn: synthetic SDF1 control (provided by R. Sadir, purchased from F. Baleux, Pasteur Institute), SDF1-C-His: SDF1 with histidine tag at the C-terminal, N-His-SDF: SDF1 with histidine tag at the N-terminal, D-SDF1-C-His: Deuterated SDF1 with histidine tag at the C-terminal, D-N-His-SDF: deuterated SDF1 with histidine tag at the N-terminal. NB: Higher than 100 % migration and errors are due pipetting inaccuracies during chemotaxis setup.*

### 4.3.5. NanoSight™ particle tracker: monitoring changes from liposomes to proteoliposomes during CXCR4 CFPS

The stock of liposomes used in CFPS and freshly-made by sonication and passed through a 0.22 µm filter is homogenous with the liposomes having average diameter of 58 nm, and most them being between the ranges of 10 - 100 nm (Figure 4.12). For the purposes of the CFPS reaction, this range is considered to be adequate. The lysate was, as expected, found to be a highly heterogeneous sample. At mostly less than 120 nm in diameter, these particles were expected to be residual insoluble suspensions that were derived from lysate production. The addition of liposomes to the reaction at the start of CFPS was found to bring about an increase in the number of particles from 23 x $10^{13}$/ml (total number of particles in liposome and lysate) to about 120 x $10^{13}$/ml (Figure 4.13). The addition of liposomes to the CFPS reactions results in the precipitation of the liposomes and other CFPS components. This is feature is readily observed as the reaction changes from a translucent liquid to opaque white. With the addition of liposomes to lysate, an increase in average diameter of the liposomes (148 nm) was observed. It is highly likely that coalescence of liposome-to-liposome and proteins-to-liposome accounts for this observation. Indeed, as used here, phosphatidylcholine-rich liposomes have been described to fuse spontaneously and form giant unilamellar liposomes in the presence of divalent ions such as $Mg^{2+}$ and $Ca^{2+}$ (Akashi *et al.*, 1998), which are both present in the typical CFPS reaction.

At the end of CFPS, the number of particles decreased again to less than 20 x $10^{13}$/mL (Figure 4.13). Following the CFPS reaction, the average particle diameter was recorded to be about 200 nm, and ranging between 100 – 500 nm. The decrease in numbers of particles could be explained by further fusion of proteoliposomes to form larger particles. Even more fusion, and concomitant decrease in the number of proteoliposome particles, was brought about after subsequent sucrose-gradient purification, centrifugation washes and resuspension into HEPES buffer, as described in the previous chapter. Nevertheless, it should be mentioned that the loss of proteoliposomes during each of these steps (due to manipulation during dilution, resuspension, recovery...) is contributing to an overall reduction in the number of particles.

With insertion of membrane proteins into the liposomes, the resulting proteoliposomes increase in size. This was noticed here as a significant difference in the diameter of CXCR4-proteoliposomes compared to those of negative control. At the end of CFPS incubation, proteoliposomes from the CXCR4 reaction had a larger mean diameter of 220 nm as compared to the negative control of 199 nm. Following

purification, CXCR4 proteoliposomes were found to be comprised much more distinct particles, namely at 42, 82, 133, 196, 290 and 415 nm (Figure 4.12). In the case of negative control, the NanoSight™ profile indicated less distinct peaks for particles, with peaks at 125, 174 and 205 nm (Figure 4.12).

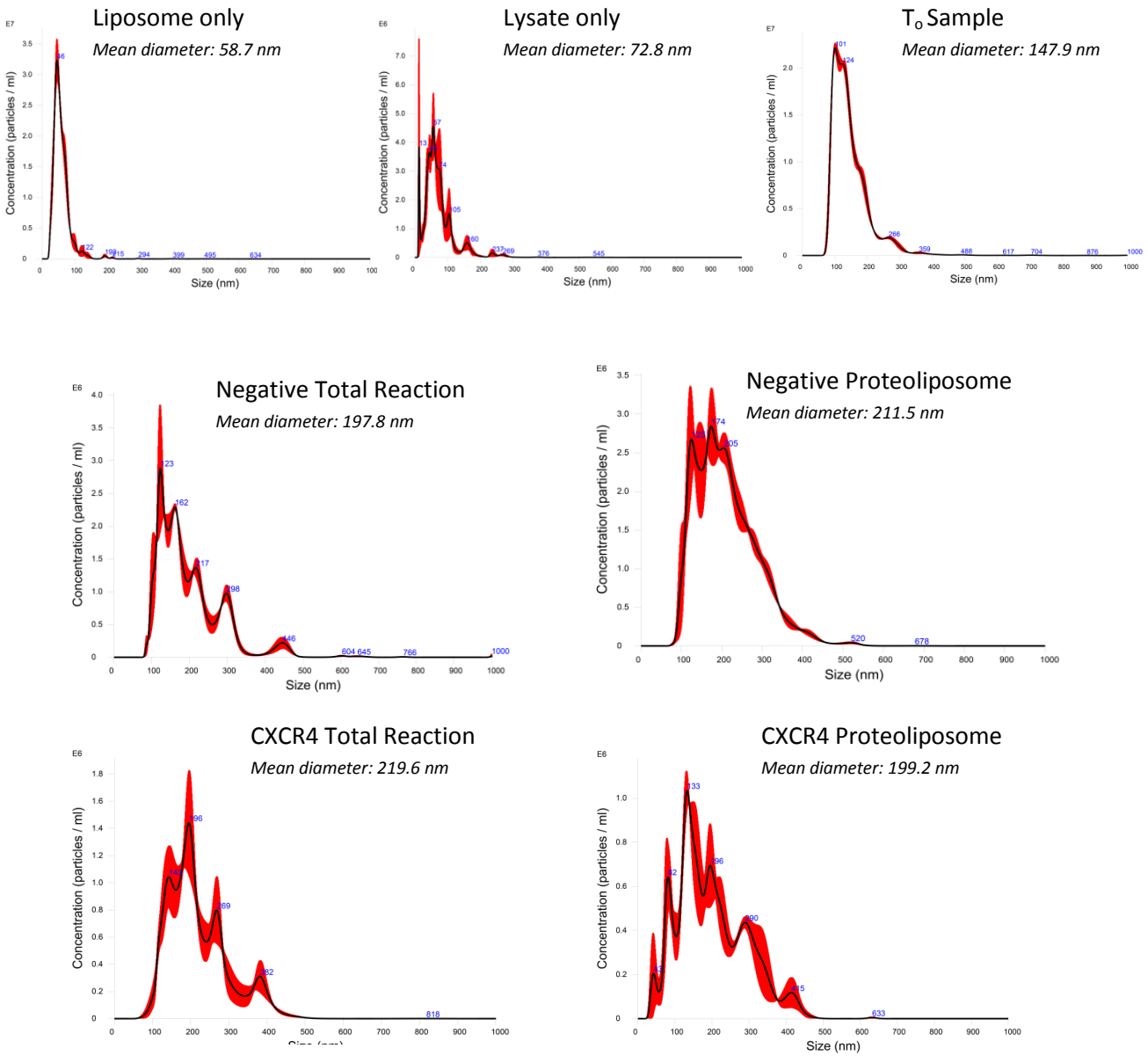



***Figure 4.12:*** *Size distribution (concentration of particles/mL against particle diameter/nm) of CXCR4-proteoliposomes and negative reaction before and after CFPS reaction, and after sucrose-gradient purification, as recorded by the NanoSight™ NS300 particle tracking analyses. The graphs indicate the shift in relatively homogenous liposome sample to highly heterogeneous CXCR4-proteoliposome sample at the end of CFPS reaction and purification. Data for each run shown is comprised of merged data from three separate runs.*

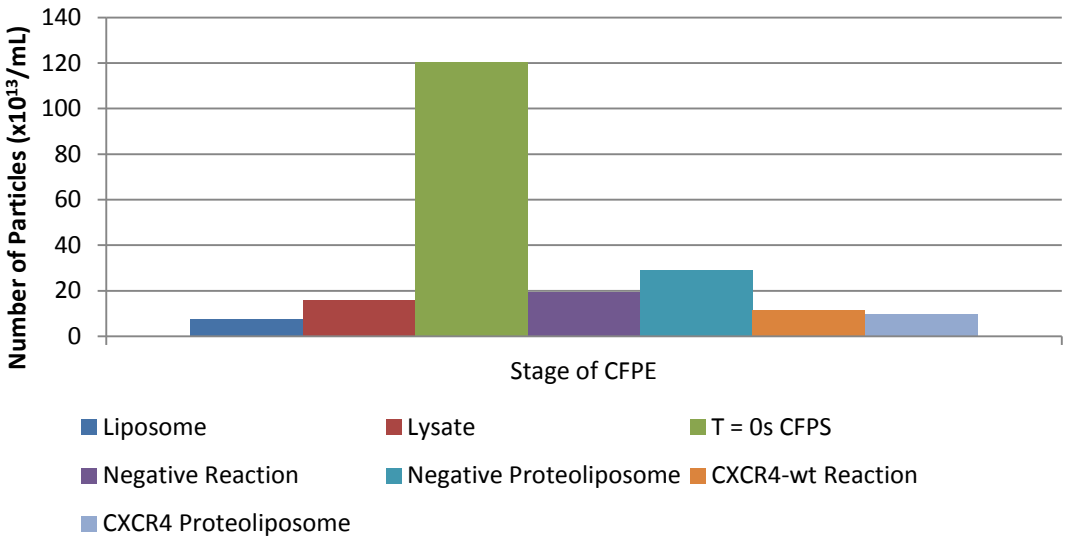



***Figure 4.13:*** *NanoSight™ NS300 recorded data of changes in particle sizes and number before, during and after CXCR4* CFPS *reaction and purification. It should however be noted that measurement of the total number of particles is not only due to liposomes/proteoliposomes, but also due to insoluble particles and aggregates.*

### 4.3.6. Transmission electron microscopy of CXCR4 and full-length CD4 proteoliposomes

Interpretation of the results from TEM proved to be quite difficult since the results show that the staining during sample preparation is likely to have disrupted the proteoliposomes (Figure 4.13). In order to provide the electron scattering properties of a sample, TEM requires the sample to be treated with heavy metal stains which interact with the particles in the sample. In all cases tested here, the use of uranyl acetate (UA) however proved to be less destructive to proteoliposomes compared to phosphotungstate acid (PTA). With the use of PTA, all of the proteoliposomes were destroyed leaving a dense sea of phospholipids (Figure 4.13, B and D). Using UA instead, whilst admittedly still destructive, proteoliposomes and proteoliposome fragments of respectively ≈ 100 nm and ≈ 25 nm diameter could be observed (Figure 4.13, A, C, E and F). In both cases, the presence of large aggregates is seen. These aggregates are likely to be a combination of proteoliposomes and insoluble particles.

Particles in a sample have varied affinity for the heavy metal used in TEM stains, and therefore can appear as electron-dense (high affinity) or electron-lucent (low affinity) and are observed as dark or light structures on the resulting TEM image. As compared to CXCR4- and CD4-embedded proteoliposomes, appearing as dark particles, negative control proteoliposomes were observed as lighter particles, surrounded by a dark rim. It is very plausible that due to active protein expression and membrane protein insertion into the liposomes during CFPS, the resulting proteoliposomes is substantially enriched in proteins which contribute to high electron-density. In the case of the negative control, lower protein amounts and a higher lipid proportion gives an electron-lucent structure and reduced TEM staining. The higher amount of electron dense CXCR4-proteoliposome fragments compared to that of CD4 can also be noted. This can be explained by the lower temperature use for CFPS of CD4 with respect to the production of CXCR4, which results in decreased overall protein expression, and reduces the non-specific adhesion of lysate protein to the proteoliposomes. These results were shown by the western blot in the previous chapter (Figure 3.9) where less aggregates were observed from the proteoliposome CFPS reaction which had been incubated at 20$^{o}$C compared to that incubated at 30 $^{o}$C.

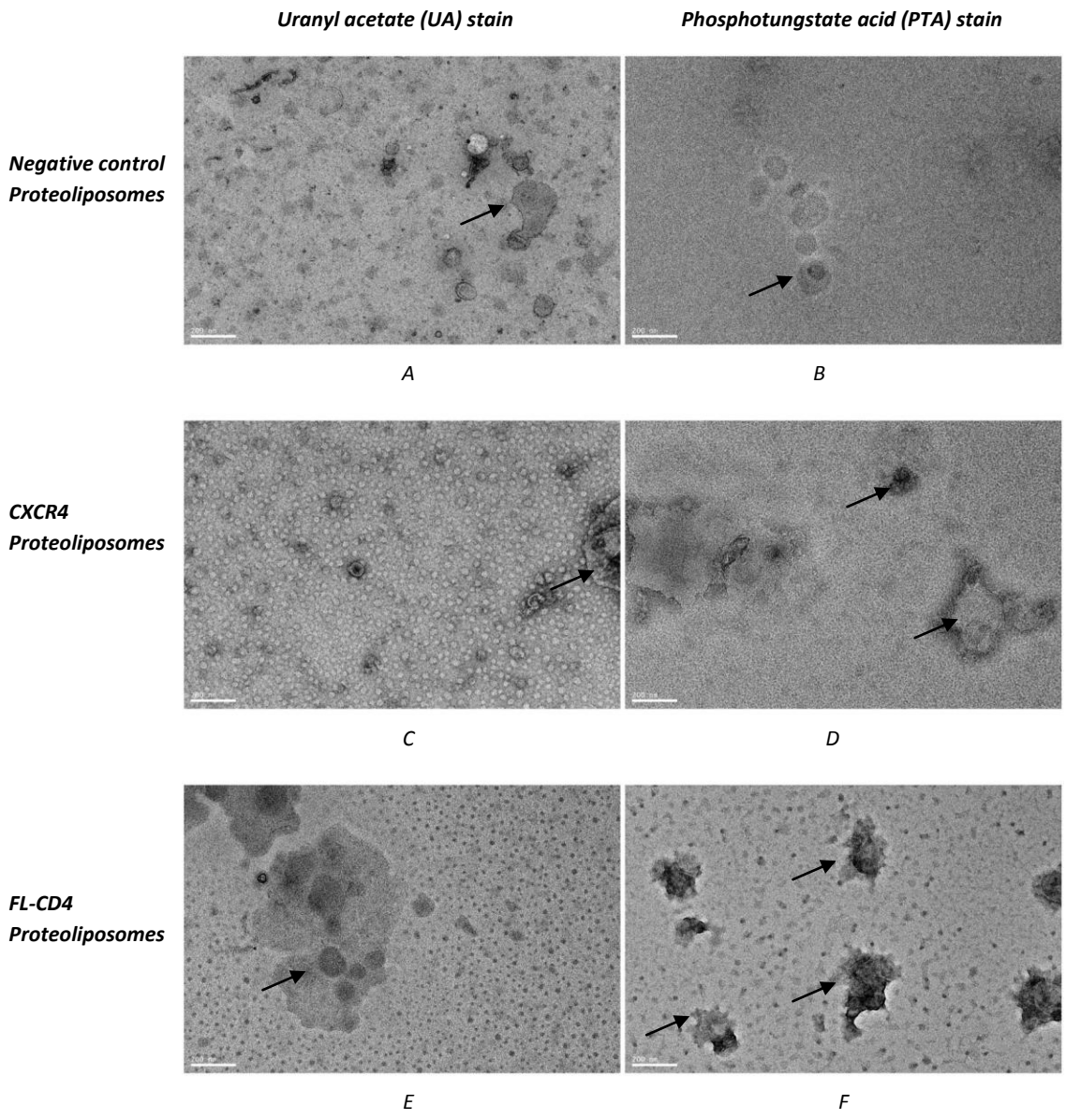



***Figure 4.13:*** *TEM images of negative control proteoliposomes. 1:100 dilution, UA stain (A); 1:100 dilution, PTA stain (B); 1:10 dilution, UA stain (C); 1:10 dilution, PTA stain (D). CXCR4 proteoliposomes at 1:10 dilution, UA stain (E); CD4 proteoliposomes at 1:10 dilution, UA stain (F). Scale bar: 200 nm. Example of proteoliposome aggregates are marked with arrows.*

## 4.4. DISCUSSION

### 4.4.1. Indication of the structural integrity of CFPS CXCR4

It has been reported previously that nearly 50 % of detergent-solubilised mammalian-cell expressed CXCR4 showed poor binding to the conformation dependent 12G5 antibody (which targets the extracellular loop 2 of CXCR4 and shown to bind to functional native mammalian expressed CXCr4 (Baribaud *et al.*, 2001) during SPR analysis (Kofuku *et al.*, 2009). Additionally, in the same paper, detergent-solubilised CXCR4 was reported to be extremely unstable and prone to aggregation. Here however, SPR of CFPS CXCR4 in proteoliposome using 12G5 indicated that the majority of the protein is folded correctly. The SPR results here therefore provide evidence to the structural integrity, and consequent functionality of CFPS CXCR4 in proteoliposomes. As such, CFPS seems to be a superior method for the production of active CXCR4.

### 4.4.2. Easy expression and recovery of active full-length CD4

Due to the requirement of the lipid bilayer for supporting the transmembrane domain, recombinant expression of FL-CD4 has been very challenging. Instead, active soluble 2dCD4 lacking the transmembrane domain have been expressed using eukaryotic cells (Brady *et al.*, 1993, Hussey *et al.*, 1988). Analogous to the CD4 in proteoliposomes that was produced here, attempts to stabilise recombinant baculovirus insect-cell expressed FL-CD4 have for instance, involved the electro-insertion of the molecule into the cell membrane of red blood cells (Zeira *et al.*, 1991). The modified red blood cell showed viral inhibitory properties when tested *in vitro*. Here, full-length CD4 (FL-CD4) expressed in proteoliposomes, using a relatively easy CFPS method, was shown to be functionally active in permitting the binding of GP120 and ibalizumab (Figure 4.8). Incidentally, ibalizumab being a conformational-dependent antibody (Song *et al.*, 2010, Freeman *et al.*, 2010, Burkly *et al.*, 1992), evidence for the structural integrity of CD4 is further demonstrated.

### 4.4.3. Confirming the expression of active unlabelled and labelled SDF1-α

The lack of cellular boundaries and the openness of the CFPS system provide a relatively easy way to produce isotopic-labelled proteins, especially in the field of NMR (Kigawa *et al.*, 1999, Matsuda *et al.*, 2007, Staunton *et al.*, 2006). By incorporating deuterated amino acids in the CFPS reaction, the deuteration of SDF1-α could be achieved (Figure 4.9, Tables 4.2 and 4.3). Furthermore, using chemotaxis assays, it was also shown that both the unlabelled and deuterated SDF1 were biologically active.

The production of deuterated protein *in vivo* in *E. coli* requires the adaptation of the transformed *E. coli* to grow and express the protein of interest in deuterium-based minimal media (Haertlein *et al.*, 2016, Dunne *et al.*, 2017a). However, this task can be quite labour-intensive, time-consuming and costly – especially when it concerns difficult proteins. The use of CFPS for the production of deuterated proteins can prove advantageous over the traditional *in vivo* system. Firstly, direct incorporation of deuterated amino acids in the reaction allows for deuterated protein production. Secondly, the reaction is not limited by cellular viability. This can be particularly problematic when growing cells in $D_2O$ based media. These two advantages, when coupled together, give CFPS a potential advantage as a quick method for the expression of deuterated protein.

Using CFPS for the expression of SDF1, a maximum 77 % deuteration level for non-exchangeable hydrogen was obtained (Table 4.3). As described by Etezady-Esfarjani, the use of CFPS conducted in $D_2O$-based reaction mixtures produced chaperonin GroEL with 95 % deuteration for the non-exchangeable hydrogen (Etezady-Esfarjani *et al.*, 2007b). The reaction described here was conducted in $H_2O$-based conditions, and therefore back-protonation of $^2H$ to $^1H$ is significant and thus explains the relatively lower deuteration levels (Etezady-Esfarjani *et al.*, 2007b, Tonelli *et al.*, 2011). As shown by NMR, aspartic acid, asparagine, glutamine, glutamic acid are particularly prone to back-protonation in both the α- and β- positions, and thereby contributing to lower deuteration levels (Etezady-Esfarjani *et al.*, 2007b). However, with the use of $D_2O$ medium, the yield of protein expression is reduced and requires supplementation with exogenous polymerases (Etezady-Esfarjani *et al.*, 2007b, Hohlefelder *et al.*, 2013). Interestingly, the rate of transcription is increased in $D_2O$, but the translation and folding is compromised (Tonelli *et al.*, 2011). Nevertheless, deuterated SDF1-α produced here (95:5 daa:haa) is labelled at a level that optimally matches matched-out conditions for SANS experiments, as described by Dunne *et al.* 2017.

### 4.4.4. Observing changes from liposomes to proteoliposomes

Whilst CFPS has been used to successfully produce membrane proteins in proteoliposomes (Nozawa *et al.*, 2011, Goren and Fox, 2008), the reaction itself remains a "black-box". The NanoSight™ particle tracker proved to be a very efficient, easy-to-use and highly visual-based device in order to obtain information on size distribution of the particles throughout CFPS. Firstly, it was confirmed that the stock of in-house made liposome was homogenous, with particles less than 100 nm. The fusion of liposomes, likely due to presence of divalent ions (Akashi *et al.*, 1998), then occurs when they are added to the CFPS reaction. The formation of proteoliposome is characterised by an increase in the size and heterogeneity of liposomes upon the insertion of membrane proteins. DLS studies on the reconstitution of purified bacterial membrane protein OmpF into liposome reported the shifts of mean vesicles size from 100.1 ±4.5 nm to 230.2 ±71.4 nm (Neves *et al.*, 2009). Here, when compared to negative control proteoliposomes where diffuse profile and relatively smaller particles were observed (peaks at 125, 174 and 205 nm), the insertion of CXCR4 into the liposomes resulted in more distinct profile and larger-sized proteoliposomes (82, 133, 196, 290 and 415 nm) (Figure 4.12).

Changes in the size and distribution of the positive and negative proteoliposome controls (Figure 4.12, 4.13) can, however, also be attributed to the interaction of CPFE reagents and non-specific adhesion of lysate proteins to the liposomes. The attachment of these proteins leads to the formation of protein corona, which would explain the diffuse distribution of proteoliposome diameters, particularly observable in the negative control proteoliposomes. An analogous observation has been reported with non-membrane blood serum proteins immediately adhering to liposomes, and other colloidal nano-particles (Diederichs, 1996, Pareek *et al.*, 2018, Zhdanov and Cho, 2016). Understanding the formation of the protein corona, and modification of the liposome composition to reduce protein adhesion has been an important topic in the development of "stealth" nano-particles used as drug delivery agents. For instance, PEGylated liposomes have often been described to reduce protein adsorption to its surface (Papageorgiou *et al.*, 2018, Sangra *et al.*, 2017). Lessons learnt from these studies, coupled with the use of the NanoSight™, can help to establish methods which will reduce the amount of adhering contamination proteins in the quest for pure, homogenous proteoliposomes samples. The NanoSight™ particle tracker could become a useful tool for assessing, improving and standardising the quality of CFPS reactions, the subsequent purification and resuspension buffers. This is of particular importance in the industrial setting and could be implemented for protocol refinement and quality control procedures.

Additionally, the study carried out here can serve as a reference for the use of the NanoSight™ particle tracker which has multiple wavelength lasers and fluorescence detector system. The improved device would therefore have the capacity of detecting fluorescent tags and fluorescent antibodies which have been used to label specifically CXCR4. This could potentially answer the very common question asked by Synthelis customers: how many correctly inserted and oriented CXCR4 molecules are present in the proteoliposomes? Subsequently, the device could be used to characterise the binding of SDF1-α to CXCR4 after each of these have been specifically labelled with a fluorescent tag or antibody.

Potentially complementing the Nanosight studies and with the increase in resolution limits, cryo-EM has rapidly become a very successful tool for studying isolated membrane proteins, membrane proteins embedded in the membrane and the dynamics of the membrane (Goldie *et al.*, 2014, Abe and Fujiyoshi, 2016, Mitsuoka and Gerle, 2016, Rawson *et al.*, 2016). As a preliminary step towards high-resolution cryo-EM studies on the proteoliposomes, standard negative-staining for TEM was initiated in this work. Negative-staining for TEM however remains a challenging technique for studying liposomes and other soft-matter due to its artefact-introducing and destructive propensity (Franken *et al.*, 2017). TEM, admittedly with the use of optimised freeze-fracture protocols, has however been used for visualisation of either endogenous bacterial proteins or CFPS membrane proteins re-inserted into the liposomes (Klammt *et al.*, 2004b, Costello *et al.*, 1984). One example is the *E. coli* transmembrane carrier protein *lac*. This was inserted into liposomes, and the morphology, assembly and distribution, occurring as studs of less than 10 nm, was described using TEM. The re-insertion of a membrane protein into the liposome, however, provides a clean picture of only the membrane protein, and therefore excludes the observation of contaminating proteins. As indicated in chapter three (Figure 3.14), this is not the case when using CFPS proteoliposomes here, where contaminating proteins adhering to the proteoliposomes are constantly present. Furthermore, freeze-fracture is a laborious process requiring skilled personnel (Carson, 2014) which would limit it as a routine analytical technique for CFPS. Possibly, with the use of optimised staining protocols and electron microscopy techniques such as for liposomes, such as in-situ TEM (Hoppe *et al.*, 2013), TEM could be further used in our case.

## 4.5. CONCLUSIONS

Preliminary CFPS tests, using ELISAs, SPR and chemotaxis assays, with CXCR4, CD4 and SDF1-α as model proteins, produced either in proteoliposomes or directly as soluble proteins, provided an indication on the tools that could used to biophysically and functionally assess CFPS proteins.

Whilst the Synthelis CFPS system specialises in the expression of proteins as proteoliposomes, it was also shown that system could be adapted for the production of active soluble proteins without of liposomes. The ease of expression of soluble CD4 or SDF1-α here, depicts the potential of the system by producing active proteins of value in fields of HIV or cancer research. Additionally, the ability of CFPS to produce labelled proteins with ease offers potential in advanced structural studies. As described in chapter six, the deuterated CFPS expressed 2dCD4 was used for SANS studies.

It is necessary to validate the functionality of proteins prepared using CFPS via easy, quick, reliable and economical tools. This is especially important in the context of small industrial setups where resources can be limited, and self-sufficiency from larger businesses is certainly advantageous. As demonstrated in this chapter, ELISA, SPR and NanoSight™ are tools which could be relatively easily be implemented on the bench-side as part of the quality control battery in an industrial setting. This is usually not the case using methods such as TEM or X-ray crystallography, which are only accessible by larger companies or through institutions and are risky and slow (X-ray crystallography). Of these methods, ELISA is certainly the most accessible, versatile and economical method, of which the main consumable costs are the antibodies and ligand. Validation of the functionality of the protein would thus bring about added market value to the protein products. Requiring high initial investment cost, additional bench-side analyses such as SPR (45,000 – 275,000 €) and the NanoSight™ (30,000 – 100,000 €) can be introduced later on with industrial growth, in order to provide highly accurate quantitative results, for even further added market value of protein products. These would additionally permit troubleshooting of the protocols and products used, that will, eventually lead to better standardisation, reproducibility and higher quality control.

Due to their complexity and heterogeneity, characterising proteoliposomes remains a challenge. Well-adapted methods need to be developed and specific equipment is usually required. In the context of industry, tools for such characterisation which can be implemented as easy-to-use routine

procedures, would serve as additional quality control checks. Here, a study with the aim of assessing the capability of the NanoSight™ particle tracker as an easy-to-use tool for monitoring the changes in liposomes/proteoliposomes during a typical liposome-based CFPS reaction and purification was explored. Incidentally, the study provides a method which can be used to deepen the understanding CFPS in order assist in improving the homogeneity of the liposome/proteoliposome samples.

Finally, as overall summary of chapters two to four, Table 4.4 provides the pros and cons of CFPS against classical *in vivo* expression system for the expression of soluble and membrane proteins (Table 4.4 and 4.5). With the openness of the CFPS, the system can be directly modified and optimised in order to produce either functional membrane or soluble proteins. Expression *in vivo* on the other hand can be less subject to modulation due the presence of cellular membranes and the need to maintain cells alive. The immediate production of functional membrane and soluble proteins furthermore avoids the labour-intensive refolding step which is occasionally required with the expression of non-functional, insoluble recombinant proteins in *E. coli*. Additionally, CFPS allows the production of labelled protein with relative ease by substituting the amino acids in the reaction for labelled amino acids.

Overall, despite the higher production costs of CFPS, a key advantage of the system compared to *in vivo* production, is the speed at which functional protein, both unlabelled and labelled, can be expressed and recovered (provided that lysate is available already). CFPS can therefore be a viable alternative for protein production, in both the industrial and academic contexts, where tight deadlines need to be met.

| Criteria of soluble protein expression | CFPS | *In vivo* bacterial expression | *In vivo* eukaryotic (mammalian) expression |
|---|---|---|---|
| **Setup time before protein expression** | 15 min: combining pre-made CPFE ingredients | 2 days: transformation of cells and pre-culture | 2 – 3 weeks: transfection and propagation of transfected cells |
| **Time for protein expression following setup** | Within a few hours to overnight | Overnight | Continuous, once cells are maintained |
| **Relative expression cost** | Intermediate | Low | High |
| **Laboratory needs** | Standard biological laboratory equipment | Standard biological laboratory equipment | Specialised mammalian cell culture laboratory |
| **Protein solubility** | Soluble, with optimised CFPS | Insoluble | Soluble |
| **Localisation of expressed protein** | CFPS reaction medium | In bacterial cells, as inclusion bodies | Secreted into supernatant |
| **Purification** | Direct nickel-affinity chromatography of CFPS reaction | Requires cell lysis, extraction of cytosol, separation of soluble and insoluble cytosol, solubilisation of insoluble cytosol with solubilisation agent (e.g. 8 M urea), refolding and nickel-affinity chromatography | Direct nickel-affinity chromatography of supernatant |
| **Folding state of purified protein** | High chance of correct folding | Denatured, requires specific and optimised refolding protocol | Folded |
| **Time for recovery of folded protein** | 1 day | 3 – 4 days | 1 day |
| **Deuteration of expressed protein** | Easy and immediate possibility by incorporating labelled amino acids in CFPS reaction | Requires cell adaptation in minimal media growth prior to deuterated protein expression | Relatively unfeasible due to poor adaptation of mammalian growth in deuteration media |
| **Cost of protein deuteration** | High, requires use of costly deuterated amino acids | Low | NA |

***Table 4.5:*** *Comparison of various expression systems for the expression and purification of a typical soluble protein.*

| Criteria of membrane protein expression | CFPS | *In vivo* bacterial expression | *In vivo* eukaryotic (mammalian) expression |
|---|---|---|---|
| **Setup time before protein expression** | 15 min: combining pre-made CPFE ingredients | 2 days: transformation of cells and pre-culture | 2 – 3 weeks: transfection and propagation of transfected cells |
| **Time for protein expression following setup** | Within a few hours to overnight | Overnight | Continuous, once cells are maintained |
| **Relative expression cost** | Intermediate | Low | High |
| **Laboratory needs** | Standard biological laboratory equipment | Standard biological laboratory equipment | Specialised mammalian cell culture laboratory |
| **Protein solubility** | Solubilised with detergents | Insoluble | Insoluble |
| **Localisation of expressed protein** | Inserted into liposomes when expressed as proteoliposomes; In reaction medium when solubilised with detergents | Generally as inclusion bodies in bacterial cells | Inserted into cellular membrane when expressed as proteoliposomes |
| **Purification** | **Proteoliposome:** sucrose gradient, ultracentrifugation, and centrifugation **Detergent solubilised:** direct nickel-affinity chromatography of CFPS reaction | Requires cell lysis, extraction of cytosol, separation of soluble and insoluble cytosol, solubilisation of insoluble cytosol with solubilisation agent (e.g. 8 M urea), refolding and nickel-affinity chromatography | Requires recuperation of cell membranes, solubilisation of membrane protein with detergents, and nickel-affinity chromatography of supernatant |
| **Folding state of purified protein** | High chance of correct folding when in proteoliposomes | Denatured, requires specific and optimised refolding protocol | Folded |
| **Time for recovery of folded protein** | 1 day | 3 – 4 days | 1 day |
| **Deuteration of expressed protein** | Easy and immediate possibility by incorporating labelled amino acids in CFPS reaction | Requires cell adaptation in minimal media growth prior to deuterated protein expression | Relatively unfeasible due to poor adaptation of mammalian growth in deuteration media |
| **Cost of protein deuteration** | High, requires use of costly deuterated amino acids | Low | NA |

***Table 4.6:*** *Comparison of various expression systems for the expression and purification of a typical membrane protein.*

# CHAPTER FIVE

## X-RAY CRYSTAL STRUCTURE OF CPFS PRODUCED SDF1-α

## ABSTRACT

Following the CFPS and purification of SDF1-α, the protein was screened for suitable crystallisation conditions. After the determination of crystallisation conditions, the single crystals obtained were then subject to X-ray diffraction at the European Synchrotron Radiation Facility (ESRF). The crystal structure of cell-free expressed SDF1-α was successfully determined to a resolution of 1.74 Å. The structure of CFPS SDF1-α showed strong similarity to previous published structures, which indicates a high degree of structural integrity. This result reinforces the validity of the Synthelis CFPS system in the production of structurally valid proteins. Furthermore, it demonstrates the establishment of the Synthelis CFPS system as further developed in this thesis for future structural applications.


## 5.1. INTRODUCTION

### 5.1.1. Industrial interest for a crystal structure of a CFPS product

Structural validation of the CFPS product is an important quality control to support the validity of the Synthelis CFPS procedures to produce high quality proteins, equivalent to those produced with classical methods. This is critical to instil confidence in the technology. A key objective of this thesis work was therefore the crystallisation and X-ray structure determination of a protein expressed using the CFPS procedure.

Near of the end of 2018, with 146,266 biological macromolecular structures present in the RCSB Protein Data Bank (PDB, *www.rcsb.org*) only about 80 of these were derived from proteins prepared using CFPS methods. This is perhaps surprising given the potential ease of use of CFPS to produce proteins, but clearly shows the lack of take-up of the technique for structural biology. Nonetheless, with the technological advancement and optimisation of laboratory methods relating to CFPS, making it easier to access and exploit the technology, the technique has potential to expand the repertoire of deposited protein structures by becoming a more routinely exploited protein production tool.

Currently, there are 16 deposited structures of full-length human SDF1-α in the RCSB PDB, of which ten are NMR solution structures and the remaining six are X-ray crystal structures. With the exception of

one structure which was chemically synthesised (Dealwis *et al.*, 1998), all of these SDF1-α structures were produced using traditional cellular-based *E. coli* expression systems. To produce soluble and biologically active SDF1-α in *E. coli* there was, however, the need to refold the insoluble protein produced – a laborious and time-consuming task, which does not necessarily guarantee structural and functional integrity of the protein with an associated risk to interpretation of results based on structural and functional analyses. Additionally, unless the protocol for the refolding process has been fine-tuned for the refolding of a specific protein, the yields of the recovered soluble protein can be very low. In the case of the CFPS presented here, despite the use of *E. coli* lysate, an optimised reaction which allows the direct expression of active and soluble SDF1-α was established (chapters three and four). With the high expression yields and the high recovery rates of pure soluble and bioactive CFPS SDF1-α, the protein was therefore suitable as a target for an X-ray crystal structure. Furthermore, a previously published SDF1-α structure determined to a high resolution of 1.6 Å provides a good comparative model (Murphy *et al.*, 2010). This therefore suits the goal in comparing an *in vivo* expressed protein against CFPS using high-resolution structural results. Finally, the implication of SDF1-α in cancer pathology provides pharmaceutical and clinical relevance for studying the protein and provides an interest for a simplified and more effective preparation procedure for the protein (Hendrix *et al.*, 2000).

### 5.1.2. Structure of SDF1-α

As mentioned in chapter one, SDF1-α, the natural ligand of CXCR4, is a pro-inflammatory chemokine that is known to be involved in the migration of chemotactic cells (Kucia *et al.*, 2004). SDF1-α is also critical to the homing and sequestration of circulating haematopoietic into the blood marrow (Moepps *et al.*, 2000). Alternative roles of SDF1-α involve the guiding of embryological precursor cells towards their correct destination. Resulting in prenatal death, mice embryos with disrupted the CXCR4 and SDF1-α interactions, lacked lymphatic, myelopoietic, neural and cardiac development (Nagasawa *et al.*, 1996, Zou *et al.*, 1998). Under pathological circumstances, SDF1-α and its receptor CXCR4 have a high implication in the cancer growth, survival and metastasis (Domanska *et al.*, 2013, Orimo *et al.*, 2005, Jin *et al.*, 2006, Muller *et al.*, 2001).

The family of chemokines are structurally characterised by the presence of four conserved cysteine residues which form two pairs of disulphide bonds. The classical topology of chemokines involves a

flexible N-terminal region of three to eight amino acids, a 10-20 residue N-terminal loop, a short $3_{10}$-helix, three β-strands and an α-helix. The two major families of chemokines are referred to as "CC" and "CXC", taking their names from the sequence around the first conserved cysteine residues. Further minor families of chemokines include "CX3C" and "XC". SDF1-α, as part of the CXC family, has a proline residue between the two first cysteine residues.

Using the crystal structure 3GV3 as a representative model for the ensemble of crystal structures available of recombinantly *E.coli* expressed SDF1-α, the structure reveals an extended disordered N-terminal and follows with a β-β-β-α topology (Figure 5.1) (Murphy *et al.*, 2010). The structure can be described as resembling a "forefinger protruding from a closed fist" (Ryu *et al.*, 2007). Consisting of a total of 68 residues, the anti-parallel β-strands are formed by, respectively, residues 24-28, 38-42 and 46-50, and the α-helix of residues 56-66. Notably, the occurrence of the two disulphide bridges (Cys9-Cys34 and Cys11-Cys50) indicate folding integrity and provide strong conformational constraints to the local residues.

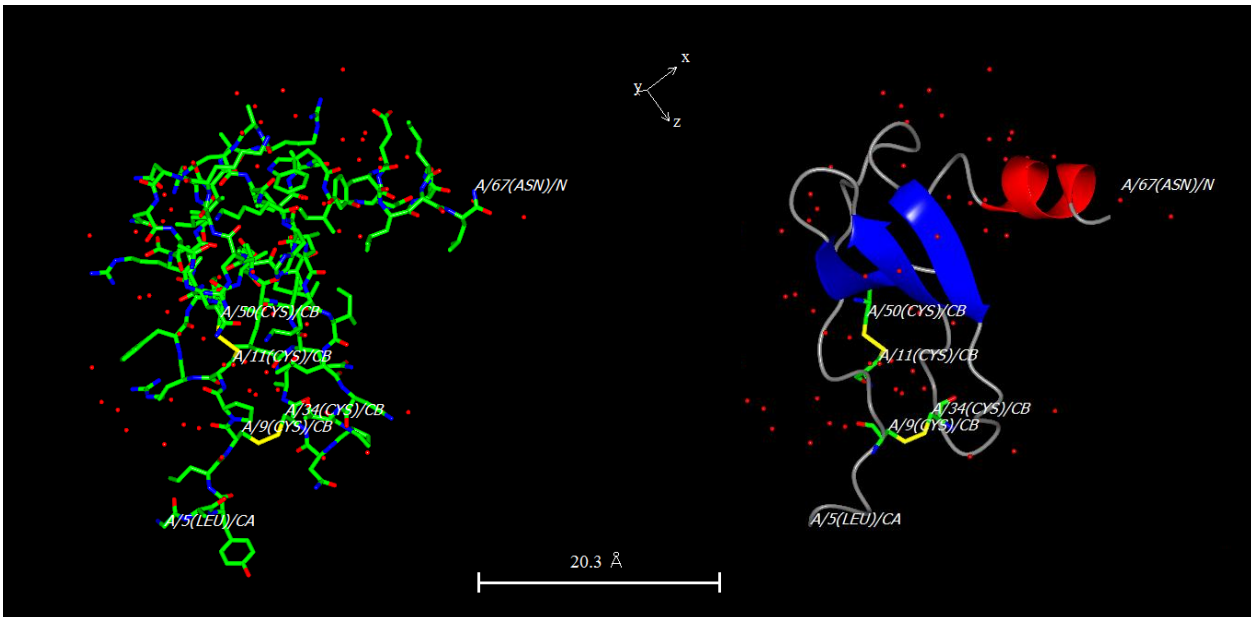



***Figure 5.1:*** *Crystal structure of SDF1-α (PDB: 3GV3, Murphy et al., 2010). (A) Protein chain of SDF1-α, showing Leu5 at the N-terminal, Asn67 at the C-terminal, cysteine residues involved in disulphide bonds (yellow), and the characteristic C-X-C motif, with intervening proline. On the protein chain, oxygen atoms are shown in red, nitrogen atoms in blue and disulphide linkages in yellow. Red dots represent water molecules. (B) Ribbon diagram of SDF1-α in the same orientation is in (A), with α-helices shown in red and β-sheets in blue. Red dots represent water molecules.*

The interaction of chemokines with their target GPCR is suggested to occur by a two-step mechanism (Siciliano *et al.*, 1994, Kofuku *et al.*, 2009). The first step of the interaction occurs between the β-sheets and N-terminal of SDF1-α with the extracellular region of CXCR4, leading to the anchoring of SDF1-α to CXCR4. In the second step, the highly flexible N-terminus of SDF1-α binds to residues buried within the transmembrane domains of CXCR4. This leads to the conformational changes in CXCR4 that are required to induce GPCR signalling. Solution NMR studies have pin pointed the interaction of SDF1-α with CXCR4 to the first 17 residues of the CXCR4 N-terminal, with the main contribution coming from the first six residues (Gozansky *et al.*, 2005). In the same study, based on isotopic chemical shifts, the interaction of the N-terminal of CXCR4 was found to occur beneath the antiparallel β-sheet of SDF1-α. The N-terminal of CXCR4 is then sandwiched by the N-terminal and α-helix of SDF1-α on either side. In another solution NMR study, dimeric SDF1-α bound to a short peptide with the sequence of N-terminal CXCR4 indicated the importance of the sulphotryosine residues present on CXCR4 in maintaining the high affinity ligand-receptor interaction (Figure 5.2) (Veldkamp *et al.*, 2008a). The use of a small molecule which could bind to SDF1-α inhibited the SDF1-CXCR4 interaction by intercalating between the binding site of SDF1-α to specifically the sulphotryosine residue 21 of CXCR4 (Smith *et al.*, 2014). A systematic NMR study on the monomer-dimer equilibrium indicated that the dimerisation of SDF1-α was influenced by non-acidic conditions, and by the presence of divalent anions (Veldkamp *et al.*, 2005). Dimer formation is also promoted by the binding of heparin to SDF1-α (Murphy *et al.*, 2007).

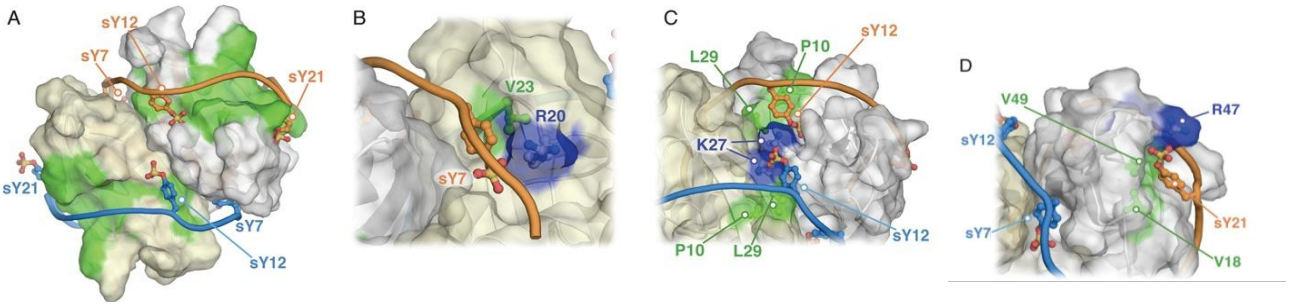



***Figure 5.2:*** *The SDF1-α dimer bound to the short peptide of N-terminal of CXCR4 (orange and blue strands) (PDB: 2KO3). The overall SDF1-α dimer interacting with the two N-terminal of CXCR4 (A). The important interactions made by the sulphotyrosine residues (sY7, sY12, sY21) in the N-terminal of CXCR4 with SDF1-α are shown. Zoom on (B) sY7, indicating the interaction with R20 and V23 of SDF1-α; (C) on sY12, indicating the interaction with P10, L29 and K27 of SDF1-α; (D) on sY21, indicating the interaction with V18, R47 and V49 of SDF1-α (image from: Veldkamp et al., 2008a).*

### 5.1.3. The principle of X-ray crystallography and crystallisation

#### *Basic principles of protein crystallisation*

A crystal is formed of particulate constituents (atoms, molecules or ions) which are arranged in a well-ordered and, generally, closely-packed lattice structure. Crystallisation is a separation technique that involves the separation of a solid phase from a liquid phase, which happens as a crystal is formed (Figure 5.3). In the case of protein crystallisation, highly homogenous, pure and concentrated soluble or solubilised (as in the case of detergent solubilised membrane proteins) protein solution is required. In the vapour diffusion setup ("hanging-drop" or "sitting-drop"), the protein solution is added into a droplet of crystallisation buffer which will permit precipitation of the protein as crystals (Figure 5.3A, B). Without previous knowledge, intensive screening for suitable crystallisation conditions, with different buffer and precipitants at different temperatures and volumes, are usually needed to identify the crystallisation conditions (Figure 5.3C). Today, the screening process is very much facilitated through automation, thanks to crystallisation service platforms which make use of crystallisation robots. The droplet, containing the protein, is enclosed in an air-tight chamber that also contains a reservoir of the crystallisation precipitation buffer at higher concentration than that of the droplet. In order to achieve equilibrium in terms of buffer concentration, net evaporation and diffusion of vapour from the droplet occurs to the reservoir. As a result, the concentrations of the precipitant and protein in the droplet are gradually increased to a metastable, and eventually to a super-saturated state (Figure 5.3D, E). While in the super-saturated state, the protein is not at equilibrium with its diluent. Therefore, to achieve equilibrium the protein is "pushed" out of the solution. The initial transition of the protein molecules from solution, to a disordered and widely-separated molecular state, to crystalline, a regularly and closely-packed molecular state, is referred to as nucleation. Subsequently, the crystal nucleus undergoes growth with the net addition of protein molecules along the crystal lattice.

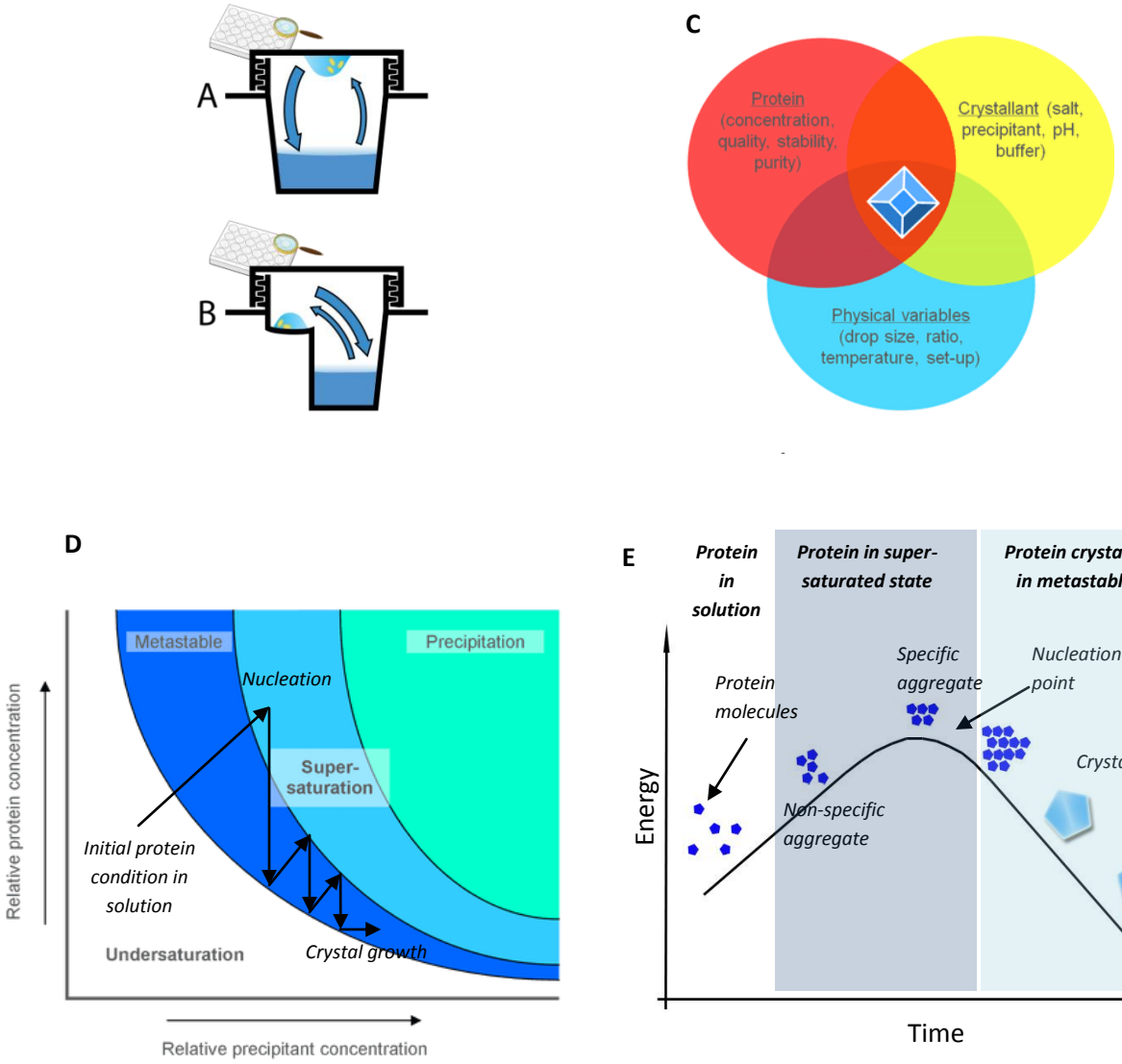



***Figure 5.3:*** *Diagram illustrating crystallisation of a protein. The main vapour diffusion techniques for protein crystallisation are either the hanging- or sitting-drop method (A and B,* By Adenosine - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9818231). *The crystallisation of a protein is complex process that is governed by multiple factors (C;* https://research.csiro.au/crystal/user-guide/protein-structure-via-crystallisation/). *Essentially, with the evaporation and diffusion of vapour from the droplet, the protein concentration is gradually increased into a metastable/super-saturated state (D). The protein molecules form primitive low-order aggregates which gradually increase in order (E). At the critical nucleation point, being in a high energy and unfavourable state, crystal growth ensues with the net addition protein molecules across its lattice.*

### *Basic principle of X-ray crystallography*

Due to their high degree of order and packing, with inter-particular distances close to that of the wavelength of X-rays (around 1 Å), crystals permit the diffraction of incident monochromatic X-rays at specific angles. Brought about by the electrons of the particles present in the crystal, the diffracted X-rays which interact constructively with each other, are observed as reflection spots on a detector screen, and thus form a diffraction pattern (Figure 5.4A). The constructive interference of the diffracted waves is governed by Bragg's Law, which is given by:

$$2dsin\theta = n\lambda$$

**Equation 5.1**

Where $n$ is a positive integer, $\lambda$ is the wavelength of the incident wave, $d$ is the inter-planar distance between lattice planes and $\theta$ is half of the scattering angle between the incident and diffracted waves (Figure 5.4B).

The location and intensity of each of these reflections correspond to a set of evenly spaced electron density clouds from a single plane. By rotating the crystal over a range of angles and subjecting it to X-ray diffraction, the data from the set of generated diffraction patterns from multiple planes can be compiled and processed to allow the 3D reconstruction of an electron density map, which subsequently can enable the determination of the crystal structure at very high resolutions (order of 1 Å).

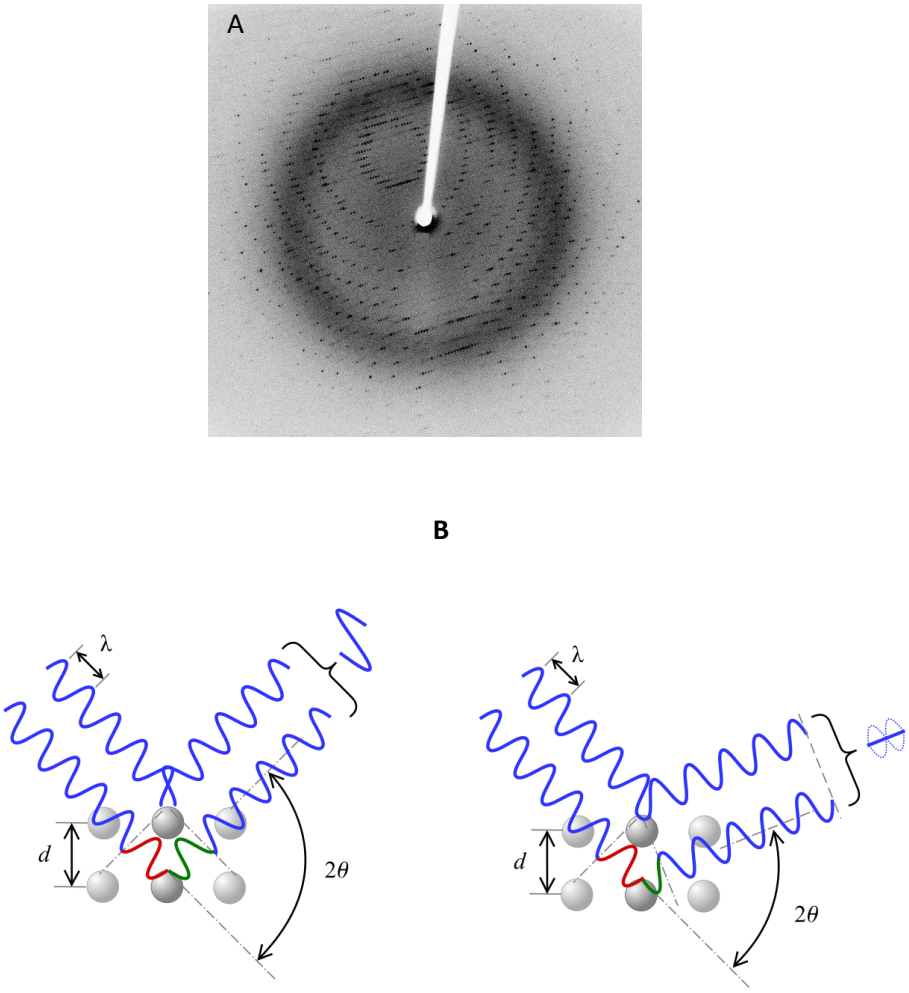



***Figure 5.4:*** *Diagram illustrating the principle of X-ray diffraction in a crystal. (A) The diffraction pattern, which is a collection of reflection spots, can be processed to give the position (an electron density map) of the diffracting electron density in the sample. Here, X-ray diffraction pattern of crystallised a protease (3Clpro, 2.1 Å) is shown as an example (original author: Jeff Dahl, from: https://commons.wikimedia.org/wiki/File:X-ray_diffraction_pattern_3clpro.jpg). (B) The formation of reflection spots is due to constructive interference of waves (left), which is given by Bragg's Law (Equation 5.1). Conversely, destructive interference results in clear regions on the detector (right). Picture by: Loi_de_bragg.png: User: Cdangderivative work: Gregors (talk) - Loi_de_bragg.png, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=14524146.*

### *Molecular replacement*

Data processing, leading to the integrated intensity of each diffracted spot using standard software packages, enables the interpretation of the diffraction patterns and the determination of information about the crystal structure (Figure 5.4): indexing the diffraction spots, identifying the unit cell and space group of the crystal, and recording the integrated intensity of each reflection. In order to be able to calculate an electron density map, the phase of each reflection is required. However, X-ray crystallography experiments do not provide this phase information directly. Several methods exist to overcome this, one of which is the molecular replacement method. The method involves the use of a known protein structure (be it derived from X-ray crystal data, NMR data or other source) which is expected to be very similar to the target structure due to high sequence homology of the known and target protein amino acid sequences. Using software, the model "probe" protein structure is correlated with the target protein data; where there is a high correlation, the resulting calculated phases can provide a sufficiently good starting basis for calculation of an electron density map for model building of the target protein. The atomic model of the studied crystal is then refined over a series of cycles gradually improving the fit of the structure to the data whilst respecting known protein geometry (bond lengths, bond angles, torsion angles). This refinement is followed by two quality control measures known as the R-factor and $R_{free}$-factor which measure the fit of the model structure to the data. The $R_{free}$ factor is calculated from a set of X-ray reflections (typically 5 % of reflections) which do not drive the refinement process and are therefore independent. In the work for the CFPS SDF1- α crystal structure, the existing high resolution of crystal structure *E. coli* expressed SDF1-α (PBD ID 3GV3, at 1.6 Å resolution) (Murphy *et al.*, 2010) was used as the probe for molecular replacement .

## 5.2. MATERIALS AND METHODS

### 5.2.1. Crystallisation

As described in chapter three, SDF1-α variants NHis-SDF1 and CHis-SDF1 were successfully expressed under cell-free conditions and purified. NHis-SDF1 and CHis-SDF1, exchanged into buffer composed of 50 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM β-mercaptoethanol and 3 % glycerol, were concentrated to 5 mg/mL of protein in approximately 90 – 100 µL of volume and sent to the High-Throughput Crystallisation Laboratory (HTX Lab, PSB/EMBL, Grenoble) for crystallisation trials (Dimasi *et al.*, 2007). The crystallisation experiments were performed with a Cartesian PixSys 4200 crystallisation robot (Genomic Solutions, U.K.). Standard drops were set up with 100 nL of the sample mixed with 100 nL of precipitant solution. Drops were set in 96-well crystallisation screen plates using a range of precipitants: Classics Suite (Quiagen/Nextal), JCSG+ (Molecular Dimensions), PACT *premier* (Molecular Dimensions), PEGs-I (Quiagen/Nextal), Salt Grid (Hampton Research) and Wizard II&II (Rigaku Reagents). Crystallisation experiments were performed at 20 $^{o}$C.

### 5.2.2. X-ray diffraction and data collection

X-ray diffraction of the crystals was carried out at the ESRF on the "MASSIF-1" beamline ID30A-1 (Bowler *et al.*, 2015, Svensson *et al.*, 2015, Svensson *et al.*, 2018). The experiment and beamline parameters of CHis-SDF1 crystal X-ray diffraction data collection are shown below (Table 5.1). The NHis-SDF1 crystal did not lead to useful X-ray diffraction.

| Parameter | Value |
|---|---|
| **Number of images** | 1360 |
| **Wavelength** | 0.966 Å |
| **Energy** | 12.835 keV |
| **Flux** | 1.3 x $10^{11}$ photons/sec |
| **Oscillation range** | 0.1° |
| **Overlap** | 0° |
| **Exposure time/image** | 0.175 s |
| **Total exposure time** | 238 s |
| **Detector distance** | 200.82 mm |
| **Xbeam** | 128.91 mm |
| **Ybeam** | 146.86 mm |
| **Resolution edge** | 1.76 Å |
| **Resolution at corner** | 1.28 Å |

***Table 5.1:*** *Experiment parameters used for X-ray diffraction data collection of the CHis-SDF1 crystal at the MASSIF-1 beamline (ESRF).*

### 5.2.3. Data processing, molecular replacement and refinement

The automatic data processing workflow (Bowler *et al.*, 2015, Brockhauser *et al.*, 2012), based on the EDNA (Incardona *et al.*, 2009), greNADES (Monaco *et al.*, 2013) processing software, was used to process the X-ray diffraction data frames of the CHis-SDF1 crystal.

Molecular replacement was performed using MOLREP (Vagin and Teplyakov, 2000) as implemented in the CCP4i software package (Potterton *et al.*, 2003) using the SDF1-α crystal structure with PDB accession code 3GV3 (Murphy *et al.*, 2010) as the search probe. Subsequent refinement was done with REFMAC5 (Murshudov *et al.*, 2011, Murshudov *et al.*, 1997), also as implemented in CCP4i, and using WinCOOT (Emsley and Cowtan, 2004) for map and structure visualisation and manual adjustment of the protein model. Additional visualisation software was used which were CCP4MG (version. 2.10.10) and DeepView/Swiss-PdbViewer.

## 5.3. RESULTS

### 5.3.1. Obtaining diffracting crystals of SDF1-α

Single crystal formation was observed for CHis-SDF1 (Figure 5.5) and NHis-SDF1 (Figure 5.6) in 0.1 M tri-sodium citrate pH 5.6, 20 % (v/v) isopropanol and 20 % (w/v) PEG 4000 after around 60 and 90 days respectively. Crystals were harvested after 90 days and were cryo-cooled in preparation for X-ray diffraction. Useful X-ray diffraction was obtained with only CHis-SDF1 crystals (Figure 5.7 and Table 5.2). NHis-SDF1 crystal gave poor diffraction which could not be interpreted (not shown).

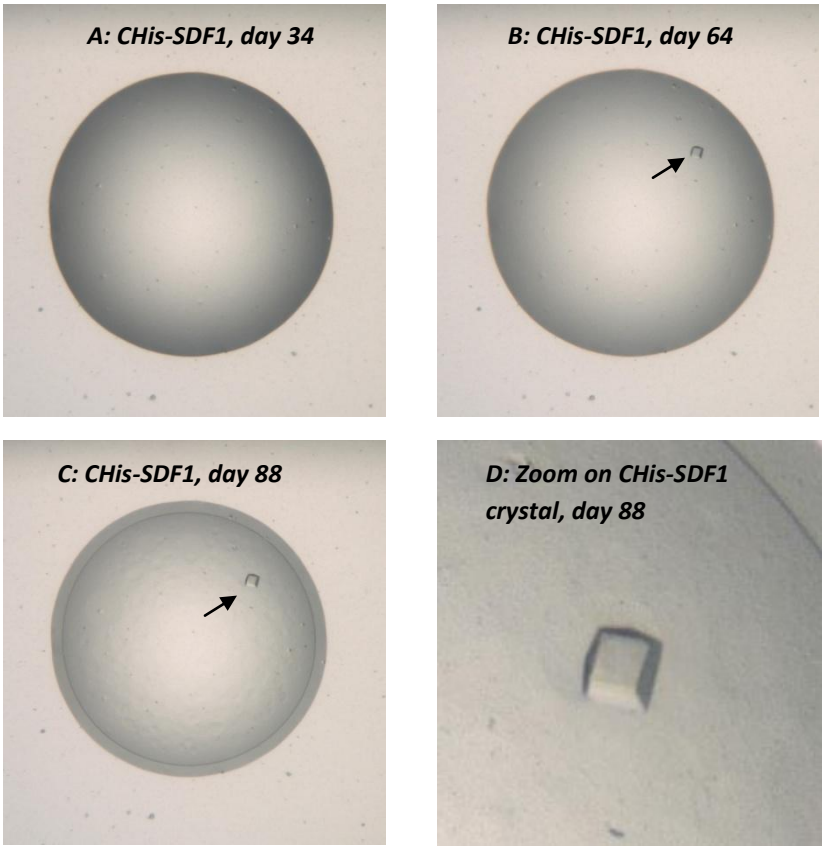



*Figure 5.5: Photographs of CHis-SDF1 (A – D) in 0.1 M tri-sodium citrate pH 5.6, 20 % (v/v) isopropanol and 20 % (w/v) PEG 4000 following crystallisation plate setup. CHis-SDF1: A, day 34; B, day 64; C, day 88; D, zoom on crystal. Arrows indicate crystals.*

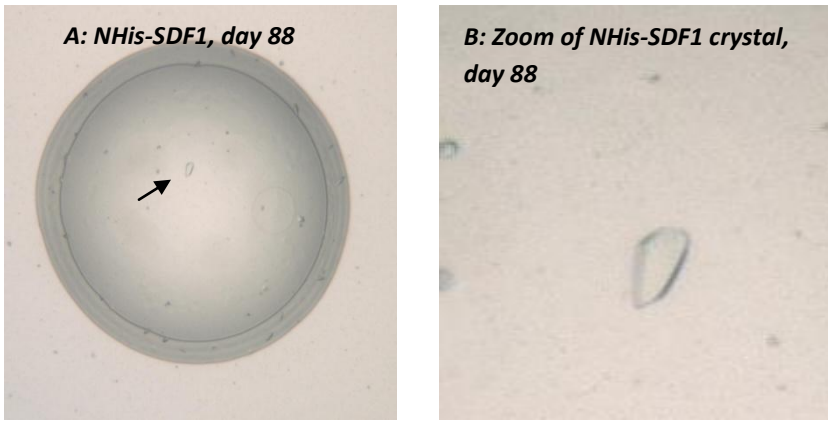



***Figure 5.6:*** *Photographs of NHis-SDF1 in 0.1 M tri-sodium citrate pH 5.6, 20 % (v/v) isopropanol and 20 % (w/v) PEG 4000 following crystallisation plate setup. A, day 88; B, zoom on crystal. Arrow indicates the crystal.*

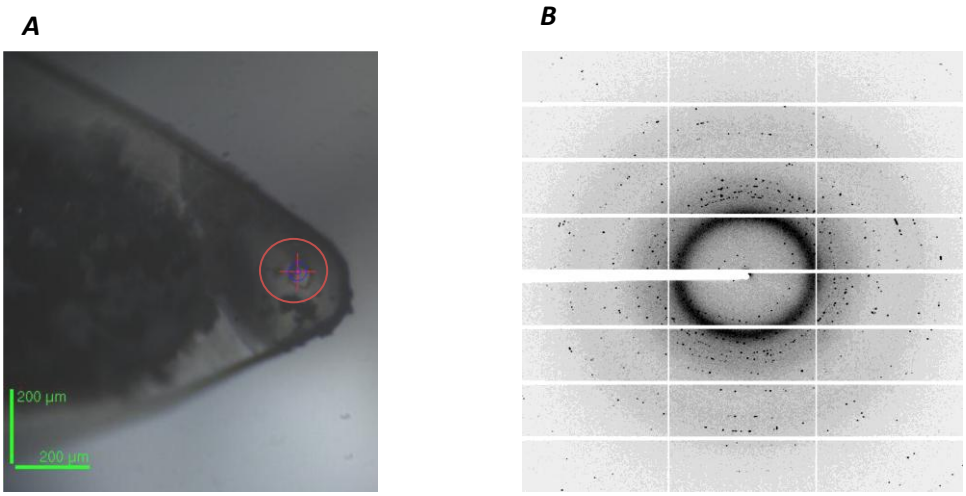



***Figure 5.7:*** *X-ray diffraction of CHis-SDF1 crystal. (A) Cryo-cooled crystal of CHis-SDF1 (pink reticule, indicated by crosshair) prior to X-ray diffraction mounted on ESRF beamline ID30A-1 using MASSIF-1 instrumentation (Bowler et al., 2015, Svensson et al., 2015, Svensson et al., 2018); B, snapshot of diffraction pattern of CHis-SDF1 crystal.*

### 5.3.2. Description of the structure of CFPS SDF1-α

The structure of CFPS CHis-SDF1-α was determined by molecular replacement, using the SDF1 structure with accession code 3GV3 from RCBS PDB as the reference model (Figure 5.8 and Table 5.2). Similarly to 3GV3, CFPS CHis-SDF1-α occurs in the $P3_221$ crystal form, with unit cell a = b = 55.86 Å and c = 45.34 Å, and with a single monomer in the asymmetric unit.

As with previously published structures of SDF1-α derived from *E. coli* expression (Ryu *et al.*, 2007, Murphy *et al.*, 2010), the typical chemokine topology consisting of three anti-parallel β-strands (β1 – β3) followed by α-helix is observed (Figure 5.7A). CFPS CHis-SDF1-α consists of a long and extended N-terminal; the β-strands, which are comprised of residues 22 – 29, 37 – 43 and 47 – 51; and the α-helix, which is made up of residues 57 – 67. The crystal structure of CFPS CHis-SDF1-α furthermore shows that the disulphide bonds (Cys9 – Cys34 and Cys11 – Cys50) are present. The application of crystal symmetry operations allows the visualisation of the characteristic CXC family chemokine dimer (Figure 5.9A) (Clark-Lewis *et al.*, 1995). The dimer of SDF1-α formed by the interaction of the anti-parallel β1-strands of each sub-unit, includes hydrogen bonds (Figure 5.9B, C, D). With a saddle-shape appearance, the dimer can be considered as having two surfaces, with six β-strands forming the saddle, and two α-helices the bottom layer (Figure 5.9C). As described in previous publications, these features are important structurally as they are the localisation sites for the binding of heparin sulphate (Murphy *et al.*, 2007) and the N-terminal of CXCR4 (Veldkamp *et al.*, 2008a).

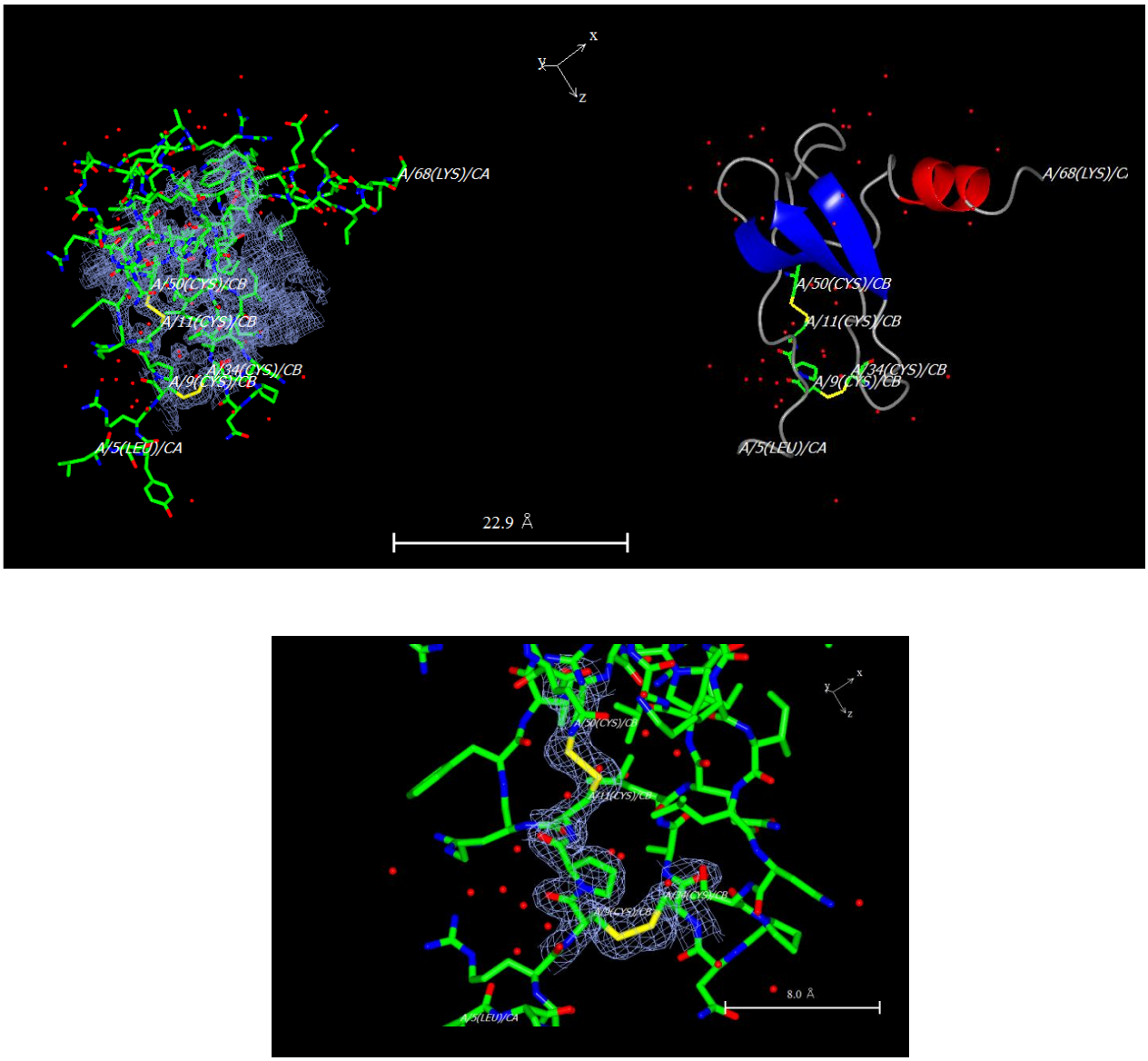






***Figure 5.8:*** *X-ray crystal structure of CFPS SDF1-α (A) Protein chain of SDF1-α with overlay of the $2F_o$-$F_c$ electron density map (contoured at 1 sigma), showing Leu5 at N-terminal, Asn67 at C-terminal, cysteine residues involved in disulphide bonds. (B) Ribbon diagram of CFPS SDF1-α, with α-helices in red, and β-sheets in blue. Red dots represent water molecules. (C) Close-up view of the disulphide bridges showing the $2F_o$-$F_c$ electron density map (contoured at 1 sigma). On the protein chains, oxygen atoms are shown in red, nitrogen atoms in blue and disulphide linkages in yellow. Red dots represent water molecules.*

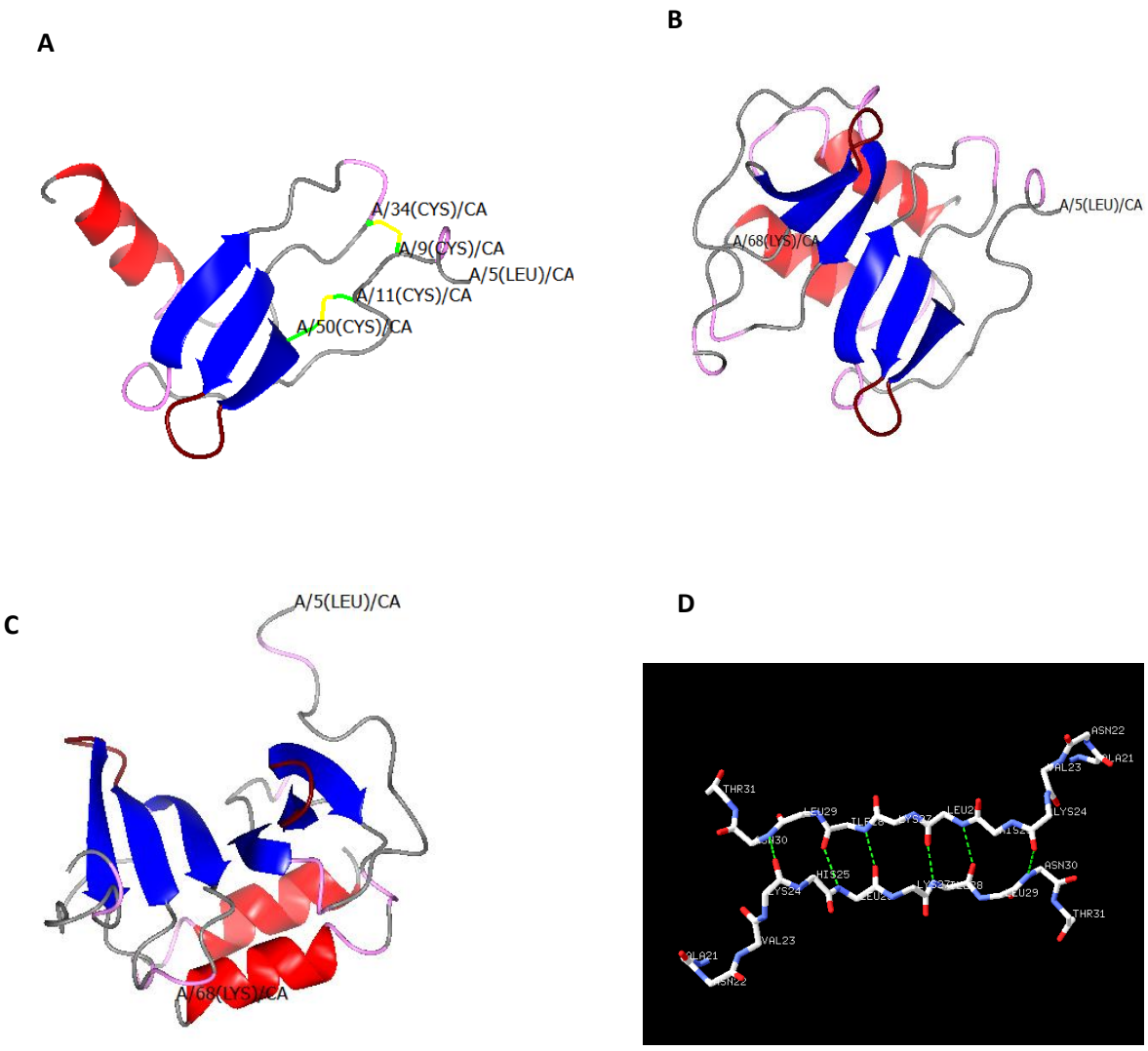



***Figure 5.9:*** *Ribbon diagrams of* CFPS *CHis-SDF1-α, coloured according to secondary structure (blue: β-stands, red: α-helix), showing N-terminal leucine and C-terminal lysine residues. A: monomer of* CFPS *CHis-SDF1-α, showing the presence of the two expected disulphide bonds. B and C are dimeric forms of* CFPS *CHis-SDF1-α. D: zoom on the β1-stands in* CFPS *CHis-SDF1-α dimer showing hydrogen bonds between main carbon chain.*

| Data Collection | |
|---|---|
| **Space-group** | $P3_2\ 2\ 1$ |
| **Unit cell dimensions** | |
| ***a, b, c (Å)*** | 55.86, 55.86, 45.34 |
| ***α, β, γ (°)*** | 90, 90, 120 |
| **Overall resolution (Å)** | 48.49 - 1.74 (1.83 – 1.74) |
| **Number of unique reflections** | 9148 |
| $R_{merge}$ ***(%)*** | 6.2 (96.8) |
| $R_{p.i.m.}$ ***(%)*** | 2.4 (37.6) |
| $R_{meas}$ ***(%)*** | 6.7 (104.0) |
| **I/σ(I)** | 13.8 (1.9) |
| **$CC_{1/2}$** | 0.999 (0.668) |
| **Completeness (%)** | 99.9 (99.9) |
| **Mosaicity (°)** | 0.18 |
| **Multiplicity** | 7.2 (7.5) |
| **Refinement** | |
| **Resolution limits (Å)** | 33.196 – 1.745 |
| **Number of reflections** | 8684 |
| **Number of used reflections** | 8277 |
| **R-*factor*/ R-*free* *(%)*** | 18.8/24.0 |
| **Number of atoms** | |
| ***Total number of atoms*** | 583 |
| ***Number of protein atoms*** | 537 |
| ***Number of water atoms*** | 46 |
| **B-factor ($Å^2$)** | |
| ***Protein*** | 42.0 |
| ***Water*** | 53.9 |
| **RMS deviation** | |
| ***Bond length (Å)*** | 0.24 |
| ***Bond angles (°)*** | 2.25 |
| **Ramachandran statistics (number of residues/%)** | |
| ***In preferred regions*** | 60.0/98.4 |
| ***In allowed regions*** | 1.0/1.6 |
| ***Outliers*** | 0/0 |

***Table 5.2:*** *Crystal data, collection and refinement statistics. Values in parentheses are for the highest resolution shell.*

### 5.3.3. Comparison of the structure of CFPS SDF1-α to SDF1-α produced by other systems

Upon comparison to the representative crystal structure of SDF1-α obtained from refolded *E. coli* expressed protein (PDB ID: 3GV3), the structures of CFPS SDF1-α were found to be morphologically similar and to display identical secondary structure (Figure 5.10, 5.11, Table 5.3). The CFPS structure contains two additional residues which could be modelled at the N-terminal of the protein, but, as is typical in X-ray structures, the His-tag could not be located in the electron density map and was therefore not modelled. The overall root-mean square displacement (RMSD) of the superimposed carbon-α atoms (62 atoms) and of the protein backbone atoms (248 atoms) of *E. coli* SDF1-α to that of CFPS SDF1-α was found to be both 0.39 Å. Inspection of the RMSD per residue plot indicated generally low RMSD values (< 0.30 Å) throughout the structure of CFPS SDF1-α compared to *E. coli* SDF1-α (Figure 5.12). Particularly high RMSD values were recorded for residues Leu5, Ser6 and Lys56. Leu5 and Ser6 are located at the beginning of the molecule and lack well-defined electron density and have correspondingly high *B*-factors in both the representative and CFPS crystal structures. In the reference *E. coli* SDF1-α model, the side-chain of Lys56 was not modelled (presumably due to lack of electron density) and therefore, leads to the high RMSD value here.

Overall, the structures of SDF1-α derived from various sources show similar morphology to CFPS SDF1-α (Figure 5.10, 5.11, Table 5.3). Differences in the local conformation of between CFPS SDF1-α and eukaryotically expressed SDF1-α (PDB ID: 1QG7, 2.00 Å) were most significant, with the RMSD of the superimposed carbon-α atoms (61 atoms) of 1.25 Å. Synthetic SDF1-α, compared to CFPS SDF1-α (PDB ID: 1A15, 2.20), had intermediate RMSD difference value of 1.06 Å (59 carbon-α atoms).

```
>CFPE SDF1-α
----LSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALNK

>3gv3 (E.coli refolded)
----LSYRCPCRFFESHIARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKAL—

>1a15 (Synthetic, chain A)
KPVSLSYRCPCRFFESHVARANVKHLKILNTPACALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN-

>1a15 (Synthetic, chain B)
-------RCPCRFFESHVARANVKHLKILNTPACALQIVARLKNNNRQVCIDPKLKWIQEYLEK-----

>1qg7 (Eukaryote, chain A)
-----SYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN-

>1qg7 (Eukaryote, chain B)
-PVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
```

***Figure 5.10:*** *Alignment of SDF1 structures derived from different expression hosts. β-strands are shown blue and the α-helix in red. Green indicates point mutation difference between the two sequences which was present in the E. coli refolded SDF1-α.*

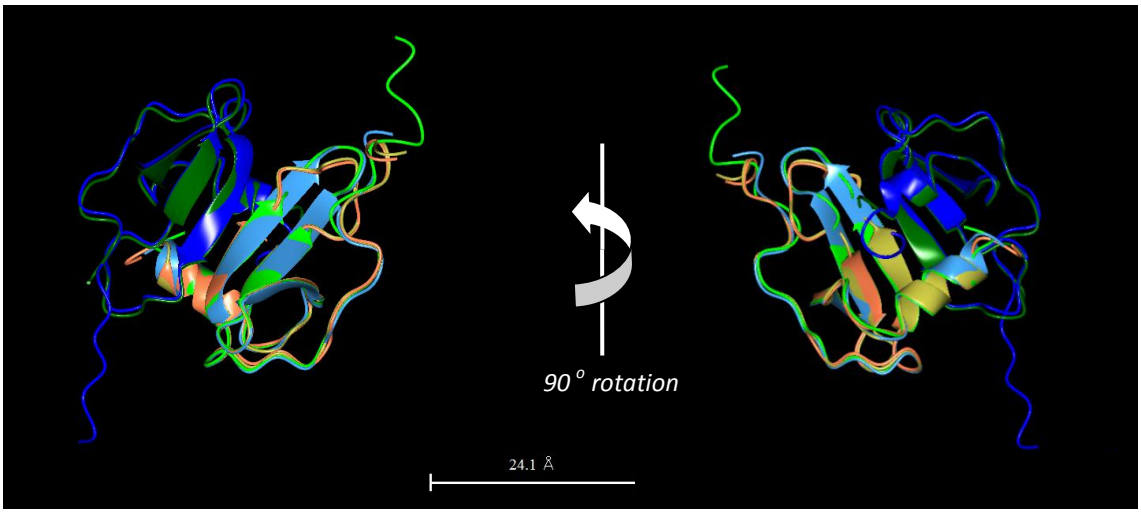



***Figure 5.11:*** *Super-position of SDF1 structures derived from different expression host: CFPE (orange), refolded E.coli (yellow; PDB: 3VG3), eukaryote (pale blue, chain A; dark blue, chain B; PDB: 1QG7) and synthetic (pale green, chain A; dark green, chain B; PBD: 1A15).*

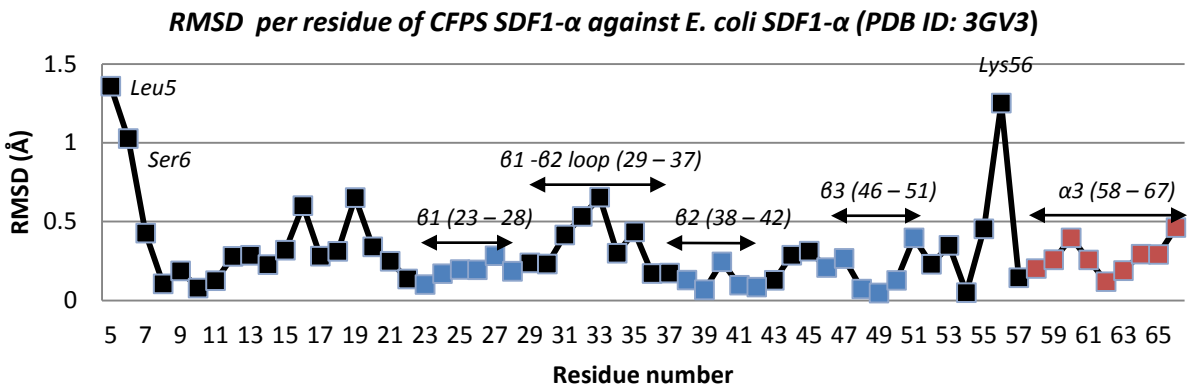



*Figure 5.12: RMSD per residue (all atoms) of CFPS SDF1-α against the SDF1-α crystal structure with highest resolution (1.6 Å, E. coli refolded, PDB ID: 3GV3). β-strands are indicated in blue and the α-helix in red. Also labelled are peaks with high RMSD.*

| Criteria | CFPS SDF1-α | *E. coli* SDF1-α (PDB: 3GV3) | Eukaryote SDF1-α (PDB: 1QG7)[a] | Synthetic SDF1-α (PDB: 1A15)[a] |
|---|---|---|---|---|
| **Space group** | $P3_221$ | $P3_221$ | $P2_12_12_1$ | $P2_12_12_1$ |
| **Resolution (Å)** | 1.74 | 1.60 | 2.00 | 2.20 |
| **No. of residues in construct** | 74 | 63 | 89 | 67 |
| **No. of residues in model** | 64 | 62 | 62/66 | 67/57 |
| **Residues composing secondary structure elements:** | | | | |
| **β1** | 23 – 28 (6) | 23 – 28 (6) | A: 23– 31 (9)/ B: 23 – 29 (7) | A: 23 – 29 (7)/ B: 23 – 29 (7) |
| **β2** | 38– 42 (5) | 38– 42 (5) | A: 35 – 42 (8)/ B: 38– 42 (5) | A: 37– 42 (6)/ B: 38– 42 (5) |
| **β3** | 46 – 51 (6) | 46 – 51 (6) | A: 47 – 51 (5)/ B: 47 – 51 (5) | A: 46 – 51 (6)/ B: 46 – 51 (6) |
| **α** | 58 – 67 (10) | 58 – 65 (8) | A: 55 – 66 (11)/ B: 55 – 59 (5) | A: 58 – 64 (7)/ B: 58 – 61 (4) |
| **RMSD[b]/Å** | | | | |
| **Vs. CFPS SDF1-α** | Reference | 0.39 (62) | 1.25 (61) | 1.06 (59) |
| **Vs. *E. coli* SDF1-α** | 0.39 (62) | Reference | 1.12 (59) | 1.03 (59) |
| **Vs. Eukaryote SDF1-α** | 1.25 (61) | 1.12 (59) | Reference | 0.57 (118) |
| **Vs. Synthetic SDF1-α** | 1.06 (59) | 1.03 (59) | 0.57 (118) | Reference |

*Table 5.3: Summary of crystal structure and secondary structure elements of the X-ray crystal structures of SDF1-α derived from CFPS and traditional E. coli expression systems. Values in parentheses give the number of residues between strands. [a]: Chains A and B are included as the crystal structure is a dimer in the asymmetric unit. [b] :Values in parentheses give the number of C-α atoms considered for RMSD.*

## 5.4. DISCUSSION

### 5.4.1. CFPS: alternative system for the production of SDF1-α for crystallisation

Over the years, the production of SDF1-α for the purposes of X-ray structure determination has entailed the use of several expression techniques. Resolved to 2.20 Å resolution, chemically synthesised SDF1-α led to the first reported crystal structure of SDF1-α (Dealwis *et al.*, 1998). However, chemical synthesis at the time was capable of efficiently producing peptides only up to 40 – 50 amino acids. To cope with this problem, the N-terminal and C-terminal peptides of SDF1-α were first generated and then ligated to each other in order to produce the entire molecule (Ueda *et al.*, 1997). Nevertheless, in order to permit the ligation, chemically synthesised SDF1-α had to include the mutation of Asp 34 to Ala 34, which deviates the molecule from the wild-type. Chemical synthesis of proteins, particularly for the purposes of X-ray crystallography, remains a challenging method. The production of longer proteins results in aggregation of the peptide, steric crowding, uncoupling of ligation reactions and the accumulation of unwanted truncated by-products. This eventually results in low recovery and purity of the full-length proteins, and high expense.

Resolved to 2.00 Å resolution, SDF1-α was also produced in eukaryote cells using viral vectors expressing human SDF1-α (Ohnishi *et al.*, 2000). Typical of recombinant eukaryotic cell expression, soluble and bioactive SDF1-α was secreted directly into the supernatant and could be recovered easily. The challenge of the expression in mammalian cell, however, remains the need for specialised laboratory, equipment and trained personnel. Asides from the heavy investment costs, the running costs of the eukaryotic laboratory remain quite high due the use of expensive reagents and consumables. Additionally, the setting up of eukaryotic laboratories, involving the use of viruses, requires adherence to strict health, safety and ethics regulations. These disadvantages can limit the industrial application of eukaryote expression systems for the production of proteins for structural biology.

SDF1-α, eventually resolved to 1.60 Å resolution, was derived from classical *E. coli* recombinant expression (Murphy *et al.*, 2010). SDF1-α in *E. coli* is the cheapest (in terms of consumables, but neglecting time) and highest yielding method when considering only the expression stage of the protein preparation. However, SDF1-α is expressed as insoluble and aggregated protein the form of inclusion bodies. As a result, a labour-intensive and time-consuming extraction and refolding protocols are

required for the recovery of soluble and bioactive SDF1-α. Furthermore, refolding protocols need to be very well optimised in order to guarantee the quality and yield of the output refolded protein.

Here, CFPS was successfully and efficiently used to produce SDF1-α for the purposes of generating a crystal structure. As described in the discussion of chapter four, CFPS presents several advantages over the other modes of expression. Namely, the short time required and the ease of production are key aspects which put CFPS ahead. Soluble and bioactive SDF1-α could be produced directly in the CFPS reaction, be immediately purified and thus within two days, be ready for crystallisation.

### 5.4.2. CFPS SDF1-α monomer show strong resemblance to previous SDF1-α crystals

Obtaining a crystal structure of SDF1-α has provided a confirmation of the application of CFPS in X-ray crystallography; a path which Synthelis has been striding for. Another goal was to validate the quality of the CFPS product using the X-ray crystal stucture. Here, with an RMSD of between only 0.38 – 1.23 Å, it was shown that CFPS SDF1-α very closely resembles previously published SDF1-α crystal structures, which have been derived from *E. coli*, eukaryotic and synthetic sources. CFPS SDF1-α showed highest structural similarity to the refolded *E. coli* SDF1-α (PDB: 3GV3), which currently is the crystal structure of SDF1-α with the highest resolution (1.60 Å).

## 5.5. CONCLUSIONS

Obtaining a high resolution X-ray crystal structure of CFPS SDF1-α has validated that the CFPS system as developed by Synthelis can express proteins in suitable quality and quantity for protein structural biology. The structure is essentially identical to those produced by classical methods. Furthermore, the results of this chapter demonstrate the applicability of the CFPS system as an accessible tool which can be compatible for the expression of proteins for exploitation in structural biology.

# CHAPTER SIX

## SMALL-ANGLE NEUTRON SCATTERING OF CFPS-PRODUCED 2DCD4 IN COMPLEX WITH GP120

## ABSTRACT

This chapter focuses on solution structural studies using SANS of the complex formed by deuterated 2dCD4 and GP120 dimer. The project faced key challenges at the outset. Firstly, with the use of GP120 dimer, it can be difficult to assure the homogeneity of the sample can to ensure useful structural information can be extracted from the SANS data. Moreover, GP120 is a highly glycosylated protein (37 % by weight) and the glycans have a slightly different match-out level to proteins. Moreover, glycans are typically highly flexible components of protein structures which can make the interpretation of SANS data challenging. The objective of this preliminary study was therefore to gain insights into the limitations and challenges of the exploitation of CFPS produced labelled 2dCD4 in the context of a challenging protein complex.

Two types of CD4-GP120 complexes were formed: firstly using fully hydrogenated proteins, and secondly complexes in which the 2dCD4 component was deuterated. This allowed SANS studies to be used in conjuction with solvent contrast variation to visualise (i) the entire complex, (ii) the complex where the GP120 was matched out (rendered “invisible”) and (iii) the complex where the 2dCD4 was matched out. Whilst SANS analysis showed that CFPS-produced deuterated 2dCD4 was effectively matched-out, its signal in non-matched conditions was very low, which limited the interpretation of the data. Structural interpretation from the complex was mostly dominated by data recorded for the GP120 dimer. While extraction of the basic structural parameters was possible using these data, advanced modelling using known crystal structures was compromised due to the poor signal-to-noise in the available data. Nonetheless, an initial model of glycosylated GP120 dimer was generated.

## 6.1. INTRODUCTION

### 6.1.1. Demonstrating the use of deuterated CFPS products for SANS

CFPS is a powerful technique that enables the rapid recombinant expression of proteins. Biophysical characterisation and functional assessments provide powerful information on the structural and functional integrity of the final purified CFPS products – measures that in the context of industry are relatively standard and used on a daily basis. However, to demonstrate the quality of CFPS products as equivalent to those produced in classical *in vivo* expression systems, advanced structural methods can add significant value and confidence in the product delivered. Moreover, as described in chapters three and four, CFPS was shown to be an advantageous tool in the production of deuterated proteins, which are otherwise typically produced *in vivo in E. coli* or *P. pastoris* expression systems (Dunne *et al.*, 2017b, Laux *et al.*, 2008, Haertlein *et al.*, 2016). CFPS-produced deuterated 2dCD4, for which the production is described in chapter three, was extensively produced, biophysically and functionally assessed, and used successful for SANS analysis previously (Channell, 2018). Driven by the success of these experiments, it was therefore sought to analyse the 2dCD4-GP120 complex using CFPS deuterated 2dCD4 in complex with hydrogenated GP120. These experiments were of interest in establishing the use of CFPS as an efficient system for routine application for the production of deuterated protein for of SANS studies of biomolecular complexes.

### 6.1.2. The use of SANS for structural characterisation of 2dCD4-GP120 complex

The crystal structure of the complex shows that CD4 is inserted into a depression in the GP120 (Kwong *et al.*, 1998, Liu *et al.*, 2008, Diskin *et al.*, 2010, Chen *et al.*, 2009). This depression is formed at the interface of the inner and outer domains of GP120, with a bridging sheet stretching between the two domains. From these crystal structures, the critical residues Phe 43 and Arg 59 on CD4 were identified. These are involved in a number of contacts with well-conserved residues of GP120: Asp 368, Glu 370 and Trp 427.

Despite having much lower resolution than X-ray crystallography, the key advantage of small-angle scattering (SAS) lies in its ability to analyse the native CD4-GP120 complex while in solution. Typically, crystallisation of the CD4-GP120 complex is only possible with the use of a truncated "core" of GP120, which lacks a number of flexible loops and is devoid of glycosylation. By concealing conserved antibody recognition sites on the GP120 core, these loops and sugar decorations are critical in HIV's immune escape mechanisms (Sanders *et al.*, 2008, Doores *et al.*, 2010). Furthermore, it has been suggested that glycosylation assists in strengthening the GP120-CD4 interaction (Dirckx *et al.*, 1990, Li *et al.*, 2008, Wilhelm *et al.*, 2012). As such, SAXS was used to seek information on conformation changes in native GP120 induced by CD4 binding (Ashish *et al.*, 2006). These changes can be ascribed to the conformational changes in GP120 leading to the exposure of CD4-induced epitopes on GP120 (Sullivan *et al.*, 1998, Zhang *et al.*, 1999, Sattentau and Moore, 1991). With regards to CD4, conformational changes have been described in the molecule upon GP120 binding (Denisova *et al.*, 1997). This leads to the exposure of previously concealed epitopes on CD4 that are recognised by anti-CD4 antibodies in the GP120-CD4 complex (Denisova *et al.*, 1997). However, these changes in CD4 are difficult to identify and studies have mostly focused on the significant changes in GP120. Therefore the objectives of the SANS with contrast variation studies were targeted to gain further insights into the specific conformational changes of the individual CD4 and GP120 components which form the complete complex.

Previous work on the GP120 and CFPS deuterated 2dCD4 had identified that the match points of these were respectively at 45 % and 92 % of $D_2O$ in the solvent (Channell, 2018). Therefore, using contrast variation and selective deuteration, SANS could be used to gain insight into the complex consisting of deuterated 2dCD4 and GP120. Hence SANS studies of 2dCD4-GP120 complex in solvent conditions of 0 %, 45 % and 92 % $D_2O$ would respectively provide scattering information arising from the entire complex, 2dCD4 alone and GP120 alone (Figure 6.1).

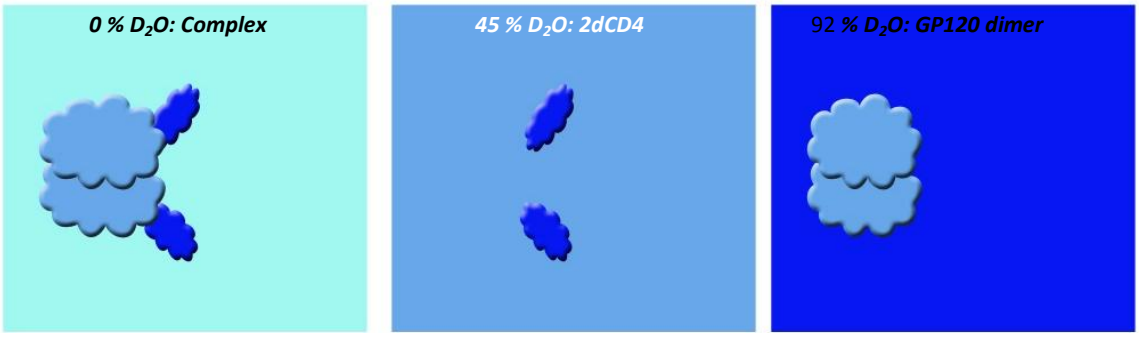



*Figure 6.1: Diagram of experimental strategy used for SANS studies of the 2dCD4-GP120 complex in different solvent contrast regimes. Left: the entire complex should be visible in buffer conditions of 0 % $D_2O$. Centre: only 2dCD4 should be visible in buffer conditions of 45 % $D_2O$. Right: only GP120-dimer would be visible at 92 % $D_2O$.*

## 6.2. MATERIALS AND METHODS

### 6.2.1. Preparation of 2dCD4-GP120 complex for SANS

Deuterated 2dCD4 complexed to dimeric GP120 was prepared in hydrogenated buffer (50 mM Tris.HCl pH 7.5, 300 mM NaCl, 5 % sucrose) as described in chapter three. Using ProtParam (https://web.expasy.org/cgi-bin/protparam/), the molar extinction coefficient of the proteins, which was used to calculate the complex concentration, was determined. Following SEC, samples consisting of dimeric GP120 complexed to deuterated 2dCD4 were pooled and concentrated to approximately 10 mg/mL in 600 µL using a 100 kDa MWCO Amicon® Ultra-4 centrifugal filter unit (Merck Millipore). The sample was stored on ice until SANS analysis (approximately 72 hours).

The 600 µL of GP120-CD4 complex was split into two aliquots of 300 µL each. The first 300 µL aliquot, consisting of GP120-CD4 complex in 0 % $D_2O$/100 % $H_2O$ buffer, was further maintained on ice until SANS analysis. This sample, in which the entire 2dCD4-GP120 complex would be visible under SANS analysis, is referred to as *SampleComplex*. The second 300 µL sample was extensively dialysed (12 hours) against two sequential fresh 300 mL batches of cold 45 % deuterated buffer (50 mM Tris.HCl pH 7.5, 300

mM NaCl, 5 % sucrose, in 45 % $D_2O$/55 % $H_2O$). This sample, in which GP120 would be matched-out and only 2dCD4 would be visible, was called *Sample2dCD4*. Following SANS analysis of *SampleComplex*, *Sample2dCD4* was analysed by SANS. *SampleComplex* was then was extensively dialysed (12 h) against two sequential fresh 300 mL batches of cold 92 % deuterated buffer (50 mM Tris.HCl pH 7.5, 300 mM NaCl, 5 % sucrose, with 92 % $D_2O$/8 % $H_2O$). This sample, in which 2dCD4 would be matched-out and only GP120 would be visible, was called *SampleGP120*.

However, due to a very low signal recorded from *Sample2dCD4*, all of the remaining GP120-CD4 samples from SANS measurements were pooled, concentrated to 300 µL volume using a 100 kDa MWCO Amicon® Ultra-4 centrifugal filter unit (Merck Millipore), dialysed extensively against 45 % deuterated buffer and subject to SANS analysis (*Sample2dCD4-conc*). Immediately prior to each SANS analysis, the concentration of sample was determined using a NanoDrop® ND-1000 UV-Vis Spectrophotometer (ThermoScientific).

### 6.2.2. SANS of 2dCD4-gp120 complex at the KSW-2

SANS measurements were carried out the Forschungs-Neutronenquelle Heinz Maier-Leibnitz (FRMII, Munich, Germany) using the KWS-2 instrument (Houston *et al.*, 2018, Radulescu *et al.*, 2016). The neutron wavelength was set to 5 Å. The samples were measured in 1-mm-path-length Suprasil quartz cuvettes (Hellma). Data were recorded at collimation lengths of 8 m, 20 m and 8 m, with respective sample-to-detector distances of 1.675 m, 7.675 m and 7.675 m to provide a full q-range (0.001 Å – 0.5 Å) from the Guinier region of the proteins to the solvent/lysate level of incoherent scattering. The raw data were reduced using local QtiKWS software (http://iffwww.iff.kfa-juelich.de/~pipich/dokuwiki/doku.php/qtikws). Data reduction included thickness and transmission scaling, empty cell and blocked beam subtractions, calibration to absolute intensity using incident flux measured at sample position, azimuthal averaging and merging to produce neutron scattering data.

### 6.2.3. Treatment of neutron scattering data and model building

The neutron scattering data recorded from low, intermediate and high q measurements were merged using National Institute for Science and Technology Centre for Neutron Research SANS reduction macros on IGORpro (Kline, 2006). The neutron scattering data was generated and visualised using the same software, as well as PRIMUS from the ATSAS suite (Petoukhov *et al.*, 2012, Konarev, 2003) and scÅtter (Robert Rambo, Diamond Light Source, Didcot, UK). The radius of gyration and distance distribution functions for the corresponding model protein were estimated using automated auto*Rg* and autoGNOM functions (Svergun, 1992) from the ATSAS suite. Provided by AutoGNOM, an estimate of the quality of P (r) plots based on the fitting of points against experimental scattering data, are given as well: with 0 being poor quality and 1 being ideal quality. Where required, curve fitting for the determination of *Rg* and distance distribution was performed manually. For the calculation of scattering-derived molecular weights (MW), equation 2.4 was used. The value of $I_o$ was given by auto*Rg*. Neutron scattering contrasts (Δρ) of the complex, 2dCD4 and GP120 dimer were determined using the MULch online programme (http://smb-research.smb.usyd.edu.au/NCVWeb/input.jsp). The glycoproteins present on GP120 were accounted for and based on a previous publication in which the entire glycoproteome pattern on GP120 was determined (Panico *et al.*, 2016).

A chimeric monomer model of glycosylated GP120 in the CD4 bound state (*GlyGP120-b)* was constructed using DeepView/Swiss-PdbViewer (http://www.expasy.org/spdbv/). The chimeric model was a fusion of GP120 PDB models: 3J70 (Rasheed *et al.*, 2015), of which the unglycosylated V1/V2, N-terminal and V3 loops of GP120 were considered; 4RQS (Kong *et al.*, 2015), of which the fully glycosylated GP120 core (lacking the V1/V2, N-terminal and V3 loops) was considered; and 3U4E (McLellan *et al.*, 2011), of which the glycosylated V1/V2 loop was considered. *GlyGP120-b* was generated by fitting the glycosylated V1/V2 loop from PDB model 3U4E over the V1/V2 loop of PDB model 3J70, which corresponds to that of GP120 bound to CD4.

CRYSON was used to generate calculated neutron scattering data and to fit curves for these models against experimental data (Svergun *et al.*, 1998). Quaternary model generation and simultaneous fitting against experiment data, SASREF was used (Petoukhov and Svergun, 2005). SASREF is a programme that carries out quaternary structure modelling of complexes formed by subunits of known atomic structure, and fits the data generated against the experimental neutron scattering data.

## 6.3. RESULTS

### 6.3.1. Interpretation of neutron scattering profiles of deuterated 2dCD4-gp120 complex

#### *Overview*

The recorded 1D neutron scattering data from the 2dCD4-GP120 complex, the 2dCD4, and GP120, and their corresponding buffers, are shown in Figure 6.2, together with the buffer-subtracted curves. A large difference can be observed between the SANS curves of the *SampleComplex* and *SampleGP120*. Very low signal intensity was obtained for *Sample2dCD4-conc*, despite the use of concentrated sample of the complex (data from *Sample2dCD4* of lower concentrations not shown). This was due to the fact that the 2dCD4 is rather small and to the fact the effective concentration of the 2dCD4 contribution to the SANS data at the GP120 match point was low (Maximum absolute intensity, $I_o$ = 0.04 $cm^{-1}$ – see Table 6.1).

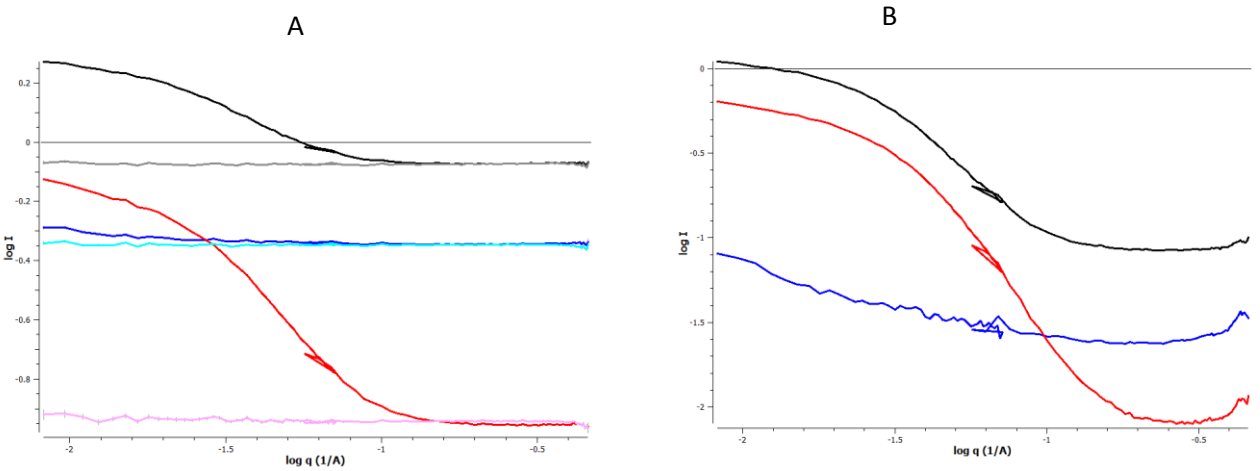



***Figure 6.2:*** *log I (q) vs log q plots of 2dCD4-GP120 complex in 0 %, 45 % and* 92 *% $D_2O$ buffers (A) and buffer subtracted plots (B). Black and grey; complex in 0 % buffer $D_2O$ buffer and its buffer – neutron scatter data from entire complex. Blue and cyan; complex in 45 % buffer $D_2O$ buffer and its buffer – neutron scatter data from 2dCD4 only. Red and pink; complex in* 92 *% buffer $D_2O$ buffer and its buffer – neutron scatter data from GP120 dimer.*

### *Analysis of size parameters from SANS data recorded for the CD4-GP120 complex*

The samples measured and their corresponding size parameters as determined from the SANS data, are summarised below:

| Sample of 2dCD4-GP120 | Concentration of sample[a] (mg/mL) | | | | *Rg* Reci.[d] (error) (Å) | Abso-lute $I_o$ (/cm) | Δρ[e] ($10^{10}$ /cm$^2$) | MW (TM)[f] (kDa) | *Rg* Real[g] (Å) | Dmax[h] (Å) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Measured[b] | Actual[c] | GP120 dimer | 2dCD4 | | | | | | |
| ***SampleComplex*** | 10.6 | 8.8 | 7.0 | 1.8 | 47 (2) | 1.12 | 2.8 | 176 (215) | 49 | 196 |
| ***Sample2dCD4*** (low conc)[i] | 9.0 | 7.4 | 6.0 | 1.5 | NA | NA | NA | NA | NA | NA |
| ***SampleGP120*** | 5.3 | 4.4 | 3.5 | 0.9 | 47 (4) | 0.62 | 3.4 | 196 (176) | 47 | 147 |
| ***Sample2dCD4-conc*** (high conc.)[j] | 17.8 | 14.7 | 11.8 | 2.9 | 19 (10) | 0.04 | -3.1 | 14 (22) | 22 | 74 |

***Table 6.1:*** *Summary of parameters derived from 1D neutron scatter data of 2dCD4-GP120 complex in in 0 %, 45 % and* 92 *% $D_2O$ buffers. [a]: Taken immediately prior to SANS analysis; [b]: Value measured; [b]: Actual concentration of sample, with correction for molar extinction coefficient of the complex (determined using Protparam as 1.21 /M cm) [d]: Guinier approximation derived Rg; [e]: Contrast of corresponding complex and components of complex (GP120 or 2dCD4) as determined using MULch; [f]: Theoretical molecular weight in brackets; [g]: P(r) derived Rg value using autoGNOM software; [h]: P(r) derived Rg value using autoGNOM software; [i]: Low concentration sample from which 2dCD4 signal was too low for interpretation; [j]: High concentration sample for repeat of 2dCD4 measurement.*

As determined on the basis of the Guinier approximation and from the P(r) function, the *Rg* values of *Samplecomplex* (*Rg* Guinier = 47 Å, *Rg* P(r) = 49) were consistent with each other. Similarly, the *Rg* values *SampleGP120* (*Rg* Guinier = 47 Å, *Rg* P(r) = 47), as determined on the basis of the Guinier approximation and from the P(r) function, were found to be consistent with each other. As such, they reinforce the confidence in the reliability of those values. Furthermore, the Guinier region for these two samples showed strong linearity, indicating the absence of aggregation (Figure 6.3).

From the a previous SANS study, the Guinier *Rg* and P(r)-derived *Rg* value for 2dCD4 was determined to be respectively 26 and 22 Å (Channell, 2018). Furthermore, a previous publication which used SAXS to study 2dCD4, the GP120 dimer and the 2dCD4-GP120 complex reported respective *Rg* values of 22, 44 and 44 Å (Ashish *et al.*, 2006). Here however, due to the low signal arising from *Sample2dCD4,* the Guinier *Rg* (19 Å) and and P(r) *Rg* value (14 Å) of 2dCD4 were severe underestimates of the previous values.

The determination of $I_o$ enabled the calculation of the molecular weight (MW) of each of the components by making use of equation 2.4 (see Table 6.1). However, all of the measurements gave underestimated MWs compared to the expected theoretical molecular weights. In the case of *Sample2dCD4*, with the largest difference in MW estimate, it is most likely due to the poor signal which makes analysis of the neutron scattering data highly error prone. In the case of *SampleComplex* and *SampleGP120* measurements, one source of discrepancy could arise from the glycans present on GP120, which was previously determined to constitute 39.6 % of the GP120 MW (Channell, 2018). Glycans have SLDs that are slightly different from that of proteins, and therefore in principle require different $D_2O/H_2O$ solvent content in order to be matched-out exactly; however, in practice the two values are too close to be able to distinguish the two components reliably. Therefore, despite the use of the glycoprotein at an averaged match point, the scattering contributions of glycans, and the imperfect matched out protein portion, lead to deviation in the recorded $I_o$. Previously published work using SAXS to study the GP120-dimer has also reported such discrepancies in the $I_o$- derived MW compared to the expected theoretical value (Guttman *et al.*, 2012).

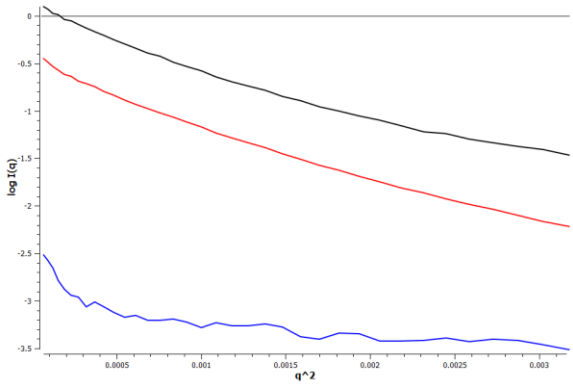



***Figure 6.3:*** *Buffer subtracted Guinier plots (Log q vs $q^2$) for the 2dCD4-GP120 complex in 0 % (black), 45 % (blue) and* 92 *% (red) $D_2O$ buffers.*

### *Analysis of P(r) and Kratky plots for the CD4-GP120 complex*

The normalised P(r) functions for CD4-GP120 complex, 2dCD4 and Gp120 with the parameters determined by autoGNOM software, are plotted in Figure 6.4A, followed by their corresponding Kratky plots (Figure 6.4B). The resulting $D_{max}$ values extracted from the Kratky plots are summarised in Table 6.1.

As expected, the P(r) plot of of *Sample2dCD4* was of poor quality (quality estimate = 0.52). Therefore, interpretation of the information from the plot is problematic. Nevertheless, compared to the other plots, a low $D_{max}$ (74 Å), reflecting measurement of only 2dCD4 domains, was obtained.

For the complex, multiple peaks in the P(r) plot were observed (quality estimate = 0.61), suggesting extended and multiple domains, with high flexibility in the structure. Conversely, with a much more uniform bell-shaped plot for GP120 (quality estimate = 0.82), a structure with less flexibility and more globularity with less distinction between domains was indicated. It should however be noted that $D_2O$ in the solvent contributes to rigidity of proteins and therefore to loss of flexibility (Cioni and Strambini, 2002). Furthermore, due to the visibility of 2dCD4 in the measurements, the $D_{max}$ of 2dCD4-GP120 complex (196 Å) was higher than that of GP120 (147 Å).

The Kratky plot of the 2dCD4-GP120 complex was found to closely match that of GP120 for low q**Rg* values. Having higher flexibility and plausibly extended conformation due the visibility of 2dCD4, the Kratky plot of the 2dCD4-GP120 complex showed a continuous rise after the peak. In both cases, the peak of the bell curve was located close to q**Rg* = √3 (1.104), which indicates the overall globularity of the structure. The Krakty plot of 2dCD4 however did not produce a typical bell-shaped profile, but showed a continuous exponential-shaped like profile after the peak, suggestive of an intrinsically disordered protein (Receveur-Brechot and Durand, 2012). However, such a profile could arise if one considers the dumbell-shape structure, which is highly flexible about its axis (Figure 6.1, Figure 2.7).

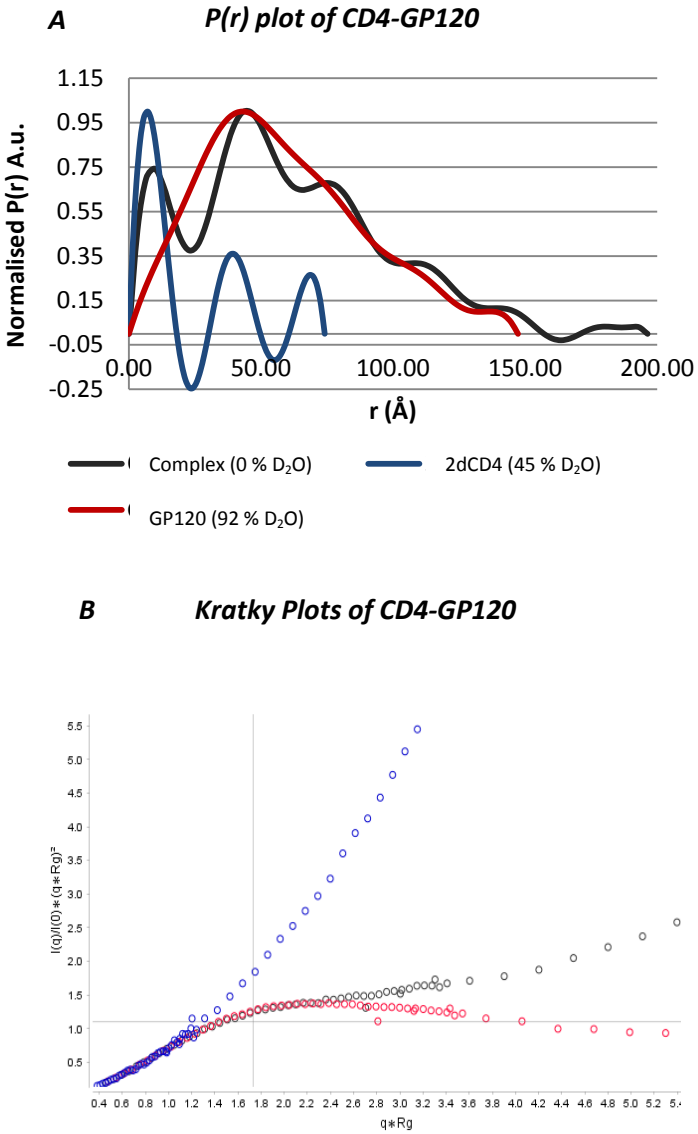



***Figure 6.4:*** *Normalised P(r) plots (A) and normalised Guinier-based Kratky plots (B) of 2dCD4-GP120 complex (black), 2dCD4 (blue) and GP120 (red) in respectively 0, 45 and 92 % $D_2O$ buffers. Due to the low quality of SANS data, the P(r) and Kratky plots of 2dCD4 in particular should be considered with caution.*

### 6.3.2. Preliminary model of GP120 dimer

Since the data quality recorded for *2dCD4* was of limited quality, the data recorded for *GP120* alone was used as the basis for inferring a model of the complex. A model of glycosylated GP120 dimer was generated in SASREF using chimeric monomer models of *GlyGP-b*. The curve of the GP120 dimer model was found to follow the initial section of the experimental data (<0.15 $Å^{-1}$), but deviated at higher Q (Figure 6.5A). Subsequently, 2dCD4 molecules were added to the GP120 dimer (Figure 6.5B). It should, however, be stressed that the models are highly speculative.

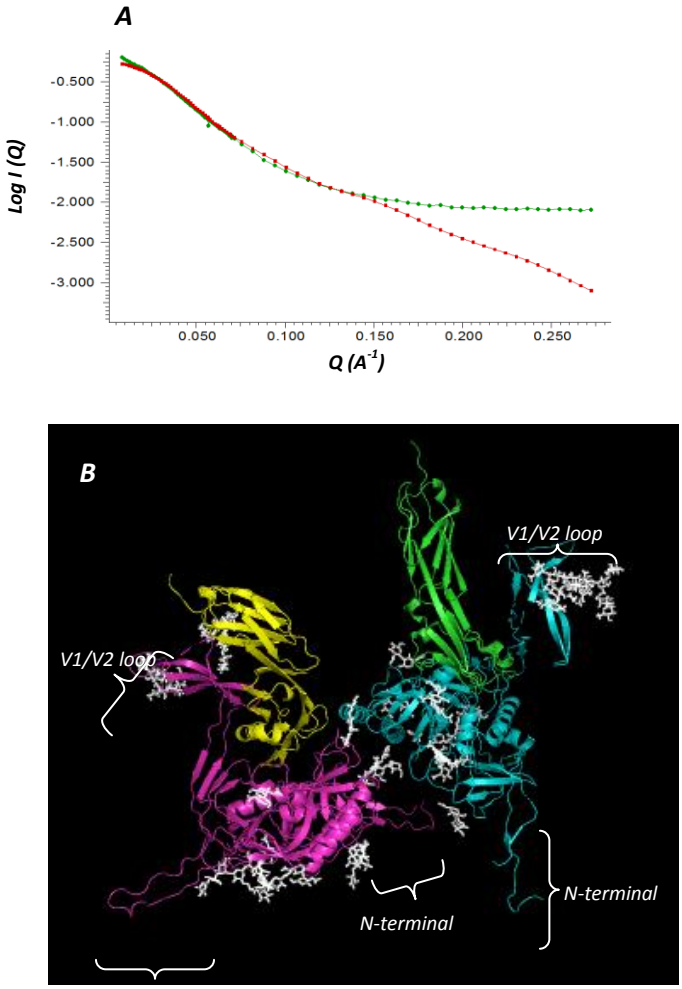



***Figure 6.4:*** *Suggested model of CD4-GP120. (A) Curve of SASREF model generated (green) was fitted against experimental data (red). (B) SASREF model of GP120-dimer, with the inclusion of 2dCD4. GP120 monomers are shown in cyan and pink, 2dCD4 are shown in yellow and green, and glycans moieties are shown in white.*

## 6.4. DISCUSSION

### 6.4.1. Confirmation of matching-out conditions for the components of the CD4-gp120 complex

In chapter three, the production of deuterated CFPS 2dCD4 was described. The 2dCD4 was then bound to dimeric GP120, and the complex was subsequently used for SANS studies. Previous work carried out by our collaborators had already established the deuteration level and match point of CFPS 2dCD4. Here, the principle of protein match-out by contrast variation in order to gain insight into the individual components of the complex was applied (Dunne *et al.*, 2017b). In previous studies of the CD4-GP120 complex where SAXS was used, the SAXS data obtained could only consider the entire complex (Guttman *et al.*, 2012, Ashish *et al.*, 2006). The experiments performed here confirmed the ability to effectively match out either the GP120 dimer at 45 % $D_2O$ or deuterated CFPS 2dCD4 at 92 % $D_2O$ (Figure 6.2). SANS is therefore an advantageous alternative that allows study of the individual components of the complex. This feature permits more accurate determination of the size, shape and mostly conformational changes in the sub-units of the complex. In this study, consistent information about the size parameters of the GP120 dimer, 2dCD4 molecules and the entire complex was extracted from 1D neutron scatter data.

### 6.4.2. Insights into 2dCD4-GP120 complex

Previous solution scattering work has mostly focused on monomeric GP120 in complex with 2dCD4 (Guttman *et al.*, 2012, Ashish *et al.*, 2006). In chapter three, as determined by SEC analyses, it was found that GP120 was being secreted as monomers and dimers. Dimeric GP120 was expressed almost twice as much as the monomer. Dimeric GP120 was selected for producing the complex with deuterated CFPS 2dCD4. The GP120 dimer is however described to be an aberrantly disulphide-linked form that is produced by over-expressing mammalian cells (Finzi *et al.*, 2010). The work here therefore provides insights into the GP120 dimer – 2dCD4 complex.

The majority of structural studies of GP120 or the CD4-GP120 complex has made use of a truncated variant of GP120 core which lacks the highly flexible loops and terminal regions (Kwong *et al.*, 1998, Chen *et al.*, 2005, Diskin *et al.*, 2010). This is due to the fact that these highly flexible regions of GP120

usually impede crystallisation. As a result of the missing loops of GP120, the comparison of core GP120 and its CD4-bound state has typically resulted in no significant conformational differences being observed (Diskin *et al.*, 2010). However, thermodynamic studies on the binding of native GP120 to CD4 have shown that major differences arise due to significant structural re-arrangement in the GP120 (Myszka *et al.*, 2000). Furthermore, with the variable loops in GP120 playing critical roles in the functionality of HIV, omitting these from studies limits the understanding of the virus.

With the use of full-length GP120, insights into the possible positioning of variable loops and terminal sections of GP120 are possible (Figure 6.5). As presented in a previous publication in which CD4-GP120 was analysed using SAXS, the V1/V2 loops were suggested to be located in the distal part of the unbound GP120 molecule, and to occupy a region similar to that of antibody B12 (Guttman *et al.*, 2012). The B12 antibody is a well-known anti-GP120 antibody that targets epitopes which becomes exposed only after CD4 binding (Zwick *et al.*, 2003). It is therefore suggested that this epitope is restricted by the presence of the V1/V2 loops prior to CD4 binding. When CD4 is bound to GP120, the V1/V2 loops undergo significant re-orientation and localises themselves close to the bound CD4. The interaction between the V1/V2 loops with CD4 stabilises the latter in its complex with GP120 since it was reported that GP120 constructs that lack the V1/V2 loops had weakened binding to CD4 (Rits-Volloch *et al.*, 2006). Consequently, the CXCR4/CCR4 co-receptor binding site located on the distal portion of GP120, which includes the V3 loop, is exposed for subsequent GP120 binding to the co-receptor (Tamamis and Floudas, 2014, Zhang *et al.*, 2001). These results are consistent with previous studies in which antibodies targeting the V1/V2 loops concealed portions of GP120, could preferably bind to GP120 in the CD4-bound state, or mutant GP120 with deleted V1/V2 loops (Wyatt *et al.*, 1995, Saunders *et al.*, 2005).

Additionally, the SANS study has demonstrated the potential of studying GP120 in its native state, with the inclusion of glycans. In particular, in addition to its intrinsic flexibility, the flexibility of the V1/V2 loops is also due the glycans found on it (McLellan *et al.*, 2011). The V1/V2 loops, and in particular the glycans that cover it, play an important role in immune evasion by shielding the inner epitopes of GP120 from neutralising antibodies (O'Connell *et al.*, 2014). Furthermore, over the past decade, there has been increasing interest in the glycosylated V1/V2 loops since they are a very sensitive target to recently identified potent broadly neutralising antibodies (Walker *et al.*, 2009, Bonsignori *et al.*, 2011, Walker *et al.*, 2011, McLellan *et al.*, 2011). The model of GP120 dimer bound to 2dCD4 presented here can therefore assist to further the understanding in intricacies of the loops of glycosylated form of GP120.

### 6.4.3. Perspectives for future work with the CD4-GP120 complex

A main limitation in the work presented here was the poor signal arising from CFPS-produced 2dCD4 in the complex at the GP120 matchpoint. The signal arising from *SampleGP120*, where only the GP120-dimer was visible, was therefore was used as the basis for inferring the structure of the entire complex. This approach may, however, introduce bias in the suggesting the binding locations of 2dCD4. Therefore, a focus on improving this signal is critical before proceeding into more advanced studies involving the CD4-GP120 complex. One could firstly increase the concentration of the sample even further in order to improve the signal. However, an alternate option is to improve the homogeneity of the sample by using monomeric GP120, instead of dimeric GP120. Since SEC cannot effectively separate these components (see Figure 3.36), *SampleGP120* as used here is likely to consist of unbound GP120, GP120 bound to single 2dCD4 molecule, or GP120 bound to two 2dCD4 molecules. With monomeric GP120 in complex with 2dCD4, which has been used in previous studies (Guttman *et al.*, 2012, Ashish *et al.*, 2006), it is more feasible to obtain a highly homogenous solution consisting of monomeric GP120 bound to a single 2dCD4 molecule. Here, due to the tight schedule with the up-coming SANS experiment, and with dimeric GP120 being available in largest quantity compared to monomeric form, dimeric was selected (chapter three, 3.3.4). The use of new mammalian expression systems which produce monomeric GP120 favourably is therefore to be considered.

In this context, future SANS studies of the complex could consider the conformational changes incurred in both GP120 and CD4 upon complex formation. CD4 has been reported to undergo conformational changes when bound to GP120 due to the reduction of the disulphide bonds in the second domain of CD4 (Cerutti *et al.*, 2014b). It is of crucial importance to define the changes in the variable loops of GP120 with more precision and to identify the key locations of the glycans. Here, selective labelling of glycans on GP120 is an interesting option. This could potentially be accomplished by using genetically engineered mammalian expression systems which express highly specific glycosyltransferases, that in turn, enable the incorporation of labelled glycans on GP120 (Almaraz and Li, 2017). Ultimately, these experiments would involve the use of CFPS CXCR4, in either proteoliposome or detergent format, where the goal would be to analyse the conformational changes that lead to the formation of the overall ensemble of the CXCR4-GP120-2dCD4 complex.

## 6.5. CONCLUSIONS

CFPS has been used for the first time to produce deuterated 2dCD4. This was complexed to GP120, and subsequently analysed by SANS. The interest of the study was the ability to use the fully native GP120 molecule - including the variable loops and glycans which are often omitted in a majority of previous studies - in the experiment and modelling trials of the CD4-GP120 complex. Despite the fact that the data were not of sufficient quality to permit definitive modelling, the study has nevertheless provided crucially important information which can drive future work in this area. These include, increasing the concentration of the sample even further, and improving the homogeneity of the sample with the use of monomer GP120 instead of dimer GP120. Additionally, selective labelling of glycans could be considered in order to enhance SANS signals, and consequently improve models. The study of CD4-GP120 by SANS could provide important insights into conformational changes that glycosylated GP120 undergoes upon CD4 binding. Whilst these are challenging prospects, the application of CFPS for the production of deuterated protein for the purpose of SANS, is promising route, particularly in the context of SANS growing as a technique used in integrated structural biology. With Synthelis seeking to expand its line of services in order to include structural characterisation, the work presented here can open the door to new markets in the field of SANS. Moreover, the structural characterisation of CFPS proteins by SANS, will provide further high-end assurance in the quality of the CFPS protein production approach.

# CHAPTER SEVEN

## FEASIBILITY STUDY OF DIRECT SANS STRUCTURAL ANALYSIS OF A MODEL PROTEIN WITHIN THE UNFRACTIONATED CFPS REACTION MIXTURES

# ABSTRACT

The structural characterisation of CFPS membrane proteins in proteoliposomes is typically hindered by low expression and purity issues (see chapter three). Due to the complexity and heterogeneity of the CFPS lysate, the structural characterisation of proteins produced within the CFPS lysate can be very challenging with most biophysical methods. Hence alternative methods have been investigated that could allow the structural characterisation of membrane proteins in proteoliposomes, despite the presence of contaminating lysate proteins.

As a first step towards this goal, a feasibility study of exploiting SANS, in conjunction with matchout deuteration approaches and contrast variation, has been carried out. The aim of this study was to identify the conditions in which model proteins immersed in CFPS lysate could be detected and imaged by SANS. Such a concept, if feasible, could open the possibility of modelling protein structures within a CFPS lysate and without purification. Furthermore, such an approach could be used more widely for the study of the effect of molecular crowding on protein structure and dynamics (Grimaldo *et al.*, 2019) in conditions that are highly relevant to the *in vivo* situation.

Two scenarios were considered – one in which a deuterated protein was produced using hydrogenated machinery (hydrogenated lysate), and one where a hydrogenated protein was produced using deuterated machinery (deuterated lysate). The key issue in both cases is the strength of the exploitable coherent SANS signal arising from the target protein in relation to the total hydrogen incoherent scattering and the total coherent signal from the machinery. The feasibility work described here was carried out to establish the limits in terms of the concentrations required that would permit such an approach. This was achieved by recording SANS data from a series of samples consisting of variously labelled lysates (hydrogenated lysate, matchout deuterated and perdeuterated CFPS lysate) to which had been added a range of different model proteins at various concentrations. It was found that SANS signals of the model proteins at the studied here, remained interpretable despite being immersed in CFPS lysate. The data quality was assessed and key SANS structural parameters (*Rg*, Dmax, MW) were extracted. *Ab-initio* models were also constructed. These were found to be consistent with reference sample control data recorded from the same target proteins in an aqueous environment, although the quality of the modelling was inevitability poorer. In conclusion, it has proved possible to obtain tractable structural information from model proteins in the presence of amounts of lysate that are equivalent to those used during CFPS reactions.

These results set out the requirements for future *in situ* CFPS SANS studies with real unpurified CFPS-generated proteins, the study of native proteins in their crowded environment, and ultimately, the study of CFPS membrane proteins in proteoliposomes.

## 7.1. INTRODUCTION

Demonstrating the structural integrity of CFPS-produced membrane proteins as equivalent to those produced using other recombinant systems was one of the core objectives of this PhD thesis work. For the purposes of many aspects of structural biology research, large amounts of highly purified and homogenous protein are typically required. As described in chapter three, despite extensive optimisation tests to increase the yields in CFPS reactions, to reduce the adherence of contaminating lysate proteins on the proteoliposomes, and to improve the purity, it was difficult to produce sufficient pure protein for advanced structural studies.

This chapter concerns the first steps that have been taken in testing the feasibility of using SANS, in conjunction with selective deuteration methods, to allow structural characterisation of proteins in their unfractionated state – *i.e.* unpurified from the CFPS reaction mixture. A major advantage of such a method is that it avoids challenging purification steps, which may result in protein loss, folding problems, or unwanted conformational changes/polydispersity. Such a technique may also have value for the characterisation of other proteins that are difficult to purify or that need a particular environment to sustain their structural/functional integrity.

### 7.1.1. Experimental rationale for structural analysis of protein within unfractionated CFPS reaction mixtures

The principle of matchout labelling (Haertlein *et al.*, 2016, Laux *et al.*, 2008, Dunne *et al.*, 2017b, Waldie *et al.*, 2019) was extended for this work. Instead of matching out a single protein, the idea was to construct a deuterium-labelled sample system in which suitable contrast (for SANS studies) could be generated between the target protein of interest and the "background" CFPS lystate. This concept is of course challenging due to the high heterogeneity, polydispersity and concentration of the lysate, which

includes proteins, RNA, DNA and other biomolecules. The aim of the experiments was to identify the conditions in which the SANS signal from a target protein immersed in such a cellular lysate could be detected and analysed to obtain structural data in a context close to that of CFPS, or even cellular conditions.

The logical starting point for the study was to produce samples containing hydrogenated lysate to which deuterated target proteins had been added. It was suggested at the outset that this approach would involve very high levels of hydrogen incoherent scattering from the lysate and that this could obscure the signal from the deuterated target. A second possibility involved the "mirror" sample configuration in which the lysate was deuterated and the target was hydrogenated; in this situation one could expect much lower levels of hydrogen incoherent scattering against which the coherent SANS signal of the target protein would have to be measured.

To this end, the following lysates were produced (following the procedures described in chapter two):

(i) Hydrogenated lysate (HL): *i.e.* normal CFPS lysate produced in unlabelled conditions

(ii) Matchout deuterated lysate (ML): CFPS lysate produced using 85 % $D_2O$-content media, with the resulting lysate deuterated so that it was, on average, as close as possible to the scattering length density of 100 % $D_2O$ (see Section 2.1.4, Figure 2.3)

(iii) Perdeuterated lysate (PL): fully deuterated CFPS lysate produced using 100 % $D_2O$-content media and deuterated carbon source. The aim of this sample was to serve as a reference to for comparison with the ML lysate (ii) above.

In the first sample arrangement, deuterated model proteins were to be added to unlabelled hydrogenated lysate (HL). Conversely, in the second setup, an unlabelled hydrogenated target protein would be added to deuterated lysates (ML or PL). In either case, the goal was to seek and quantify conditions in which the lysate was as close as possible to "invisible", leaving a visible signal from the protein of interest (Figure 7.1).

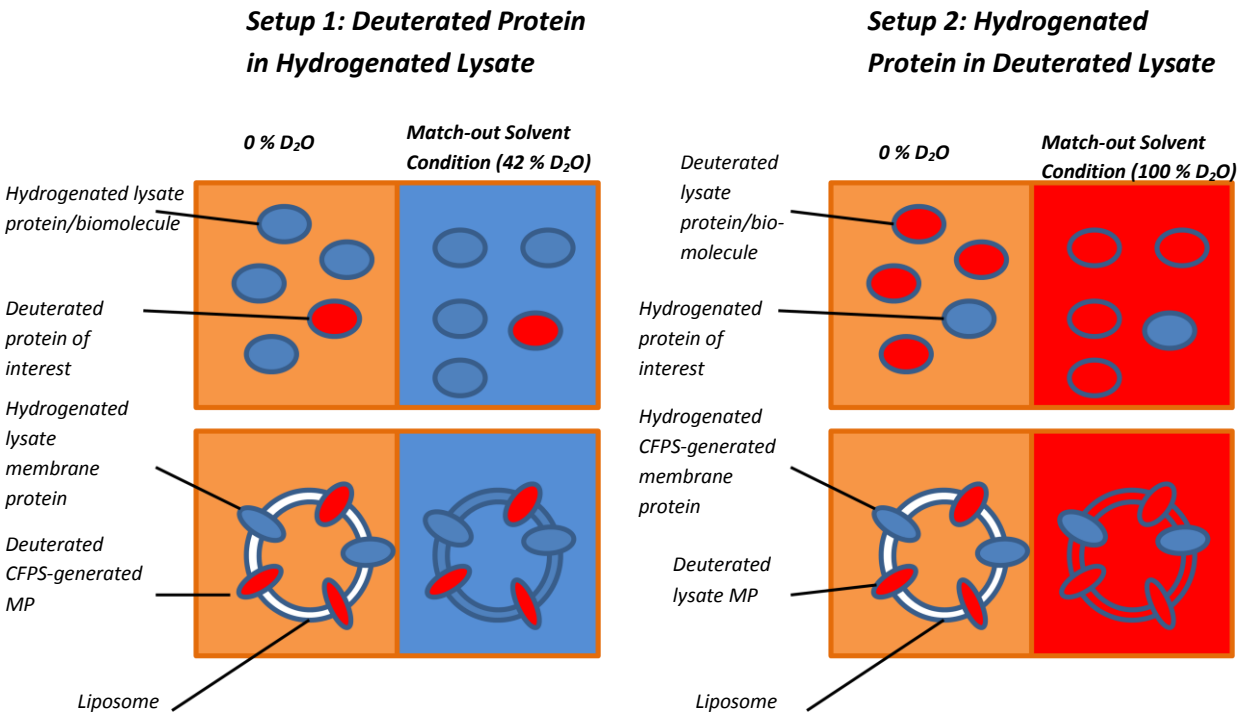



***Figure 7.1:*** *Conceptual diagram for the experimental setups for the SANS analysis of proteins in the presence of lysates. In setup 1, hydrogenated lysate is rendered "invisible" when placed in matchout conditions, while the deuterated protein of interest is "visible". Conversely, in setup 2, deuterated lysate is rendered "invisible" in matchout conditions, while the hydrogenated protein of interest remains "visible". The feasibility of this experiment may be of interest in extending the technique to CFPS proteoliposomes.*

### 7.1.2. Model proteins considered for structural analysis of protein unfractionated within CFPS reaction mixtures

This study relied on the availability of a deuterated and hydrogenated proteins to be used as model target proteins in the SANS experiments. Three model systems were selected as summarised below.

**1. Maltose binding protein (MBP), in either hydrogenated (HMBP) or perdeuterated (DMBP) forms**

MBP is a well-studied protein that is involved in the active transport of maltose across the cytoplasmic membrane and energy metabolism in *E. coli* (Spurlino *et al.*, 1991, Oldham and Chen, 2011, Oldham *et al.*, 2007, Boos and Shuman, 1998). Furthermore, DMBP has been generally used as a model protein, for which the production and characterisation by SANS has been well established (Dunne *et al.*, 2017, Haertlein *et al.*, 2016, Laux *et al.*, 2008). From these previous studies, HMBP and DMBP were established to be proteins that are highly globular, soluble, homogenous, stable and aggregate-free (even at the high concentrations), making it a good candidate as a model system.

**2. Bovine serum albumin (BSA)**

Serum albumin is a one of the most abundant proteins present in mammalian plasma (35 – 55 g/L) (Schreiber and Urban, 1978). Its main functions include the regulation of colloid osmotic blood pressure and serving as a major transport protein for a number of metabolites, hormones and drugs (Rothman and Orci, 1992, Majorek *et al.*, 2012). Bovine serum albumin (BSA) in particular has found extensive use in the field of biochemistry as being a very readily commercially available, stable, globular model protein. As a model protein, BSA has been used for several SAS studies for understanding the general properties of proteins and for establishing novel methods and protocols in the development of SAS (Mertens and Svergun, 2017, Trewhella *et al.*, 2017, Sinibaldi *et al.*, 2008, Zhang *et al.*, 2007). Bovine serum albumen (BSA) exists as monomeric and dimeric forms and as a model system represents an increase in complexity compared to MBP.

### 3. Bovine γ-Globulin (Glob)

Immunoglobulins, responsible for the neutralisation of foreign pathogens, are critical agents in the immune system and are used extensively in biotechnology and medicine (Waldmann, 2003). Particularly, immunoglobulins G (IgG) are the most well-known of the family. IgGs consist of four peptide chains (two identical heavy chains and two identical light chains) linked by disulphide bonds, resulting in the typical three-lobed overall Y-shaped structure (Figure 7.2) (Harris *et al.*, 1997, Saphire *et al.*, 2001). IgGs are therefore non-spherical proteins, with considerable structural flexibility. IgGs have been shown to exist as a monomer-dimer equilibrium (Plath *et al.*, 2016, Moore *et al.*, 1999). Commercially available bovine γ-globulin (Glob), used as the third model system here, contains an antibody mixture extracted from pooled bovine plasma, consisting of IgG (80 %), IgM (10 %) and IgA (<10 %) (Figure 7.2). The use of Glob, highly flexible, polydisperse and shifting in oligomeric state, therefore adds another level of complexity to the sample. This sample tends to represent more accurately the state of *de novo* proteins produced in an actual CFPS reaction. Furthermore, Glob has already been analysed by SAXS under protein crowding conditions (Da Vela *et al.*, 2017) providing further supporting background data for these feasibility studies.

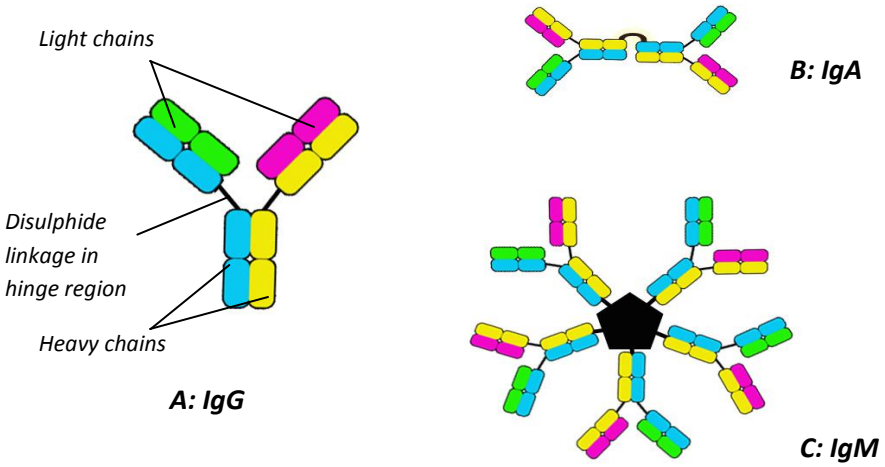



***Figure 7.2:*** *The structural arrangement of immunoglobulin (antibodies) IgG (A), IgA (B) and IgM (C) (not to scale). IgG, as a monomer, consists of two light and two heavy chains with disulphide linkages in the hinge region and adopts a Y-shaped conformation. IgA consists of two monomers of IgG linked covalently to each other at the base of the heavy chains. IgM, as a pentamer consisting of five IgG molecules, are linked together by central linker atoms.*

## 7.2. MATERIAL AND METHODS

### 7.2.1. Preparation of samples

Hydrogenated minimal media cell-free lysate (HL), perdeuterated cell-free lysate (PL) and matchout cell-free lysate (ML), as described in chapter two, were used. The lysates were centrifuged for 30 minutes at 15000 g (4 $^{o}$C). Supernatants were collected and mixed by pipetting, and stored on ice before further use. Purified hydrogenated (HMBP) and perdeuterated maltose binding protein (DMBP) protein produced in *E.coli* as described in a previous publication, were kindly provided by Dr. Martine Moulin from the ILL Life Sciences Group (Dunne *et al.*, 2017b). Bovine γ-globulin (Glob) (Braun *et al.*, 2017, Da Vela *et al.*, 2017) was purchased from Sigma-Aldrich (G5009). This sample consisted of IgG (80 %), IgM (10 %) and IgA (<10 %). Bovine serum albumin lyophilised powder (≥96 % purity, BSA) was purchased from Sigma-Aldrich (A2153).

50 mM HEPES buffers, in $H_2O$ (pH 8.2) and in 100 % $D_2O$ (pD 7.9) were prepared. 50 mM HEPES buffer at 42 % $D_2O$ was made by mixing the two afore-mentioned buffers (42 % $D_2O$/58 % $H_2O$ v/v). These buffers, to be used as solvents and dialysis buffers, were cooled to 4 $^{o}$C prior to use. Stock solutions of γ-globulin and BSA were solubilised in the 100 % $D_2O$ 50 mM HEPES buffer at 20 mg/mL. Stock solutions of HMBP and DMBP were diluted to respectively 5 mg/mL and 4 mg/mL each, using respectively the 100 % $D_2O$- and 42 % $D_2O$-based 50 mM HEPES buffers.

For SANS analysis, a range of samples were prepared with differing model protein and lysate (HL, ML and PL) concentrations. The set of samples is shown schematically in Figure 7.3. 300 µL of each sample and their corresponding lysate controls, which excluded model proteins, were made. Tables 7.1, 7.2 and 7.3 provide the details of each sample and their experimental objective. In order to attain solvent matchout conditions (Figure 7.1), the samples were dialysed extensively against at least 10 X the volume of the corresponding 42 % and 100 % $D_2O$ 50 mM HEPES buffers using a GEBAflex midi with 6 – 8 kDa MWCO. The recovered samples were centrifuged for 30 minutes at 15000 g (4 $^{o}$C), supernatants collected, mixed adequately by pipetting and stored on ice before SANS analysis.

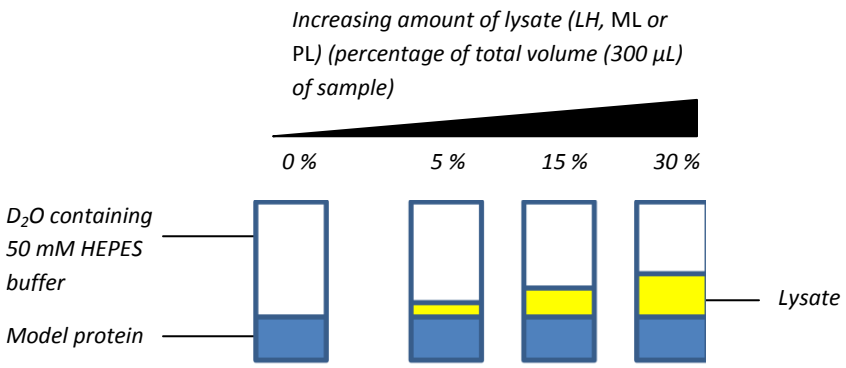



***Figure 7.3:*** *Schematic of composition of the samples prepared for SANS showing increasing lysate concentration to test the limit at where a SANS signal from the target model protein could still be recorded. In some cases, the concentration of the model protein was also varied across the lysate concentration series. The corresponding negative controls excluded model proteins.*

| Sample name | MBP (mg/mL) | Lysate amount used in sample (%) | Type of lysate | $D_2O$ in dialysis buffer (%) | Sample objective |
|---|---|---|---|---|---|
| **HMBP-100D** | 2.5 | 0 | N/A | 100 | Reference sample of HMBP |
| **HMBP-5ML** | 2.5 | 5 | Matchout – ML | 100 | Series increasing concentration matchout lysate with constant HMBP concentration, to establish the limits of the HMBP signal detection over noise. |
| **HMBP-15ML** | 2.5 | 15 | Matchout – ML | 100 | |
| **HMBP-30ML** | 2.5 | 30 | Matchout – ML | 100 | |
| **HMBP-5PL** | 2.5 | 5 | Perdeuterated – PL | 100 | Series increasing concentration perdeuterated lysate with constant HMBP concentration, to establish the limits of the HMBP signal detection over noise. |
| **HMBP-15PL** | 2.5 | 15 | Perdeuterated – PL | 100 | |
| **HMBP-30PL** | 2.5 | 30 | Perdeuterated – PL | 100 | |
| **DMBP-42D** | 1.8 | 0 | N/A | 42 | Reference sample of DMBP |
| **DMBP-5HL** | 1.8 | 5 | Hydrogenated – HL | 42 | Series increasing hydrogenated lysate concentration with constant DMBP concentration to verify limits of the DMBP signal detection over noise |
| **DMBP-15HL** | 1.8 | 15 | Hydrogenated – HL | 42 | |
| **DMBP-22.5HL** | 1.8 | 22.5 | Hydrogenated – HL | 42 | |
| **DMBP-30HL** | 1.8 | 30 | Hydrogenated – HL | 42 | |

***Table 7.1:*** *List of samples prepared for MBP (HMBP is hydrogenated, DMBP is deuterated protein).*

| Sample name | BSA (mg/mL) | Lysate amount used in sample (%) | Hydrogenated or deuterated lysate | $D_2O$ in dialysis buffer (%) | Sample objective |
|---|---|---|---|---|---|
| **BSA-100D** | 10 | 0 | N/A | 100 | Reference sample of BSA |
| **BSA-5ML** | 10 | 5 | Matchout – ML | 100 | Series increasing matchout lysate concentration with constant BSA concentration to verify limits of the BSA signal detection over noise |
| **BSA-15ML** | 10 | 15 | Matchout – ML | 100 | |
| **BSA-30ML** | 10 | 30 | Matchout – ML | 100 | |
| **BSA-5PL** | 10 | 5 | Perdeuterated – PL | 100 | Series increasing perdeuterated lysate concentration with constant BSA concentration to verify limits of the BSA signal detection over noise |
| **BSA-15PL** | 10 | 15 | Perdeuterated – PL | 100 | |
| **BSA-30PL** | 10 | 30 | Perdeuterated – PL | 100 | |
| **5BSA-30ML** | 5 | 30 | Matchout – ML | 100 | Series decreasing BSA concentration with constant ML concentration to verify limits of the BSA signal detection over noise |
| **2.5BSA-30ML** | 2.5 | 30 | Matchout – ML | 100 | |
| **1.25BSA-30ML** | 1.25 | 30 | Matchout – ML | 100 | |

***Table 7.2:*** *List of samples prepared for BSA.*

| Sample name | MBP (mg/mL) | Lysate amount used in sample (%) | Hydrogenated or deuterated lysate | $D_2O$ in dialysis buffer (%) | Sample objective |
|---|---|---|---|---|---|
| **Glob-100D** | 10 | 0 | N/A | 100 | Reference sample of Glob |
| **Glob-5ML** | 10 | 5 | Matchout – ML | 100 | Series increasing matchout lysate concentration with constant Glob concentration to verify limits of the Glob signal detection over noise |
| **Glob-15ML** | 10 | 15 | Matchout – ML | 100 | |
| **Glob-22.5ML** | 10 | 22.5 | Matchout – ML | 100 | |
| **Glob-30ML** | 10 | 30 | Matchout – ML | 100 | |
| **Glob-5PL** | 10 | 5 | Perdeuterated – PL | 100 | Series increasing perdeuterated lysate concentration with constant Glob concentration to verify limits of the Glob signal detection over noise |
| **Glob-15L1PL** | 10 | 15 | Perdeuterated – PL | 100 | |
| **Glob-22.5PL** | 10 | 22.5 | Perdeuterated – PL | 100 | |
| **Glob-30PL** | 10 | 30 | Perdeuterated – PL | 100 | |
| **5Glob-30ML** | 5 | 30 | Matchout – ML | 100 | Series decreasing Glob concentration with constant matchout lysate concentration to verify limits of the Glob signal interpretability over noise |
| **2.5Glob-30ML** | 2.5 | 30 | Matchout – ML | 100 | |
| **1.25Glob-30ML** | 1.25 | 30 | Matchout – ML | 100 | |
| **5Glob-30PL** | 5 | 30 | Perdeuterated – PL | 100 | Series decreasing Glob concentration with constant perdeuterated lysate concentration to verify limits of the Glob signal detection over noise |
| **2.5Glob-30PL** | 2.5 | 30 | Perdeuterated – PL | 100 | |
| **1.25Glob-30PL** | 1.25 | 30 | Perdeuterated – PL | 100 | |
| **Glob-22.5ML + tRNA + DNA** | 10 | 22.5 | Matchout – ML | 100 | Samples with the inclusion of tRNA (0.5 mg/mL) and plasmid DNA (15 ng/mL) to simulate the presence these components in an actual CFPS reaction, and verify the detection of Glob signal |
| **Glob-22.5PL + tRNA + DNA** | 10 | 22.5 | Perdeuterated – PL | 100 | |

***Table 7.3:*** *List of samples prepared for bovine γ-globulin (Glob).*

### 7.2.2. SANS on D22 (ILL)

Data were collected on the high-flux, large dynamic range small-angle neutron diffractometer D22 located at the ILL, Grenoble, France (www.ill.eu/users/instruments/instruments-list/d22/more/d22-manual/). Elastically scattered neutrons are detected according to the wave vector transfer q, defined as $q = 4\pi\sin\theta/\lambda$ where $2\theta$ is the scattering angle and $\lambda$ the neutron wavelength. The neutron wavelength was set to 6 Å. The samples were measured in 1-mm-path-length Suprasil quartz cuvettes (Hellma). Data were recorded at collimation lengths of 11.2 m, 5.6 m and 2.8 m and respective sample-to-detector distances of 11.2 m, 5.6 m and 2.0 m to provide a full q range (0.001 – 0.5 Å) from the Guinier region of the protein scattering to the solvent/lysate level of incoherent scattering. The raw data were reduced using the GRASP4 software package (www.ill.eu/users/support-labs-infrastructure/software-scientific-tools/grasp/), which included thickness and transmission scaling, empty cell and blocked beam subtractions, calibration to absolute intensity using incident flux measured at sample position, azimuthal averaging and then data merging to produce 1D neutron intensity plots.

### 7.2.3. Treatment of scattering data and model building

Data analysis and model building against the SANS data were carried out using programmes as described in Section 6.2.3. The crystal models used for D/HMBP, Glob and BSA were respectively PDB codes: 1JW4, 1IGT and 4F5S.

## 7.3. RESULTS

The aim of the SANS experiments carried out here was to verify the detection and the interpretability of the signal of the model protein, in a range of concentrations within CFPS lysate. The results described below are given in sections that firstly describe the overall quality of the analysis and comparison of SANS data; secondly the interpretation and verification in the consistency of protein size parameters; finally, the modelling of the protein in the presence of CFPS lysate.

### 7.3.1. Overall analysis, quality analysis and comparison of SANS data

This section deals with the overall visual inspection and assessment of the quality of the SANS data collected, as interpreted from intensity (I(q) vs q) plots. A particular focus was the verification of the protein signal as compared to its corresponding pure (no lysate) control. In addition, this section is concerned with visually assessing the Guinier region (low q) in log q vs $q^2$ plots derived from the SANS data. In the absence of aggregation or significant inter-particular interaction, log q vs $q^2$ plots give a straight line without any curvature as $q \rightarrow 0$. Good data from a well-defined Guinier region allows accurate calculation of *Rg* values (see next section).

#### *Comparing SANS signals of deuterated and unlabelled MBP without lysate present*

The log I(q) vs log q plot of pure HMBP (2 mg/mL) in 100 % $D_2O$ buffer (HMBP-100D), and deuterated MBP (2 mg/mL) in 42 % $D_2O$ buffer (DMBP-42D) were firstly verified (Figure 7.4A). The SANS signal derived from MBP could be detected against the buffer background in both cases. However, the signal difference against the solvent background, ΔI (q), of HMBP-100 was higher than that of DMBP-42 for the majority of the q-range (Figure 7.4B). Furthermore, the signal obtained from HMBP-100 was interpretable throughout the measured q range whereas that of DMBP-42 resulted in loss of interpretability due high noise at high q values.

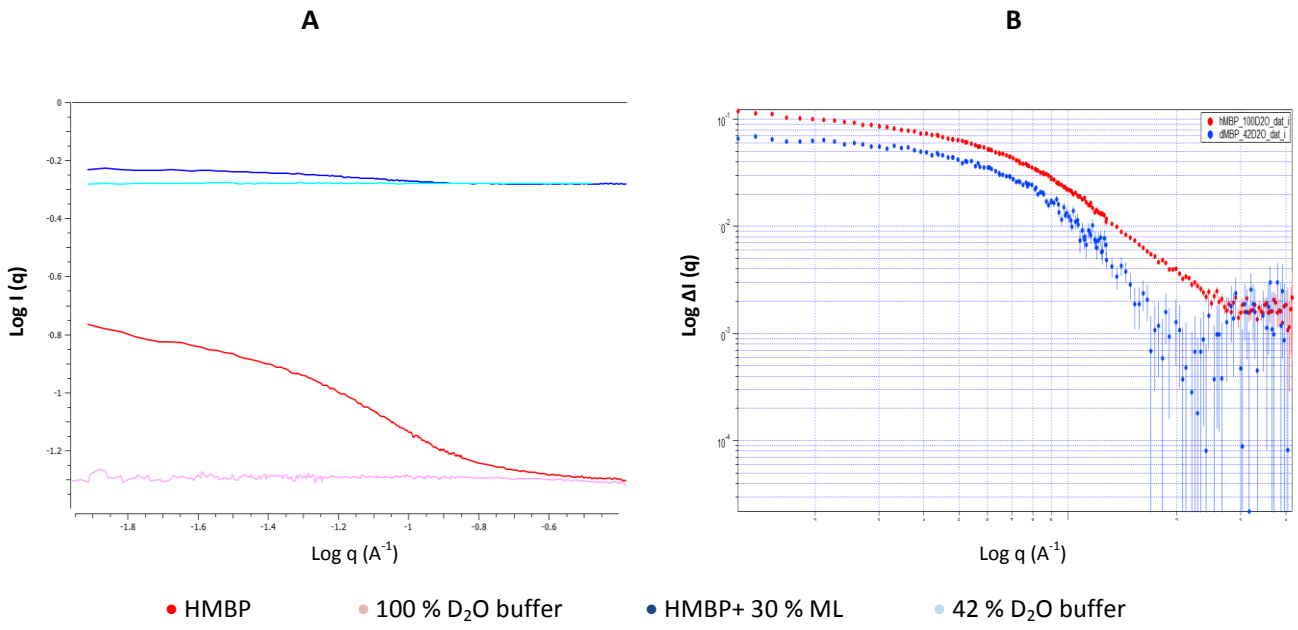



*Figure 7.4: SANS Intensity plots recorded from the model protein MBP with no lysate present.* ***A:*** *Log I(q) vs. log q, of DMBP in 42 %* $D_2O$ *buffer (red) and HMBP in 100 %* $D_2O$ *buffer (blue). Pink and pale blue are neutron intensity plots of 42 %* $D_2O$ *and 100 %* $D_2O$ *buffer respectively with no protein present.* ***B:*** *Buffer subtracted difference intensity plots, log ΔI(q) vs. log q, of DMBP in 42 %* $D_2O$ *buffer (red) and HMBP in 100 %* $D_2O$ *buffer (blue).*

## *Overall comparison of SANS signals of MBP in the presence of increasing lysate concentrations*

Here, three conditions were considered: DMBP mixed with an increasing amount of unlabelled lysate (HL), HMBP mixed with increasing amount of matchout lysate (ML), and HMBP mixed with increasing amount of perdeuterated lysate (PL). The SANS data for these are shown in Figure 7.5. The objective was to identify the conditions that permit the most effective background subtraction with the smallest deviation from the control samples. This would identify the conditions required to obtain intensity profiles of unpurified protein expressed by CFPS in the presence of the lysate (*i.e.* where the lysate of mixed proteins, RNA, DNA etc has been best made "invisible" due to the matching-out conditions).

It was found that the background subtracted profile of HMBP in ML gave the best results, at both low and high lysate percentages (Figure 7.6A). This was expected since ML represents the matchout deuteration level for proteins, which are the main components of the lysate. Inspection of the Guinier region showed a high degree of linearity, indicating the near-absence of aggregates (Figure 7.6B). At a high lysate concentration of 30 % there was a noticeable increase in the slope of the buffer subtracted plot observed (Figure 7.6B).

In the case of HMBP in PL, a much higher signal intensity ($I_o$) was recorded compared to HMBP in ML (Figure 7.5B). This can be explained by the presence of the largest amounts of highly scattering deuterium which is present in PL (perdeuterated) lysate, which is not totally matched out. Due to the poor matchout condition in PL, the buffer subtracted plots show particularly significant deviation from the control (Figure 7.6B).

In the case of DMBP in HL, increasing amounts of lysate resulted in very high increases in the background scattering (Figure 7.5C), swamping the signal from the target protein and resulting in largely mismatched buffer subtracted curves compared to the reference control (Figure 7.6A). Inspection of the Guinier plots (Figure 7.6B) reveals an increasingly significant dip in the low q region with increasing amounts of HL lysate present in the DMBP sample.

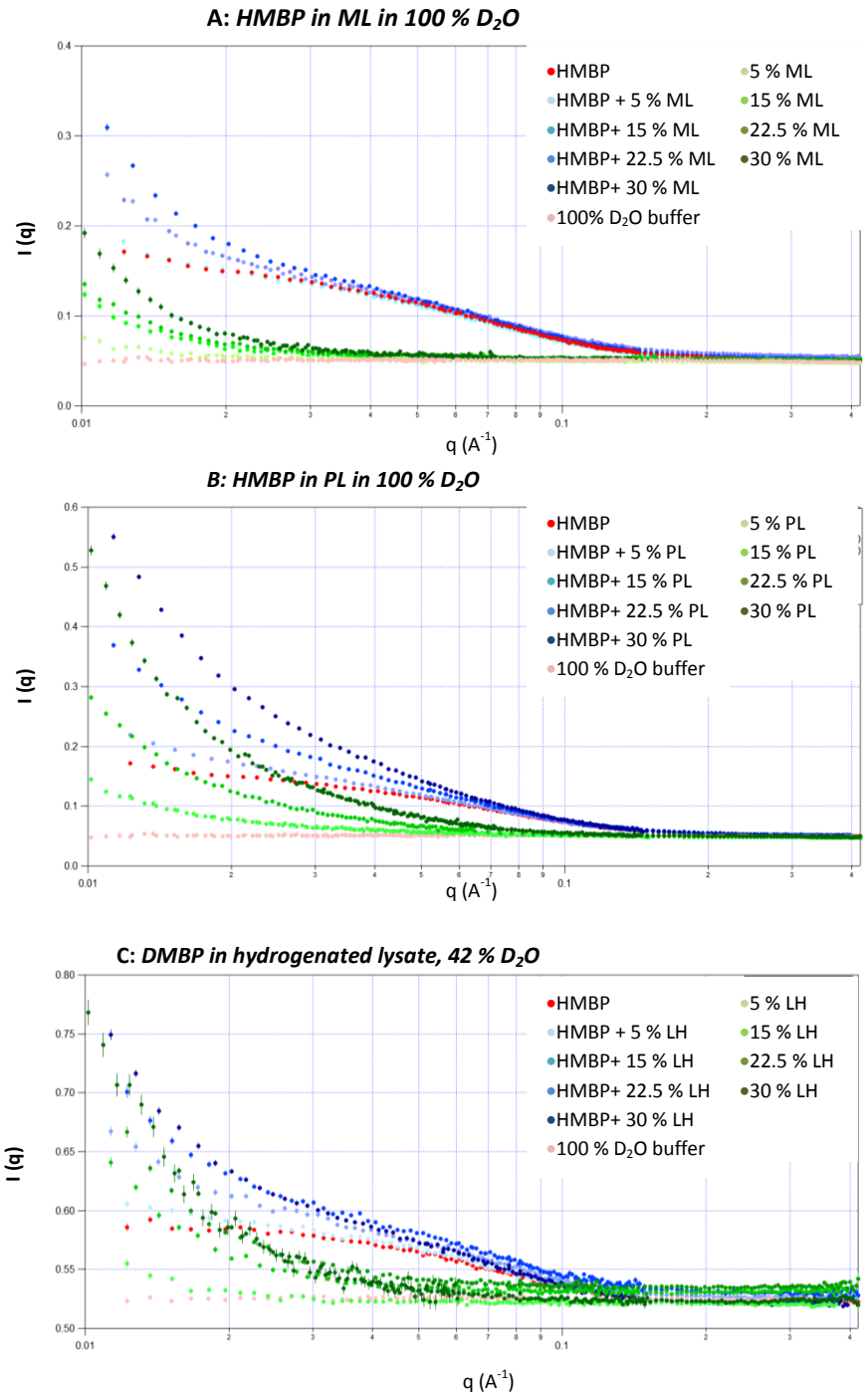



***Figure 7.5:*** *SANS Intensity (I(q) vs q) plots of SANS data for A: HMBP in ML in 100 % $D_2O$ B: HMBP in PL in 100 % $D_2O$ C: DMBP in HL in 42 % $D_2O$, and their corresponding lysate controls. Blue plots show MBP with various amounts of lysate, green shows lysate controls (i.e. various concentrations of lysate alone), red shows pure MBP control (i.e. pure MBP alone in buffer) and pink shows buffer control (i.e. no MBP and no lysate).*

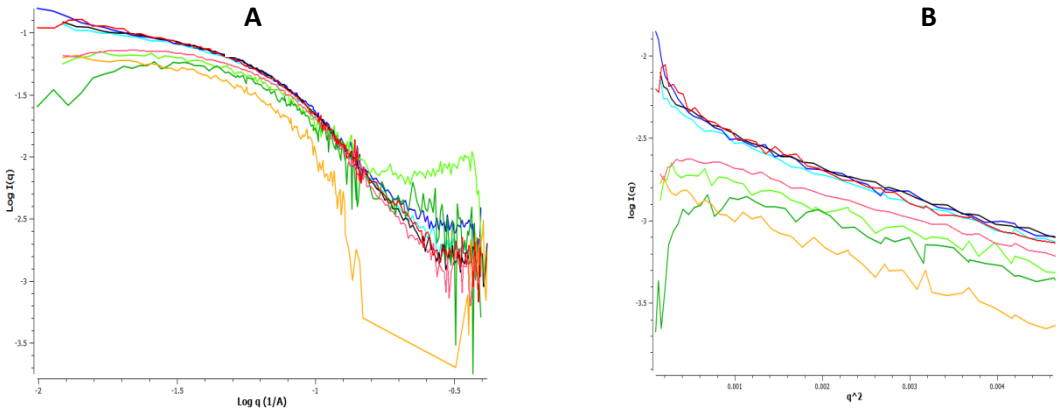



***Figure 7.6:*** *A: Buffer subtracted plots for SANS experiments using HMBP and DMBP lysates (log I(q) vs log q) B: Guinier region plot (log I(q) vs $q^2$). For ease of visualisation, not all of the difference plots are shown. Black: pure HMBP; cyan: HMBP + 5 % L85; blue: HMBP + 30 % L85; pink: HMBP + 5 % L100; red: HMBP + 30 % L100; orange: pure DMBP; green: DMBP + 5 % HL; dark green: DMBP + 30 % HL. The plots closest to the pure HMBP data (shown in black) and pure DMBP data (orange) indicate the conditions the most appropriate for minimising signal from the lysate.*

### *Detection of matched-out Glob signal in ML*

The ability to detect the coherent scattering from a high concentration sample of 10 mg/mL of Glob in CFPS lysate was evaluated (Figure 7.6A). Clear coherent scattering was detected from Glob present in solutions consisting of 30 % of either PL in 100 % $D_2O$ (sample Glob-30PL), or ML in 100 % $D_2O$ (sample Glob-30ML). The intensity plot recorded from these samples closely resemble the reference intensity plot of pure Glob in 100 % $D_2O$ (sample Glob-100D) for mid-to-high q-values (> 0.02). Deviations of the two intensity plots from the reference were noted at low q.

However, upon background subtraction, these deviations at low q were eliminated in the case of the Glob-30ML (Figure 7.6B). These results were further confirmed by the Guinier plots of this sample and the reference sample (Glob-100D) which followed each other very closely with a high degree of linearity (Figure 7.6C). These results show that ML is successfully matched-out in 100 % $D_2O$ solvent. The

deviation at low q in PL is still observed for Glob-30L100. This is because the 100 % deuterated proteins are cannot matched-out in 100 % $D_2O$.

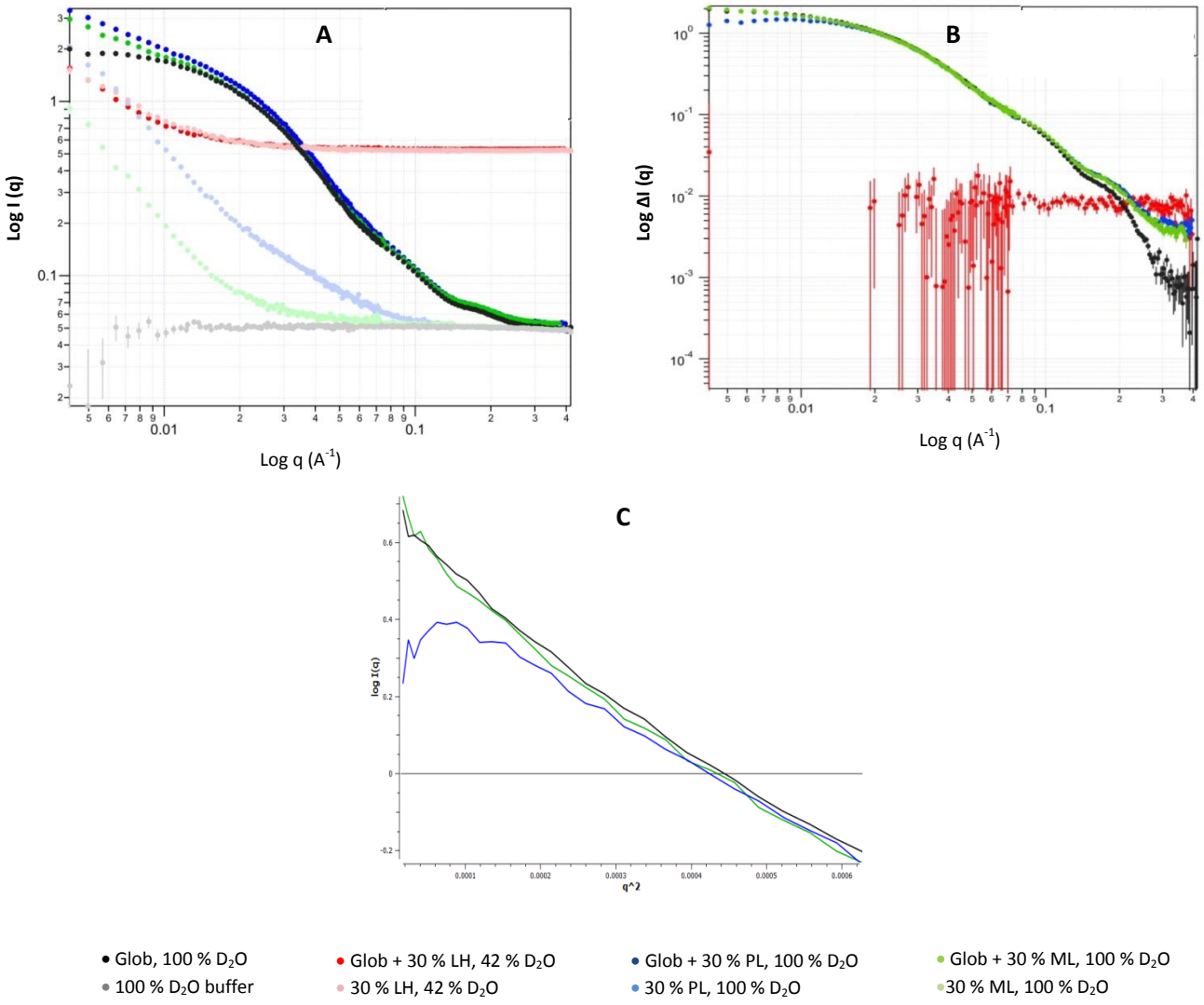



***Figure 7.7:*** *SANS intensity plots for the model protein Glob. A: Intensity plots, log I (q) vs. log q, B: Buffer subtracted difference intensity plots, log ΔI(q) vs. log q, and C: Guinier region and plot, log I(q) vs $q^2$, of Glob in 100 % $D_2O$ buffer (black), in PL in 100 % $D_2O$ buffer (blue), in ML in 100 % $D_2O$ buffer (green) and in HL in 42 % $D_2O$ buffer (red). Pale blue, pale green and pink are plots of PL in 100 % $D_2O$ buffer, ML in 100 % $D_2O$ buffer and HL in 42 % $D_2O$ buffer respectively, excluding Glob. The plots closest to the pure Glob data (shown in black) show the conditions the most appropriate for minimising signal from the lysate.*

### *Assessment of the limits of the coherent signal in ML*

The limits at which the coherent signal of the unlabelled target protein present in a deuterated lysate can be detected were verified. Since from the previous section it was found that ML was most suited for producing a signal which closely followed that of pure protein controls, the use of ML was subsequently selected for further analysis.

In the first experimental setup, the amount of the unlabelled Glob was maintained constant at 10 mg/mL, whereas the amount of lysate was increased to 30 % of the total volume (Figure 7.8A). The 30 % lysate represents the maximum amount of lysate that is used in a typical CFPS reaction. Coherent SANS intensity signals, and near identical profiles to that of pure Glob, were observed for all of the measured conditions. As expected, with the presence of increased lysate concentration in the solution, the plots showed gradual deviation from pure Glob intensity plot. The deviations, of increasing magnitude, were particularly present for low q values (<0.01 $Å^{-1}$). The profiles of controls consisting of lysate at the corresponding concentrations to that of experimental samples, but excluding Glob, served to confirm, and demonstrate the increasing magnitude of the intensity brought about by the increasing lysate concentration (i.e. increasing in the number of scattering particles).

Following the subtraction of the background signal, the plots show minimal deviations in the intensity plot at low q compared to the pure protein control (Figure 7.8B). However, a decrease in the intensity plot at low q could still be observed with increasing lysate concentration. This concentration-dependent decrease in intensity indicates increasing inter-particle repulsion brought about by a high concentration of particles upon the addition of more lysate into the solution. At high q values (>0.1 $Å^{-1}$), compared to pure Glob, the signal of Glob in lysate loses definition due to the increased interference from the high concentration of lysate particles.

Since an adequate signal was still being detected, the limits of detection of Glob were tested by varying its concentration from 10 mg/mL to 1.25 mg/mL while maintaining the amount of lysate constant (30 %) (Figure 7.8C). As the concentration of Glob present was reduced, a steady downward shift in the intensity profile was observed. The key difference between the plots of control pure Glob without lysate (Glob-100D) and those including lysate (Glob-30ML), was the increase in the intensity at low q. This deviation was effectively eliminated with the subtraction of the background (Figure 7.7D). The background-subtracted plots of Glob-100D and Glob-30ML very closely follow each other although there

was a clear deterioration in the signal-to-noise for the Glob-30ML data at high q (>0.1 $Å^{-1}$). Reducing the concentration of Glob resulted in downward shifts in the intensity profiles. However, increasingly significant deviating drops in intensity were observed at very low q with a decrease in the concentration of Glob (<0.01 $Å^{-1}$). This is due to the increase in inter-particle interactions which arise from the scattering interference coming from the lysate.

The series of experiments was repeated using BSA as the model protein (Figure 7.9A, B). Similar shifts in the intensity and buffer-subtracted data were observed.

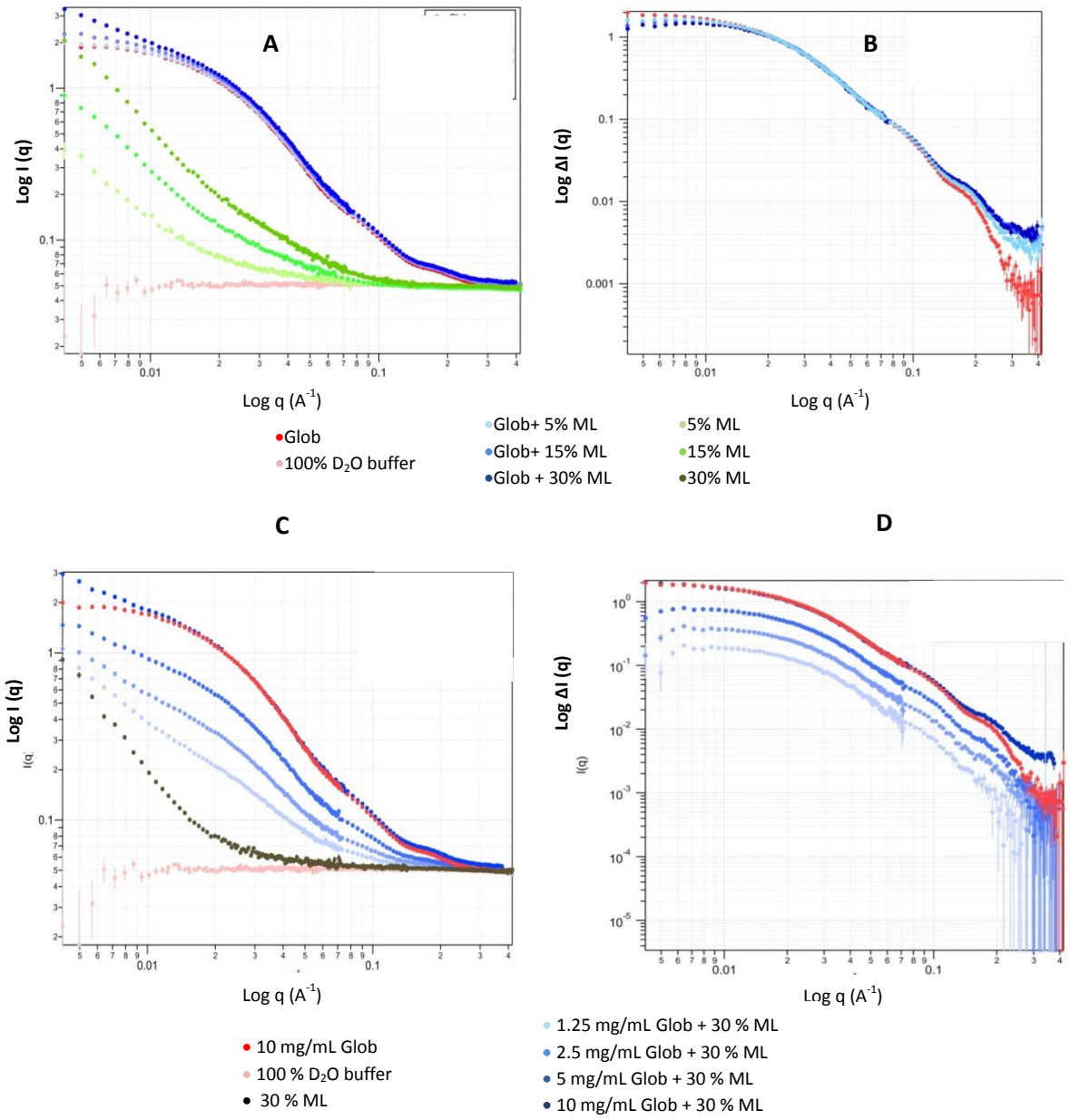



***Figure 7.8:*** *SANS intensity plots (A and C) and buffer subtracted plots (B and D) of Glob in the presence of* ML *in 100 % $D_2O$. A and B: Glob concentration is maintained at 10 mg/mL while the lysate amount is varied. C and D: Glob concentration is varied (1.25 – 10 mg/mL) while the lysate amount is maintained constant (30 %). The plots following closest to the pure Glob data (shown in red) show the conditions the most appropriate for minimising signal from the lysate.*

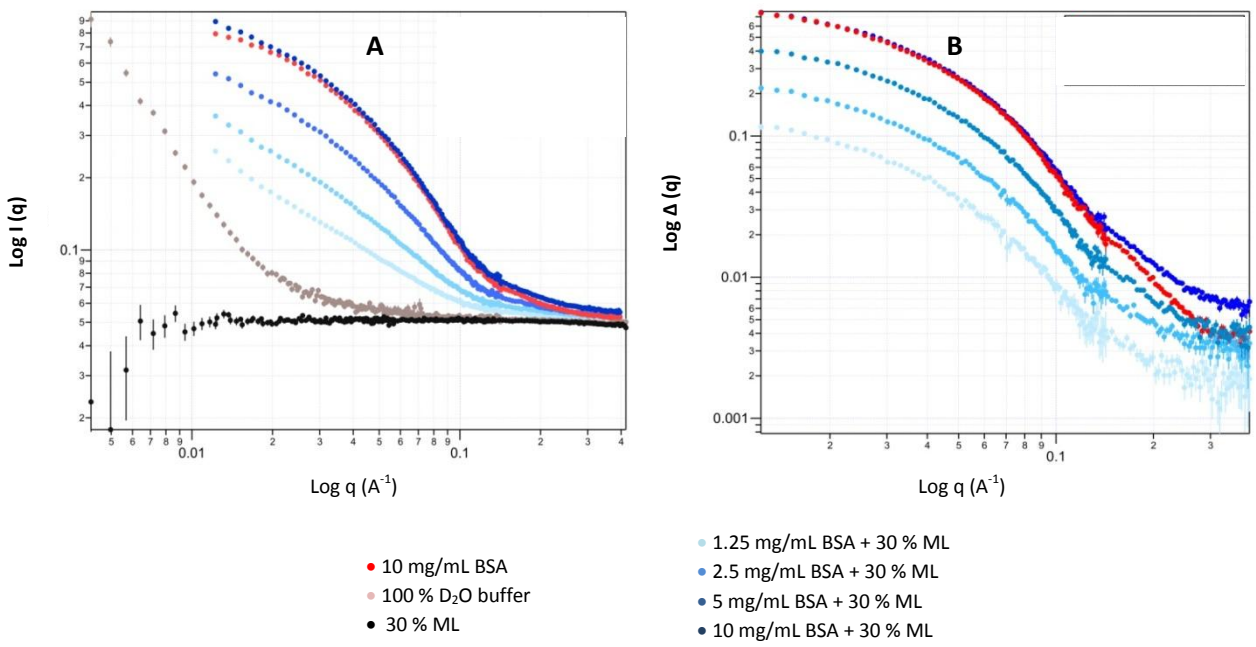



***Figure 7.9:*** *SANS intensity plots (A and B) and buffer-subtracted plots (B and D) of BSA in the presence of ML in 100 % $D_2O$. The BSA concentration is varied (1.25 – 10 mg/mL) whereas the lysate amount is maintained constant (30 %). The plots most closely following the pure BSA data (red) show the conditions the most appropriate for minimising signal from the lysate.*

### *Negligible influence of DNA and tRNA in lysate on neutron intensity profile*

In the CFPS reactions, DNA and tRNA are present at respective concentrations of approximately 15 µg/mL and 1 mg/mL. DNA and tRNA have SLDs values that are very different to those of lysate proteins (Figure 7.10A, section 2.1.4). The causes a difference in the matchout conditions for the two components of the lysate. Since lysate proteins are the overwhelming constituents of the CFPS reaction mixture, it was thought that the presence of DNA and tRNA would have a negligible impact on the data. To test this, buffer-subtracted plots of globulin in 22.5 % of lysate, with DNA/tRNA included or excluded were compared (Figure 7.10B). With both ML and PL, the plots followed each other closely, indicating that the presence of DNA/tRNA did indeed have a relatively small effect. Inspection of the Guinier region further confirmed this observation as plots followed closely overlapped each other (Figure 7.10C).

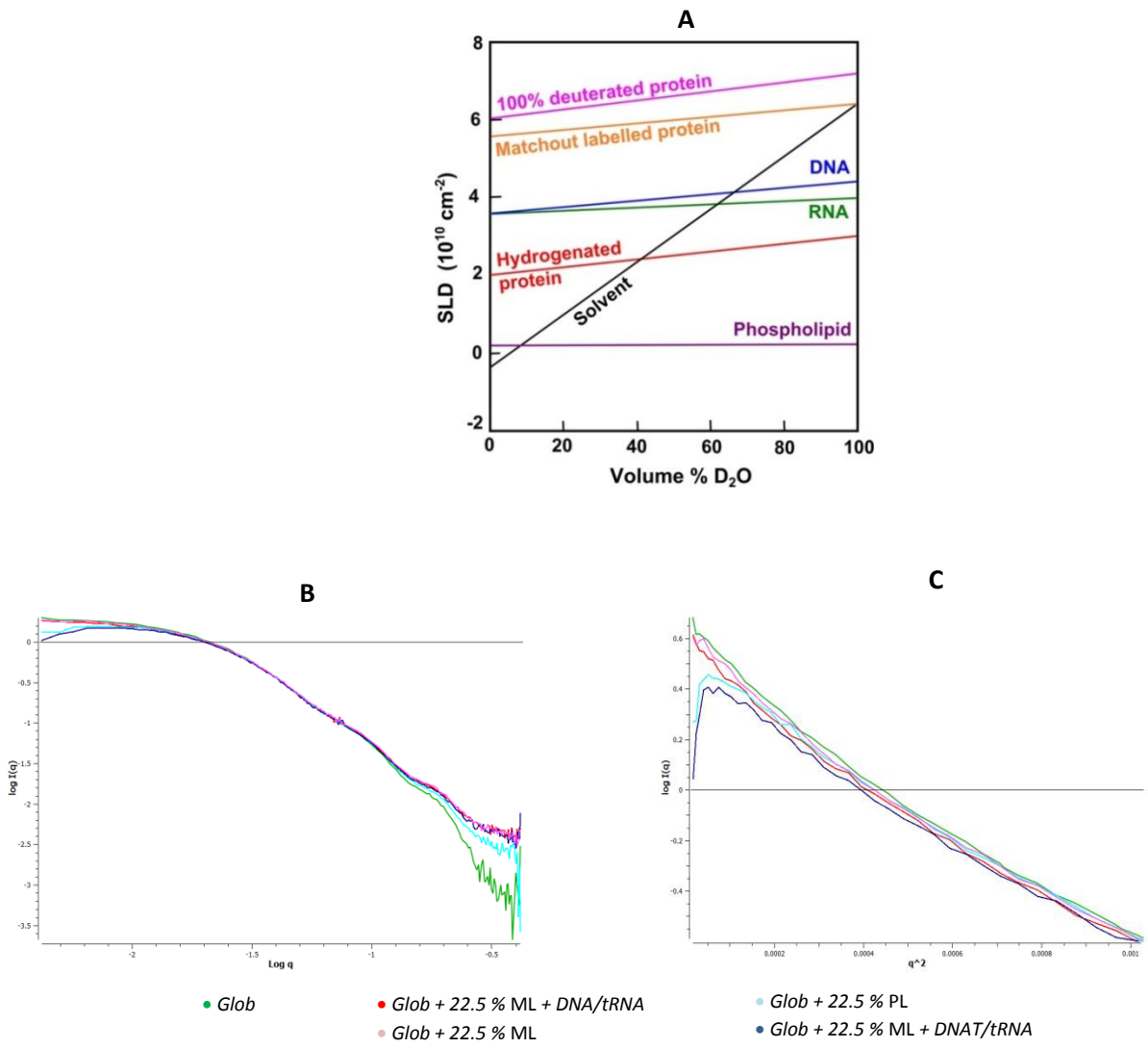



***Figure 7.10:*** *The scattering length densities (SLDs) of different biomolecules as a function of the volume percentage of $D_2O$ (A) (Dunne et al., 2017b). The black line shows the variation in SLD of the solvent as a function of $D_2O$ content. The match-point of each bio-molecule corresponds to the intersection of the solvent SLD with that for each bio-molecule. SANS intensity plots of Glob in 22.5 % of* ML *or PL, with the inclusion and exclusion of DNA/tRNA equivalent to CFPS lysates. (B) Buffer-subtracted plots (log I(q) vs q), (C) Guinier plot (Log q vs $q^2$). Green: Glob-100D; cyan: Glob + 22.5 %* PL*; blue: Glob + 22.5 %* ML *+ DNAT/tRNA; Pink: Glob + 22.5 %* ML*; Red: Glob + 22.5 %* ML *+ DNA/tRNA. The plots most closely following the Glob-100D data (shown in green) show the conditions that are most appropriate for minimising the signal from the lysate.*

### 7.3.2. Interpretation of protein size parameters

Protein size parameters (*Rg* values – both Guinier- and P(r)-derived, MW and Dmax – see the definitions given in chapter two) were extracted from the I(q) vs q data. The objective was to assess if the protein size parameters were consistent over the various lysate conditions studied. The size parameters for the pure sample control of each of the model proteins served as the references for comparison.

Upon comparison with their corresponding pure sample reference, it was found that overall there was high consistency in the protein size parameters of the MBP in the presence of lysate (Table 7.2).

Table A

| Sample | *Rg* Guinier (error)/ Å | $I_o$ /cm | MW/ KDa | *Rg* P(r) /Å | Dmax/ Å |
|---|---|---|---|---|---|
| **DMBP-42D** | 25(1) | 0.064 | 43.4 | 21 | 61 |
| **DMBP-5HL** | 21(1) | 0.071 | 48.1 | 20 | 57 |
| **DMBP-15HL** | 21(2) | 0.063 | 42.7 | 20 | 58 |
| **DMBP-22.5HL** | 20(2) | 0.061 | 41.3 | 22 | 65 |
| **DMBP-30HL** | 20(1) | 0.064 | 43.4 | 21 | 62 |

*Δρ = 3.0054, Theoretical MW: 45.3 KDa*

Table B

| Sample | *Rg* Guinier (error)/Å | $I_o$ /cm | MW /KDa | *Rg* P(r) /Å | Dmax/ Å |
|---|---|---|---|---|---|
| **HMBP-100D** | 25(3) | 0.11 | 44.8 | 24 | 76 |
| **HMBP-5ML** | 25(3) | 0.10 | 40.7 | 25 | 82 |
| **HMBP-15ML** | 25(1) | 0.10 | 40.7 | 23 | 70 |
| **HMBP-22.5ML** | 23(4) | 0.096 | 39.1 | 23 | 74 |
| **HMBP-30ML** | 25(6) | 0.10 | 40.7 | 24 | 75 |
| **HMBP-5PL** | 21(0) | 0.079 | 32.2 | 21 | 71 |
| **HMBP-15PL** | 26(4) | 0.11 | 44.8 | 24 | 76 |
| **HMBP-30PL** | 26(3) | 0.11 | 44.8 | 24 | 70 |

*Δρ = -3.289, Theoretical MW: 42.4 KDa*

***Table 7.2:*** *Summary of size parameters of DMBP (A) and HMBP (B) extracted from the SANS data.*

In the classical format of SANS matching-out experiments in which a deuterated protein is studied against hydrogenated protein background in 42 % $D_2O$ solvent conditions (Dunne *et al.*, 2017b, Haertlein *et al.*, 2016, Laux *et al.*, 2008), the presence of hydrogenated lysate (HL) did not significantly affect the calculated *Rg* value of DMBP (Table 6.2A, 21 ±1 Å for both Guinier and P(r)). This indicates that the hydrogenated lysate was well matched by the solvent SLD. With HMBP in the presence of ML in 100 % $D_2O$, a value of *Rg* near equal to that of pure sample control was obtained (23 ±2 Å). However, a small drop in the calculated MW of HMBP in ML (44.8 KDa → 40.7 KDa) was noticed.

The P(r) plots for DMBP are consistently bell-shaped, indicating high protein globularity (Figure 7.11). The normalised Kratky plots of DMBP in HL in particular shows that the peaks of the bell-curves lie at the cross-hairs which mark the position of $q^*Rg = \sqrt{3}$ (x-axis) and $I(q)/I_o \times (q \cdot Rg)^2 = 3.e^{-1}$ (y-axis). As detailed in section 2.1.5, when a normalised Kratky plot a bell-curve peak at this position, the data can be said to follow the Guinier approximation (Equation 2.2), and the protein of interest can be considered therefore to be monodisperse and globular. In the cases here, DMBP thus follows the Guinier approximation and is monodisperse and globular. However, in the case of HMBP, in either the presence of ML or PL, the Kratky plots show widened bell-curves (Figure 7.11). Increasing upturn at high q in the Kratky plots with increasing lysate concentration is furthermore observed. Deviation of the bell-curve peak from the idealised point (as shown by the cross-hairs) would normally indicate non-globular, disordered or unfolded proteins. In this case however, deviation of the protein peak from the ideal is due the loss of the clear-cut distinction at the interface of MBP and the solvent (Figure 7.12). As mentioned in section 2.1.5, high q range data from SANS experiments, provides information about the surface of the protein particle at the protein-solvent interface. The increasing concentration of lysate biomolecules leads to its increasing presence and interaction with protein of interest, which negatively affects the quality of the SANS signal at high q range (Figure 7.12, and as observed from previous 1D SANS intensity plots 7.7 - 7.9).

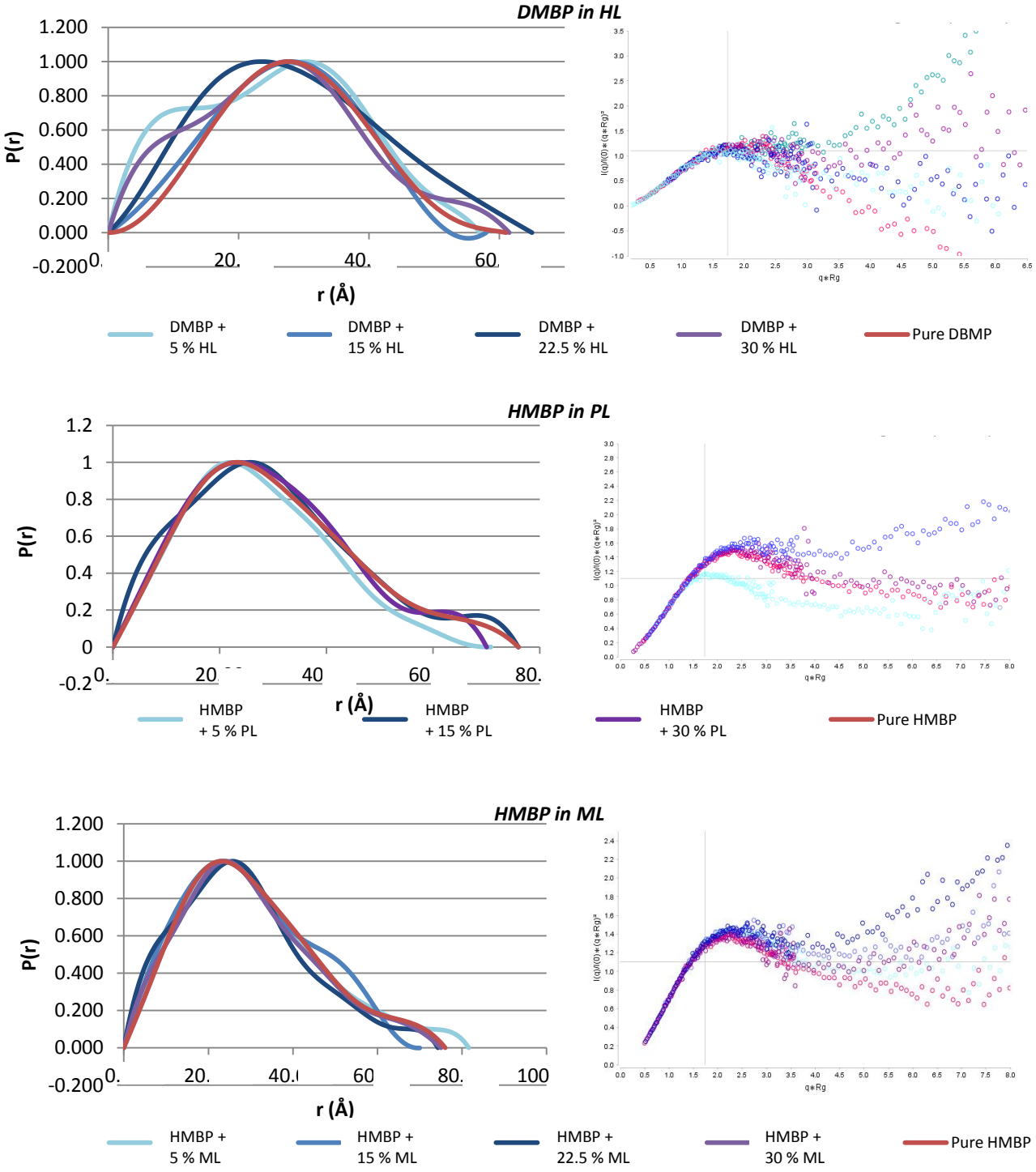



***Figure 7.11:*** *P(r) and Kratky plots of HMBP and DMBP in the presence of increasing lysate amount (5 % - 30 %). Plots should ideally follow the pure MBP control (red). Due to increasing interaction of the lysate with MBP however, the plots show deviation, which results in the loss of reliability of the information about the surface of MBP.*

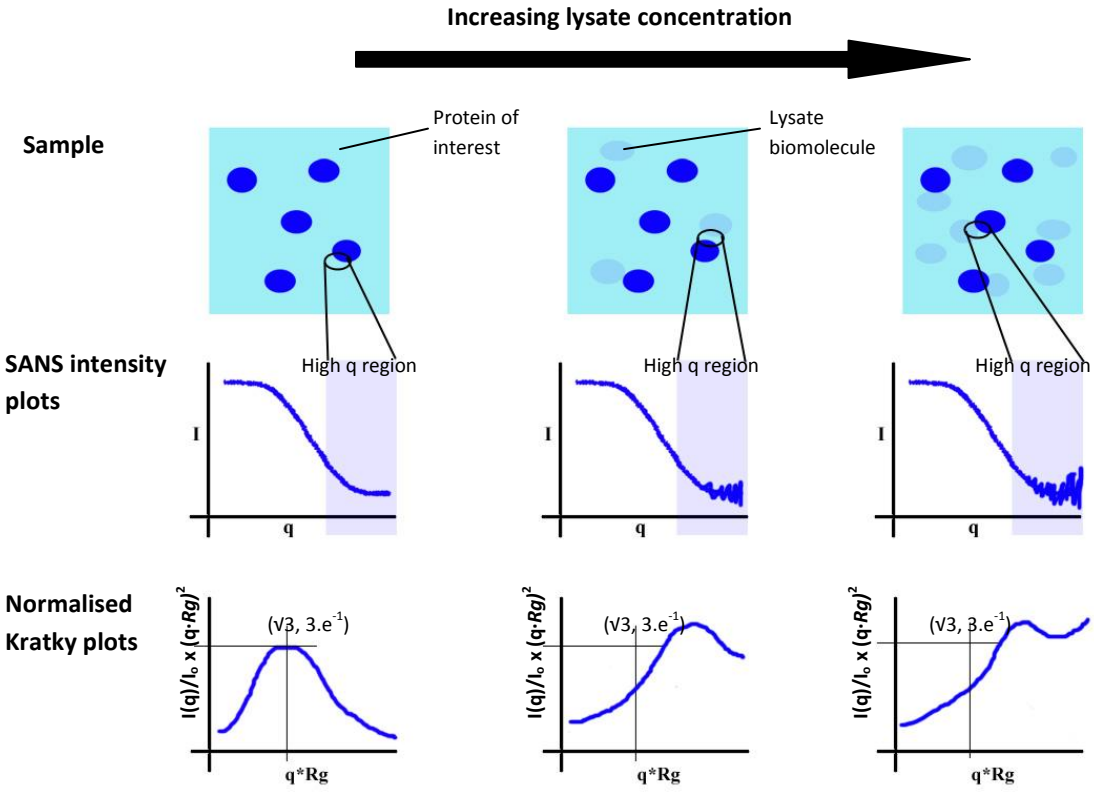



***Figure 7.12:*** *The effect of increasing lysate concentration on SANS data. As the concentration of lysate is increased, the lysate biomolecules are present in large amounts in close proximity to the protein of interest. Information about the surface of the protein of interest at the protein-solvent interface is provided by the region of high q in SANS intensity plots. With increasing lysate concentration, there is the loss in the definition of surface the protein of interest, resulting in loss of interpretability of the neutron signal at high q. This is reflected in normalised Kratky plots which deviate from an ideal monodisperse globular bell-shape (with a peak at [√3, 3.e$^{-1}$]) to give bell-curves with increasing upturn for high q·Rg values.*

With BSA at various concentrations in the presence of ML, strong consistency in the calculated parameters is noted (Tables 7.3). However, the BSA used consisted of a mixture of monomeric and dimeric forms. Furthermore, in solution, the reversible monomer-dimer equilibrium is a rapidly shifting one (Levi and Gonzalez Flecha, 2002). Thus, the parameters determined from BSA here, in particular the over-estimation in the calculated MW (69.3 – 87.5 kDa), reflect the contribution of monomeric (66 kDa) and dimeric (132 kDa) forms of BSA. The profile P(r) and Kratky plot profiles of BSA (Figure 7.13) are similar to those for MBP. Increasing upturn in at high q in the Kratky plot with decreasing protein to lysate ratio therefore indicates the loss of reliability of the information about the surface of BSA.

| Sample | *Rg* Guinier (error) (Å) | $I_o$ (/cm) | MW (KDa) | *Rg* P(r) (Å) | Dmax (Å) |
|---|---|---|---|---|---|
| **BSA-100D** | 45 (1) | 0.81 | 73.9 | 44 | 147 |
| **10BSA-30ML** | 40 (4) | 0.76 | 69.3 | 43 | 154 |
| **5BSA-30ML** | 44 (3) | 0.43 | 78.4 | 45 | 150 |
| **2.5BSA-30ML** | 44(5) | 0.23 | 83.9 | 47 | 157 |
| **1.25BSA-30ML** | 45 (1) | 0.12 | 87.5 | 44 | 147 |

*Δρ = -3.289, Theoretical MW: 66 KDa(monomer), 132 (dimer)*

***Table 7.3:*** *Summary of size parameters of BSA*

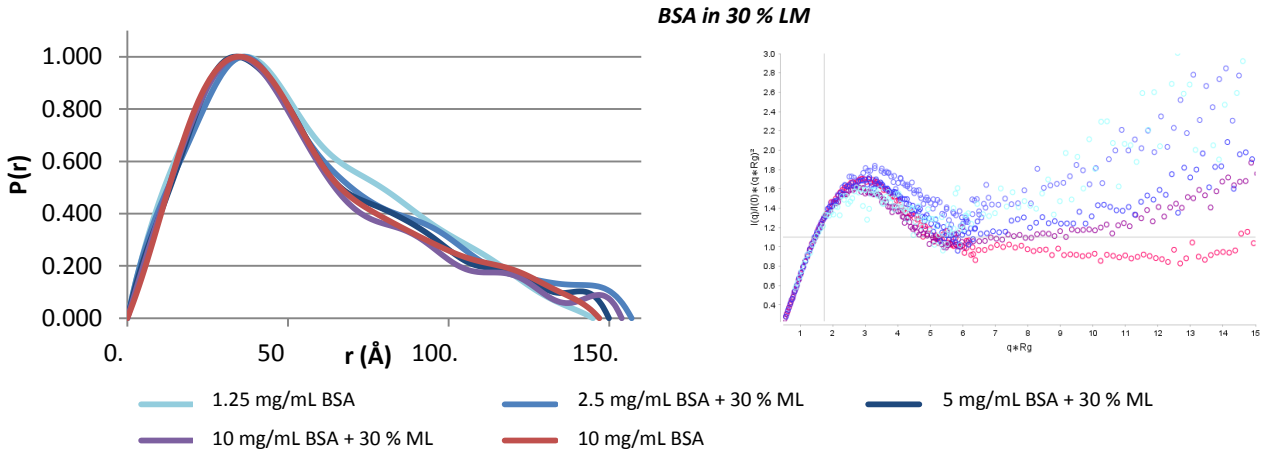



***Figure 7.13:*** *P(r) and Kratky plots of BSA in the presence of lysate. Plots should ideally follow the pure BSA control (red). In addition to the increasing interaction of the lysate with BSA, the formation of dimeric BSA leads the plots to deviate from the pure BSA control and there is the loss of reliability of the information about the surface of BSA.*

Finally, strong consistency in the size parameters of Glob in the presence of lysate is observed (Table 7.4). As reported in a previous publication which discussed the characterisation of Glob by SAXS, Glob consists of various oligomeric forms, of which the monomer and dimer of immunoglobulin G (IgG) are the most abundant (Da Vela *et al.*, 2017). In the same study, the *Rg* and Dmax of monomeric IgG were found to be respectively 53 ±4 Å and 184 Å, whereas those of dimeric IgG was found to be respectively 76 ±2 Å and 223 Å. Here, the SANS analysis on pure Glob without any lysate included, gave *Rg* and Dmax of 71 ±2 Å and 241 Å respectively. These values are consistent with the previous study and strongly indicate the dominance of dimeric forms of IgG present. The increasing presence of lysate did not change significantly the recorded size parameters. However, with lowered Glob concentration while maintaining high lysate concentration, a shift to monomeric form of IgG was observed as *Rg*, MW and Dmax were reduced.

The P(r) plots of Glob, despite the increasing presence of lysate, were consistent (Figure 7.14). The double peak and right-shifted profile is consistent with that of IgG (Da Vela *et al.*, 2017), and indicate a multi-domain protein separated by flexible linkers. The multi-lobar and high flexibility of the Glob is also reflected in the Kratky plot, and is consistent with the structure of IgG: two Fab and Fc portion that are arranged as a three-lobed structure separated by the flexible hinge region (Harris *et al.*, 1997). Once again, an increasing lysate concentration resulted in an upturn of the Kratky plots at high q, indicating the loss of reliability of information about the surface of Glob.

| Sample | *Rg* Guinier (error) /Å | $I_o$ /cm | MW/ kDa | *Rg* P(r) /Å | Dmax /Å |
|---|---|---|---|---|---|
| **Glob-100D** | 71(2) | 1.95 | 222.5 | 71 | 241 |
| **Glob-5ML** | 62(1) | 1.81 | 206.5 | 70 | 240 |
| **Glob-15ML** | 71(1) | 1.85 | 211.1 | 71 | 240 |
| **Glob-30ML** | 69(1) | 1.88 | 214.5 | 70 | 239 |
| **5Glob-30ML** | 65(1) | 0.84 | 191.7 | 64 | 220 |
| **2.5Glob-30ML** | 61(1) | 0.42 | 191.7 | 61 | 182 |
| **1.25Glob-30ML** | 60(1) | 0.21 | 191.7 | 60 | 186 |
| **Glob-5PL** | 65(1) | 1.81 | 206.5 | 67 | 230 |
| **Glob-15PL** | 62(1) | 1.75 | 199.6 | 66 | 230 |
| **Glob-30PL** | 60(1) | 1.67 | 190.5 | 63 | 230 |

*Δρ = -3.147, Theoretical MW: 150 KDa (monomer), 300 kDa (dimer)*

***Table 7.4:*** *Summary of size parameters of Glob.*

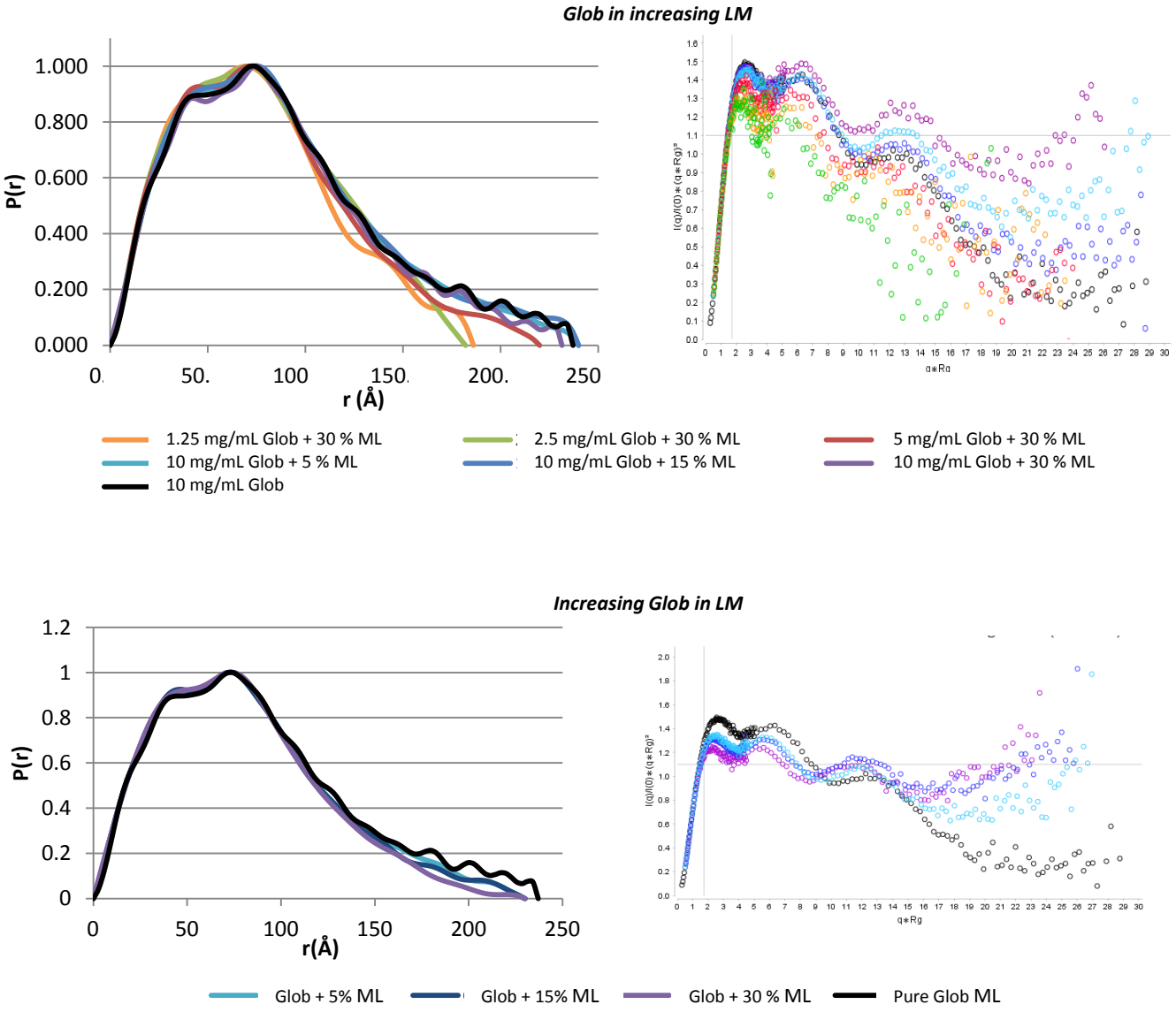



***Figure 7.14:*** *P(r) and Kratky plots of Glob in the presence of lysate. Note the right-shifted double peak that is indicative of multi-domain protein separated by flexible linkers, consistent with that of IgG as previously reported (Da Vela et al., 2017). The plots should ideally follow the pure Glob control (black). Due to increasing interaction of the lysate with Glob however, the plots show upturn deviation resulting the loss of reliability of the information about the surface of Glob.*

### 7.3.3. Capability of modelling proteins despite presence of CFPS lysate

This section focuses on the modelling of the proteins of interest as studied in the various lysate conditions. *Ab initio* modelling, rigid-body model building and fitting was carried out using the list of computational tools as described in section 6.2.3. Firstly, *ab initio* models of the protein in the presence of lysate were compared to the corresponding *ab initio* models of the pure protein in the absence of lysate. A strong resemblance between the models would tend to validate this type of approach – although it should be noted that the lysate itself may affect the protein structure – quite possibly in a way that may be of biological relevance. In addition, rigid-body modelling and fitting using a corresponding crystal structure of the protein was deployed to provide further evidence of the ability to model the proteins using the data.

#### *MBP models*

The previously determined size parameters show MBP to be (as known) a highly globular and monomeric structure, which remains relatively uninfluenced by the presence of CFPS lysate (Figure 7.15). The *ab initio* models generated of MBP reflect these criteria as strong consistency in the shape of the models was maintained despite the increasing concentration of lysate present. Particularly striking was the strong fit between pure DMBP and DMBP in the presence of 30 % unlabelled lysate. Superposition of crystal structure 1JW4 to the *ab initio* models showed high goodness-of-fit (NSD) parameters values close to one. Indicating very strong fits, the NSD values of *ab initio* models of pure MBP against their counterparts in the presence of 30 % lysate were below 1.

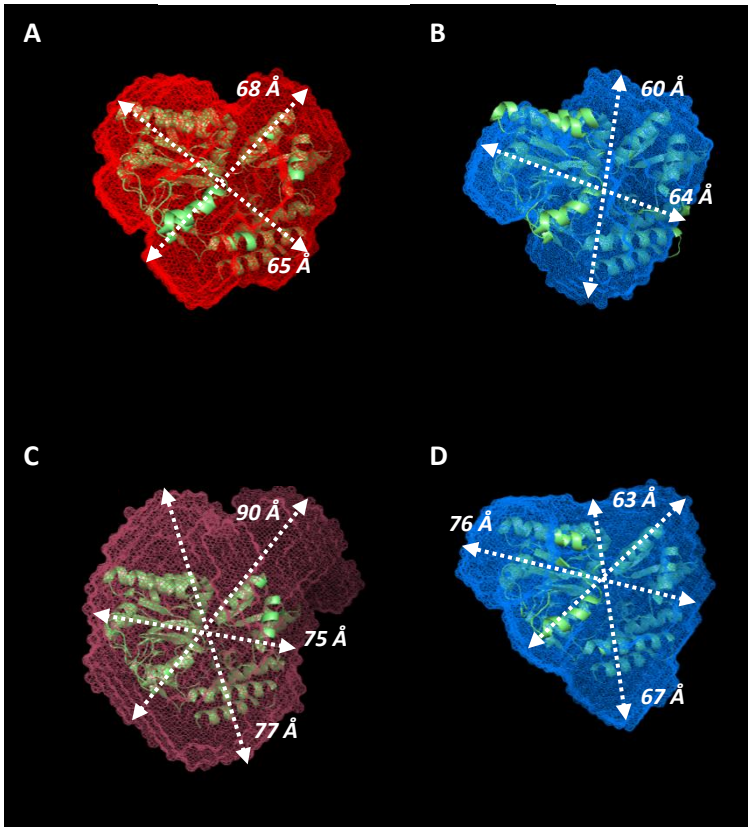



**Table E**

| Sample | Description | DAMAVER vs. 1JW4 | DAMAVER vs. pure sample |
|---|---|---|---|
| **DMBP-42D** | DMBP | 1.22 | NA |
| **DMBP-30HL** | DMBP + 30 % HL | 1.16 | 0.77 |
| **HMBP-100D** | HMBP | 3.60 | NA |
| **HMBP-30ML** | DMBP + 30 % ML | 1.51 | 0.65 |

***Figure 7.15:*** *Superposition of the crystal structure of MBP (PDB ID: 1JW4) onto ab initio DAMAVER models of MBP generated from experimental SANS data. A: DMBP-42D, B: DMBP-30HL, C: HMBP-100D, and D: HMBP-30ML. Table E shows goodness-of-fit parameters (NSD values) of models against the crystal structure, and against the corresponding pure MBP sample.*

### *BSA models*

From the calculated size parameters, BSA was shown to occur as monomers and dimers. To confirm the existence of the two species, CRYSON was used to generate and fit I(q) vs q profiles of the BSA monomer and dimer (Figure 7.16) using the crystal structure of BSA (PDB: 4F5S). The average of the two plots was found to follow and resemble best the experimental profile. The *ab initio* model calculated for pure BSA gives an elongated torpedo-shaped structure. The model shows a reasonable fit to the crystal structure of dimeric BSA 4F5S (DAMAVER NSD = 4.70). The overall shape of the *ab initio* model of BSA was well-maintained even in the presence of 30 % lysate (Figure 7.17B). Deterioration in the model was only registered at the lowest BSA-to-lysate ratio (1.25 mg/mL BSA in 30 % ML, Figure 7.17J). SASREF-generated BSA models with good fits to experimental neutron scatter data (Figure 7.17, Table 7.5), showing that in fact the monomers of BSA are quite separate from each other and display a certain freedom in rotation. This would explain the discrepancy in the recorded MW as determined previously (Table 7.3)

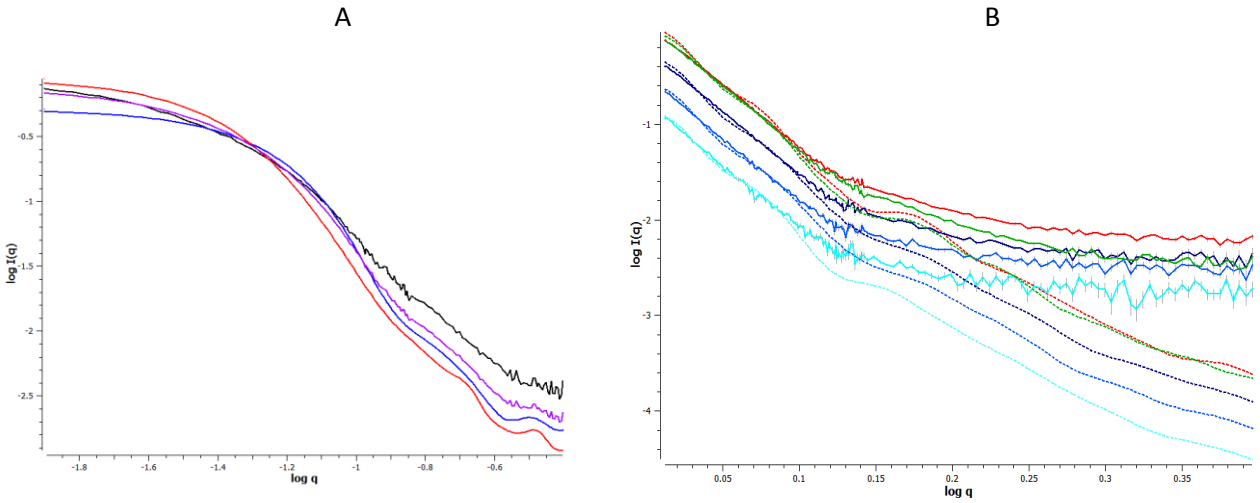



***Figure 7.16:*** *Comparison of SANS plots of experimental, theoretical and modelled BSA.* ***A:*** *I(q) vs q plot of experimental pure BSA (black), CRYSOL generated and fitted plot of monomer (blue) and dimer (red) of BSA (PDB: 4F5S), and the average plot of the monomer-dimer (purple). The average plot of the monomer-dimer plot follows experimental plot better than plots of either monomer or dimer alone, indicative of the existence of monomer-dimer equilibrium in the experimental sample of BSA.* ***B:*** *SASREF fits (dotted lines) against experimental neutron scatter data (solid lines). Red: pure 10 mg/mL BSA; green: 10 mg/mL BSA + 30 %* ML*; purple: 5 mg/mL BSA + 30 %* ML*; blue: 2.5 mg/mL BSA + 30 %* ML*; cyan: 1.25 mg/mL BSA + 30 %* ML*. For low q values, the SASREF plots closely follow the experimental plots, indicative of good modelling.*

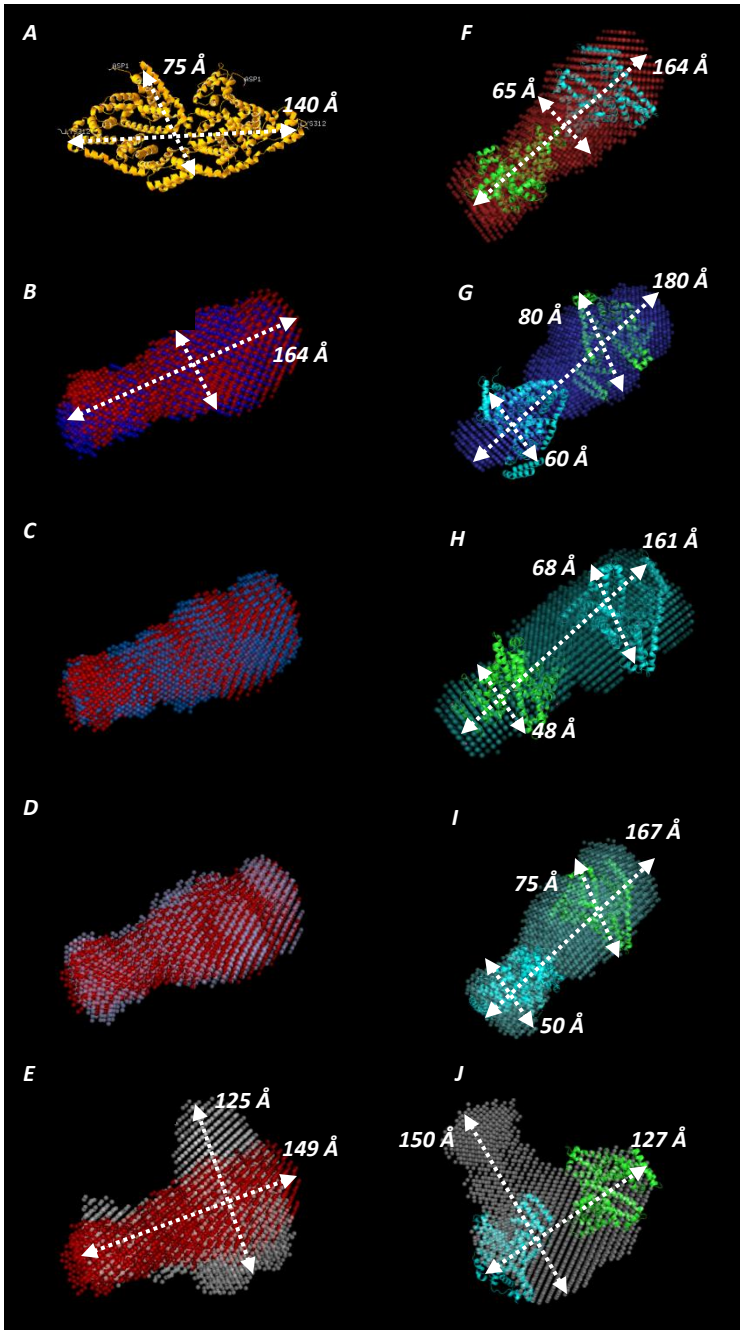



***Figure 7.17:*** *Generated ab initio model fitted and SASREF generated BSA models.* ***Left****, super-imposition of DAMAVER ab initio model of pure 10 mg/mL BSA (Red) against corresponding DAMAVER ab initio model of (B) 10 mg/mL BSA + 30 %* ML*; (C) 5 mg/mL BSA + 30 %* ML*; (D) 2.5 mg/mL BSA + 30 %* ML*; (E) 1.25 mg/mL BSA + 30 %* ML***.*** ***Right****, super-imposition of SASREF models against DAMAVER ab initio model of (F) Pure 10 mg/mL BSA, (G)10 mg/mL BSA + 30 %* ML*; (H) 5 mg/mL BSA + 30 %* ML*; (I) 2.5 mg/mL BSA + 30 %* ML*; (J) 1.25 mg/mL BSA + 30 %* ML*. (A) is PDB crystal structure of BSA 4F5S.*

| Sample | Description | DAMAVER vs. 4F5S | DAMFILT vs. 4F5S | DAMAVER vs. Pure BSA | SASREF model vs. DAMAVER | SASREF vs. DAMFILT |
|---|---|---|---|---|---|---|
| **BSA-100D** | BSA (pure) | 4.70 | 2.45 | NA | 3.51 | 1.93 |
| **BSA-30ML** | 10 mg/mL BSA + 30 % ML | 4.78 | 2.97 | 0.99 | 3.85 | 2.63 |
| **5BSA-30ML** | 5 mg/mL BSA + 30 % ML | 4.52 | 2.95 | 0.52 | 3.73 | 2.63 |
| **2.5BSA-ML** | 2.5 mg/mL BSA + 30 % ML | 5.73 | 4.38 | 0.54 | 4.29 | 2.98 |
| **1.25BSA-ML** | 1.25 mg/mL BSA + 30 % ML | 11.65 | 9.90 | 1.20 | 6.58 | 10.75 |

***Table 7.5:*** *Summary of NSD values of BSA models generated.*

## *Glob models*

As indicated by the low NSD values, the DAMAVER/DAMFILT *ab initio* models of pure Glob showed high consistency with the models of Glob which were derived from the samples in which lysate was present (Figure 7.18, Tables 7.6 and 7.7). The strong fidelity between the *ab intio* models was maintained between the pure sample of Glob and even the sample of Glob containing 30 % lysate concentration. This corresponds to the typical amount of lysate that is used during CPFS. The fidelity of the models with the pure sample was however reduced compared to samples with less than 5 mg/mL of Glob in the presence of 30 % lysate. This was due to the shift in dimeric forms of IgG to monomeric forms as the concentration of Glob was decreased. This observation is consistent with the previous data relating to the size parameters, where a significant drop in Dmax was observed as from 2.5 mg/mL concentration of Glob in 30 % lysate (Table 7.4).

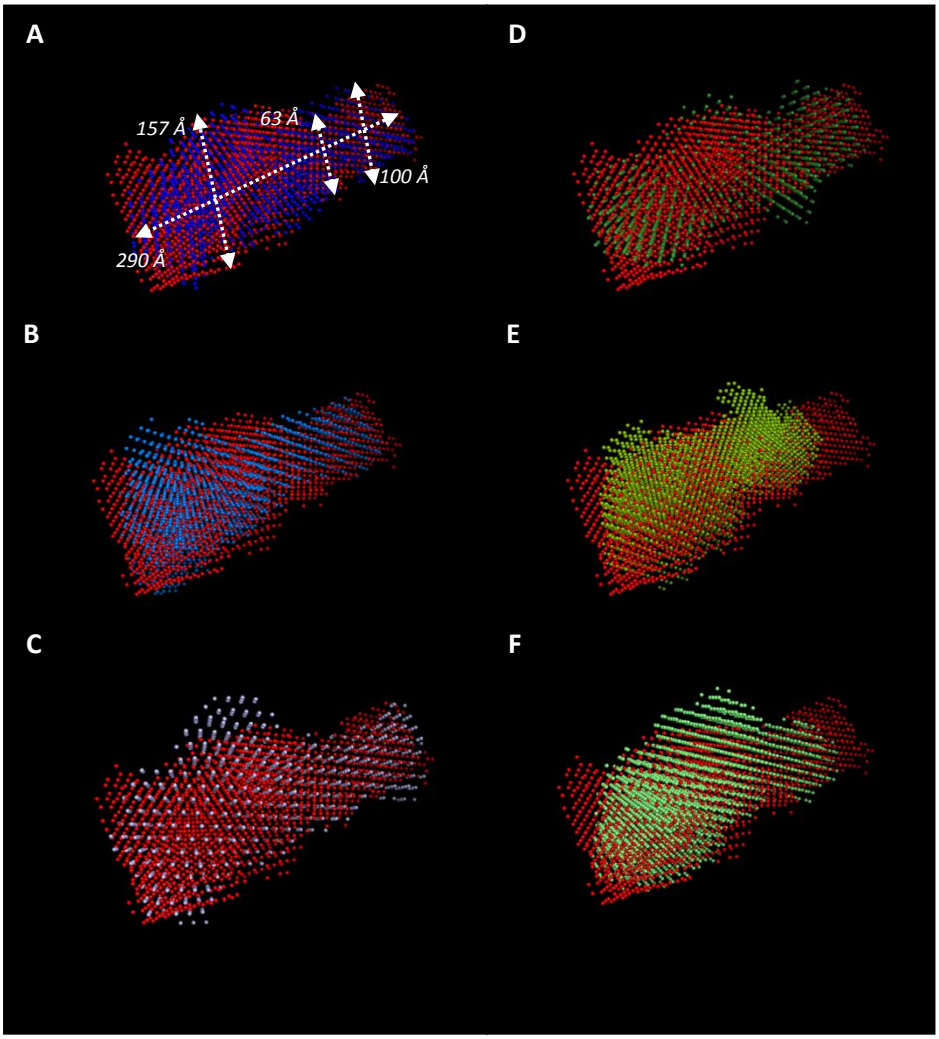



***Figure 7.18*: *Superposition of DAMAVER ab initio model of pure 10 mg/mL Glob (Red) against corresponding DAMAVER ab initio model of (A) 10 mg/mL Glob + 30 %* ML*; (B) 10 mg/mL Glob + 15 %* ML*; (C) 10 mg/mL Glob + 5 % L85; (D) 5 mg/mL Glob + 30 %* ML*; (E) 2.5 mg/mL Glob + 30 %* ML*; (F) 1.25 mg/mL Glob + 30 %* ML.

| Sample | Description | DAMAVER vs pure | DAMFILT vs pure |
|---|---|---|---|
| **Glob-5ML** | Glob + 5 % ML | 0.69 | 1.59 |
| **Glob-15ML** | Glob + 15 % ML | 1.05 | 2.21 |
| **Glob-30ML** | Glob + 30 % ML | 0.68 | 0.80 |

***Table 7.6:*** *Summary of NSD values of ab initio Glob (at 10 mg/mL) models in the presence of increasing lysate amount against ab initio model of pure Glob (at 10 mg/mL).*

| Sample | Description | DAMAVER vs pure | DAMAVER vs 2.5Glob-30ML | DAMFILT vs pure | DAMFILT vs 2.5Glob-30LML |
|---|---|---|---|---|---|
| **Glob-30ML** | 10 mg/mL Glob + 30 % ML | 0.68 | NA | 0.80 | NA |
| **5Glob-30ML** | 5 mg/mL Glob + 30 % ML | 0.77 | NA | 0.94 | NA |
| **2.5Glob-30ML** | 2.25 mg/mL Glob + 30 % ML | 1.41 | NA | 2.24 | NA |
| **1.25Glob-30ML** | 1.25 mg/mL Glob + 30 % ML | 1.01 | 0.70 | 1.72 | 0.94 |

***Table 7.7:*** *Summary of NSD values of ab intio Glob (decreasing Glob concentration, constant 30 % lysate) models against ab initio model of pure Glob (at 10 mg/mL).*

As mentioned previously, the sample of immunoglobulin used here consists of a mixture of IgG (80 %), IgM (10 %) and IgA (< 10 %). As a result the interpretation of *ab initio* models is complicated since large regions of the models can be attributed to the scatter arising from the various forms of antibodies present in the heterogeneous sample. Furthermore, due to the high flexibility of the antibodies, the protein domains have larger zones of effective occupancy in the models, which are unaccounted for when using the static and compact crystal structure of IgG. Nevertheless, using the PDB crystal structure model 1IGT, models of dimeric IgG generated by SASREF were superimposed onto DAMAVER/DAMFILT *ab initio* models (Table 7.8). Reflecting the high flexibility of the molecule, high NSD values (8.00 – 11.62), indicative of poor fitting, were reported for the DAMAVER models. However, with the exception of Glob in the presence of 15 % lysate, the DAMFILT models, which represent the “core structure” of the model, were found to give improved (lower) NSD values (2.82 – 4.51).

| Sample | Description | DAMAVER | DAMFILT |
|---|---|---|---|
| **Glob-100D** | Glob Pure | 8.86 | 2.82 |
| **Glob-5ML** | Glob + 5 % ML | 8.00 | 4.51 |
| **Glob-15ML** | Glob + 15 % ML | 11.62 | 11.97 |
| **Glob-30ML** | Glob + 30 % ML | 8.04 | 4.08 |

***Table 7.8:*** *Summary of NSD values of SASREF models against ab initio Glob models.*

Dimeric models of Glob Pure and Glob + 30 % of ML were generated using PDB crystal structure 1IGT in SASREF (Figure 7.19A, B). Contacts were shown to be made laterally between the heavy chain of one Fab of an IgG molecule, and the heavy chain of one Fab of the other IgG molecule. The antigen binding site of these Fab make further contacts with the more distal domain of the Fc portion in the hinge region of the adjacent IgG molecule. The Fab that seemingly do not contribute to any contacts between the IgG molecules are located at the opposite sides of the model, with their antigen binding site pointing away from the model. With resemblance between the models of pure Glob and Glob + 30 % ML, the ability to define structurally a protein despite the presence of lysate is further confirmed.

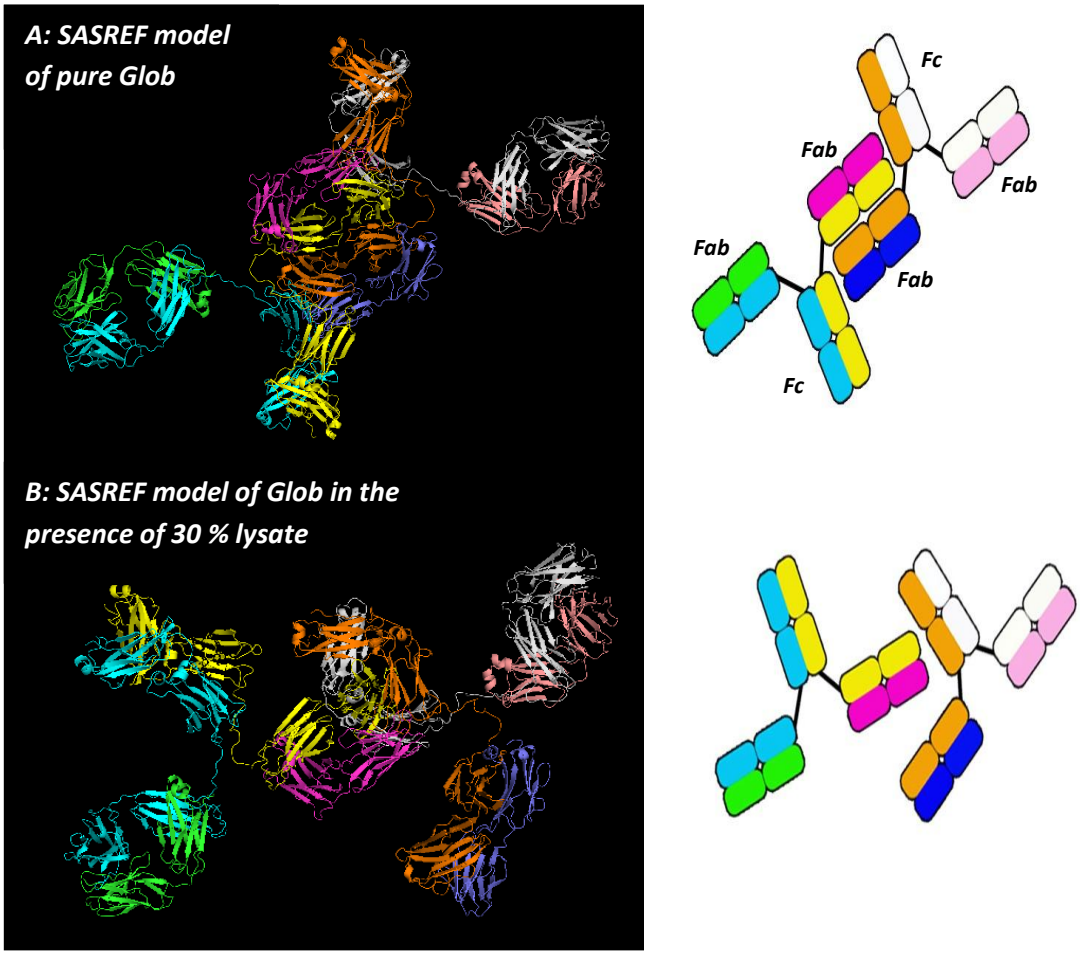



*Figure 7.19: SASREF generated models using PDB crystal structure 1IGT (left) and corresponding cartoons (right) of Glob Pure and Glob + 30 % of ML samples, as dimeric IgG. SASREF generated models indicate that in the dimer of IgG, contacts occur between one of the Fab regions of each monomer, and between the Fab region the upper portion of Fc fragment of the opposing monomer. The resemblance between models therefore indicates the ability to structurally characterise proteins despite the presence of lysate.*

## 7.4. DISCUSSION

### 7.4.1. Successful capability of matching-out CFPS lysate

Here, it is shown that the concept of contrast variation and selective deuteration can be extended to allow an effective "average matchout" of CFPS lysate components. Initially, it was suggested that the logical starting point of analysing deuterated protein in hydrogenated lysate would be problematic. This would be due to the very high levels of hydrogen incoherent scattering from the lysate and that could hide the signal from the deuterated target. However, it was shown here that unlabelled hydrogenated lysate (HL) could be matched-out in 42 % $D_2O$ to permit visualisation of deuterated model protein MBP (Figure 7.1). Conversely, the second approach of anylsing hydrogenated model protein while matching out the deuterated lysate protein, was suggested to be produce much lower levels of hydrogen incoherent scattering against which the coherent SANS signal of the target protein. It was shown that deuterated lysate could be effectively matched out in 100 % $D_2O$ to obtain a signal from a hydrogenated model protein. As expected, ML had much better matchout performance compared to PL since it has a matchout point that better corresponds to that of 100 % $D_2O$ solvent (Figure 7.5 – 7.9).

### 7.4.2. Retaining SANS signal of model protein despite presence of CFPS lysate

Upon background subtraction, it was shown that the SANS data recorded for the target proteins studied were well-conserved in the presence of the highly concentrated and heterogeneous CFPS lysate (Figures 7.5 – 7.9). Indeed, an adequate neutron scattering signal, and interpretable P(r) and Kratky plots (Figure 7.11 – 7.14) could even be obtained despite the presence of 30 % CFPS lysate (which represents the amount used for the typical CFPS reaction). For the conditions tested here, only a slight deviation in the Guinier region was perceived. The quality of the data subsequently enabled the extraction of consistent size parameters of the model proteins tested (Tables 7.2 – 7.4). Additionally, these values reflect the change in oligomeric state of the proteins due to an increasing lysate-to-protein ratio or due to decrease in protein-to-lysate ratio. These results therefore show the feasibility of obtaining reliable SANS structural information in conditions that can be compared to unpurified protein produced in a real CFPS reaction. Some deterioration of the SANS signal, mostly in the high-q regions, was noted and was attributed to the background against which the SANS data had to be measured. The high q range

provides information about the surface of the protein particle at the protein-solvent interface (see section 2.1.5).

### 7.4.3. Ability to model proteins despite the presence of CFPS lysate is dependent upon the protein of interest being observed

It was shown that even in the presence of CFPS lysate, it was still possible to obtain good *ab initio* models of the protein of interest (Figures 7.15 – 7.18). Notably, the SANS data recorded from DMBP, even in the presence of 30 % HL, could be fitted well and was closely comparable to the pure sample control (Figure 7.15).

In addition to key viability issues (e.g. solubility and concentration of the target protein in lysate conditions), the quality of the *ab initio* models obtained is dependent upon the intrinsic properties of the protein of interest. For instance, MBP is an ideal model protein since it is highly homogenous, monomeric, globular and remains conformationally unchanged even in the presence of lysate. Reproducing equivalently good models when using real CFPS reactions is certainly possible, as long as the *de novo* protein expressed fits these ideal criteria (as is usually required for high quality SANS data to be acquired as a general rule). In the case of non-ideal protein, *ab initio* modelling is affected. The data recorded from BSA, which occurs as monomers and dimers of which the equilibrium is influenced by the presence of lysate, produces an *ab initio* model that is an average of these states (Figures 6.15 and 6.16). Similarly, the non-ideal model protein Glob, which consists of a dynamic mix of globulins, results in an *ab initio* model which is an average of all the states of Glob (Figures 7.18 and 7.19). With decreasing concentrations of Glob, a shift towards monomeric IgG was favoured (Table 7.4).

In such circumstances, the use of SASREF (rigid body modelling) (Figures 7.17, 7.19) (as opposed to *ab initio* modelling alone) can permit the generation of models which fit the experimental SANS data. However, these rigid body models are based on static, closely-packed and ordered crystal structures and caution is required in extrapolating this information to the actual state of a protein in solution.

### 7.4.4. Perspectives and limits of using SANS for the direct structural analysis of unfractionated protein within CFPS reaction mixtures

The goal of the study was to establish some of the foundations relating to the possible *in situ* use of SANS for the characterisation of *de novo* expressed CFPS membrane proteins in proteoliposomes (Figure 7.1, 7.20). This feasibility study has shown that proteins can be characterised in the presence of a heterogeneous mixture of biomolecules. Such conditions could apply to CFPS produced proteins, particularly those that are liable to deterioration and precipitation when isolated and purified in a classical way. Additionally, the concept could be applied to the study of proteins under crowded condition that could be said to better relate to real *in vivo* cellular environment.

Ultimately, the goal is to establish a reliable protocol that enables the characterisation of CFPS membrane proteins embedded in the proteoliposome. As described in chapter three, one of the major problems with CFPS in its classical use is that contaminating lysate proteins adhere to the liposomes and the development of suitable protocols for completely removing these remains problematic. While work seeking to improve the purity of the CFPS protein in proteoliposomes remains important, the SANS technique described here may be of value in avoiding some of these problems. In the case of proteoliposomes, simultaneously matchout of both the lipid layer and the contaminating protein would be the objective. IIt has already been demonstrated that lipid molecules, used as “stealth liposomes”, can be accurately deuterated such as to be completely matched-out in 100 % $D_2O$ (Maric *et al.*, 2014); Maric et al 2015) and SANS exploitation of this approach has recently been demonstrated (Josts et al Structure 2018; Nitche et al 2018). In the same study, it was further shown that nanodiscs made using these matched-out lipids and matched-out belt protein, could also be rendered invisible during SANS experiments in 100 % $D_2O$. Despite being challenging, the work described in this chapter would suggest the feasibility of using the same concept to render invisible both lipid layer and contaminating proteins present in the CFPS proteoliposome samples (Figures 7.1 and 7.20).

As promising as the results and the scope of the SANS technique presented here are, there are nevertheless several limits of to the technique that need to be addressed. Firstly, the model proteins used here were studied at high concentrations that are not normally attained in a typical CFPS reaction. CFPS reactions at Synthelis can produce approximately 1 mg of *de novo* protein per mL of reaction, under ideal conditions with “well-behaving” proteins. As demonstrated in chapter three with CXCR4, SDF1-α, and 2dCD4, the amounts produced were less than 400 µg per mL of reaction. The signal

generated from such a low amount of proteins is therefore unlikely to be visible against the background arising from the lysate.

While it is difficult to produce soluble protein *in situ* using CFPS at the levels used here, one way to approach the problem would be to use the ability of CFPE to produce proteins in proteolipsomes and to partially purify these (e.g. using sucrose density gradient purification) – thereby concentrating the samples (above 10 mg/mL of CFPS membrane protein) to a level where a SANS signal could become viable (as shown in for the protein Glob and BSA in Figures 7.8, 7.9). This process simultaneously increases the concentration of the membrane protein of interest, and eliminates problems associated with the major bulk of contaminating lysate protein. Hence it is expected that the SANS signal arising from the membrane protein could be considerably better than those presented here since a major fraction of lysate will have been removed. To achieve the characterisation of membrane proteins in proteoliposomes it would be necessary to develop the protocols required for the use of matched-out liposomes during CFPS to allow membrane protein insertion. This would require the simultaneous matching of both contaminating proteins and liposome (Figures 7.1 and 7.20).

The second, but severe, limitation is the ability to generate hydrogenated protein using deuterated lysate. As shown in chapter two, the efficiency of CFPS expression using deuterated lysate was severely affected. Therefore, optimising the deuterated lysate CFPS reaction in order to restore expression will be a major task. Supplementing the lysate with exogenous polymerases could help to restore CFPS expression levels (Etezady-Esfarjani *et al.*, 2007a). For now, the SANS technique is restricted realistically to only the use of hydrogenated lysate, in which deuterated proteins can be expressed using deuterated amino acids that are supplied to the reaction mixture.

Thirdly, whilst it is possible to adequately match out the lysate, it has been shown here that the accuracy of size parameters derived from 1D neutron scattering data, and the quality of *ab initio* models generated, is impacted by the quality of protein sample. Ideally, for modelling purposes, as demonstrated with MBP, CFPS-generated proteins suitable for SANS would be highly homogenous and compact in order to give accurate models (as generally required for SAS). However, as demonstrated in chapters two and three, CFPS reactions generated a heterogeneous mix of truncated and oligomeric species of the *de novo* CXCR4. Depending on the amount of these present, these may compromise the interpretability of SANS results. This was indicated here with Glob samples, a heterogenous mix of species, which resulted inconsistencies in the recorded size parameters and reduced the quality of the models (Figure 7.18, Table 7.6).

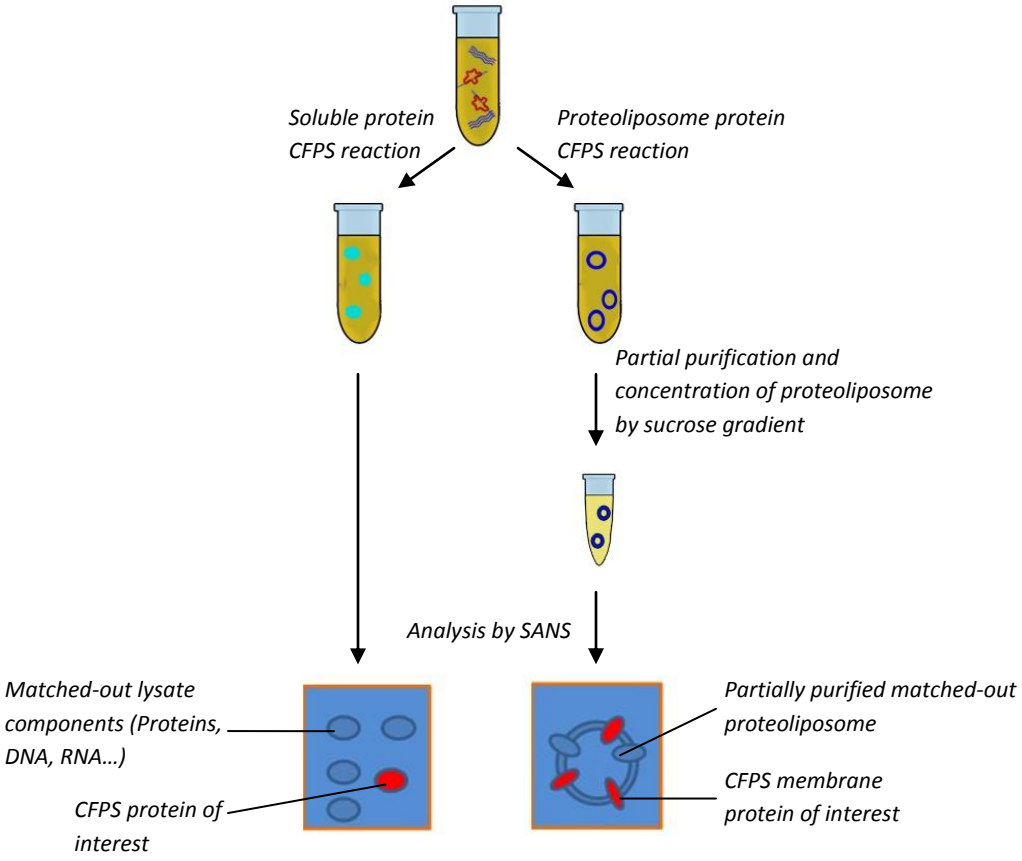



***Figure 7.20:*** *Diagram illustrating the applicability of SANS for the analysis of CFPS protein of interest in the presence of contaminating lysate proteins. Soluble CFPS proteins are found in a large amount of lysate, which can add considerable background to the recorded SANS signal. In the case of CFPS membrane proteins in proteoliposomes, of which characterisation is the ultimate goal, they can be partially purified and concentrated to enable SANS signals to be collected with substantially less noise.*

## 7.5. CONCLUSIONS

The feasibility of using SANS to matchout the cellular lysate and to obtain an interpretable SANS signal from an immersed target protein has been demonstrated. The results have the potential to open the doors to a number of highly interesting perspectives in the field of SANS. The most attractive of these is the prospective of being able to use the *in situ* structural characterisation of unpurified proteins (not necessarily CFPS-produced) in a heterogeneous mix of biomolecules. In addition to the simplicity of an application whereby complex purification steps can be avoided, this feature could be particularly useful for the characterisation of unstable proteins that deteriorate quickly following isolation and purification. Furthermore, the technique is intrinsically a means of studying proteins under crowded conditions, which most closely resemble the native conditions in cells *in vivo*.

The work described in this chapter here sets the conditions for future developments with the ultimate goal of characterising CFPS membrane protein embedded in proteoliposomes. With CFPS membrane proteins in proteoliposomes being the core selling-point of the Synthelis, the extended application of the SANS technique presented would be a valuable and elegant technique that works around the challenge to structural characterisation given by contaminating proteins. Furthermore, the technique can also benefit the structural characterisation of membrane proteins in their natural lipidic microenvironment, which remains challenging using classical methods.

# CHAPTER EIGHT

## DISCUSSION, CONCLUSIONS AND PESPECTIVES

# DISCUSSION, CONCLUSIONS AND PERSPECTIVES

The basis of the project was to demonstrate that proteins produced using the Synthelis CFPS system are equivalent to *in vivo* methods. This was tested using biophysical, functional and, finally, advanced structural tools. This goal is important for Synthelis since it reinforces the value and customer confidence in the CFPS system, and encourages the exploitation of CFPS technology for integrated structural biology more generally which is today limited in scope. There were several challenges to overcome for this. Firstly, CFPS is typically only exploited for the production of analytical quantities of protein and is not, so far, a method used to produce routinely the quantities and quality of protein required for integrated structural biology studies, whether the protein targets are soluble or membrane proteins. There was therefore a need to optimise the CFPS methods and procedures for scaled up production of proteins in suitable quantity and quality. This required analytical approaches to the production and testing of different lysates used in the CFPS reactions and the use of a series of model proteins having different properties (e.g. size, solubility etc). Secondly, the CFPS system used here had to be adapted and tested for the production of labelled proteins. This is important as neutron scattering is becoming increasingly exploited as a technique valuable within the toolbox of integrated structural biology, particularly for studying protein complexes.

The series of model protein systems selected for the project consisted of CXCR4, SDF1-α and CD4 (plus its two domain construct, 2dCD4). This series covers membrane-bound proteins (CXCR4, which is a GPCR, and CD4, produced using CFPS in the form of proteoliposomes) and soluble proteins (SDF1-α, a binding partner to CXCR4, and the two-domain construct of CD4, 2dCD4). Each of these had its own challenge in terms of protein production:

- CXCR4 and CD4: As membrane proteins, to fold into their native conformations, CXCR4 and CD4 need to be inserted into the lipid bilayer of the cell membrane. The yields of both of these recombinant proteins in mammalian cells are typically too low for structural characterisation. Recombinant expression in *E. coli* cells on the other hand does not permit the insertion of the mammalian membrane proteins into the cell membrane, thereby resulting in aggregated CXCR4/CD4. The resulting recovery of CXCR4 and CD4 requires highly specific and intensive refolding procedure.

- SDF1-α and 2dCD4: These proteins can be produced in soluble form when using mammalian host cells. However, the use of the mammalian expression system for structural purposes can be expensive and not compatible with isotopic protein labelling (e.g. deuteration). The expression in *E. coli* cells however produces SDF1-α and 2dCD4 in insoluble and aggregated forms. The inability to form correctly folded proteins in *E. coli* can be attributed to the reducing microenvironment in the cells, which limits disulphide bond formation. Both SDF1-α and 2dCD4 have two pairs of disulphide bonds which need to be formed in order to promote correct folding. To recover soluble and functional proteins, *E. coli* expressed SDF1-α and 2dCD4 need to undergo highly adapted and intensive refolding procedure. However, the yields of soluble and functional proteins at the end of the refolding process can be very low.

With several previous reports demonstrating improved CFPS protein production with the use of alternative *E. coli* strains (Krinsky *et al.*, 2016, Kwon and Jewett, 2015, Kim *et al.*, 2006a, Kigawa *et al.*, 2004, Soranzo *et al.*, 2016, Liguori *et al.*, 2008), a number of lysates derived from various *E. coli* strains (Nico21, C43, Rosetta, Rosetta-Gami and SHuffle) were produced, tested and compared for CFPS expression, as described in chapter two, with the goal of improving protein yields and quality by using BAK and VDAC as CFPS protein standards. Lysate production firstly involved the optimised culture of the *E. coli* cells as the source of the lysate. Synthelis has long established the growth profile of its reference *E. coli* strain which maximises the efficacy of the ensuing CFPS lysate. With the introduction of new *E. coli* strains for the production of lysates the growth profiles of the new strains were studied to allow cultures to be harvested reliably for maximally active lysates. These will serve as references for future cultures, providing the quality control required for an industrial process. Particularly in the context of quality control in industry, the standardisation, reproducibility, and traceability of the lysate batches exploited in CFPS reactions are crucial.

Using the references of BAK and VDAC as standards for CFPS membrane protein expression in proteoliposomes, the lysates produced from the various strains were tested. It was found that lysates derived from *E. coli* Rosetta and SHuffle strains improved the yields of protein expression. This demonstrates the future perspectives of exploiting the wide selection of commercially available *E. coli* strains, each of which having been designed specifically for improved protein expression *in vivo,* which could potentially transfer their improved properties to CFPS (http://wolfson.huji.ac.il/expression/bac-strains-prot-exp.html). Additionally, embracing the lack of barriers of the CFPS system, the idea of

mixing the lysates of derived from various *E. coli* strains, or perhaps even eukaryote-derived lysates, in order to benefit from their combined properties, is a prospect to be investigated.

As the volume of *E. coli* culture in flasks is increased, the task, which is already highly labour intensive, becomes problematic. Therefore, following the culture of *E. coli* in flasks using the well-established protocols modified with the updated growth curves, *E. coli* cultures were carried out in high-cell density cultures (HCDC) in fermenters. This procedure was important in order to achieve large scale CFPS, which is required for meeting the demands of proteins required for integrated structural biology. Upon the establishment of the feasibility of using HCDC, cultures in HCDC were also adapted for growth in deuterated media, which were used subsequently for the generation of deuterated lysate required for the purposes of SANS studies of unfractionated protein (see below). The use of HCDC, which enables automated control on the growth of *E.* coli, meets the needs of Synthelis for the routine scaling-up and automation of production. Future work will aim to increase further the volumes of *E. coli* culture in HCDC.

In chapter three, the optimisation of the expression and purification of CXCR4 and CD4, for the purposes of structural characterisation, was investigated. Three CXCR4 constructs were utilised and proteins expressed either embedded in liposomes or in solubilised form using detergent. In both cases CFPS was successful for the expression of wild-type CXCR4 or CXCR4-lysozyme mutant. Additional expression of C-terminal strep-tagged CXCR4 confirmed that CFPS enabled full-length translation of CXCR4. Working towards boosting the expression levels of CXCR4, it was determined that the lysate derived from Rosetta *E. coli* strain enhanced its expression (100 – 150 µg/mL of CFPS reaction) compared to the reference BL21 strain (50 – 70 ug/mL of CFPS reaction). Despite this step forwards and a number of optimisation experiments, the yields and the purity of recovered CXCR4 remained too low for structural characterisation. As for CD4 expression in proteoliposome, the yields even lower (20 µg/mL of CFPS reaction). In the case of liposome/proteoliposome expression, the non-specific adherence of lysate proteins to the proteoliposome was ongoing challenge which hindered the purity of CXCR4/CD4 samples. In the case of CFPS expression in the presence of detergents it was possible to obtain solubilised and purified form of CXCR4. However, the expression and purification of the protein was low and insufficient for structural studies. In the context of purification and as a means to avoid the problem of contaminating lysate proteins adhering to the proteoliposomes, the production of pure CFPS-produced membrane protein in proteoliposomes can be achieved by re-insertion of the purified detergent solubilised membrane proteins into empty liposomes (Klammt *et al.*, 2004b, Costello *et al.*,

1984). While this can be achieved, the issue of low levels of expression of CXCR4/CD4 can still be of hindrance. Nevertheless, the inability to purify CXCR4 in large quantities ultimately propelled the development of a novel method to allow structural studies despite the presence of contaminating proteins (see later).

It is important to note that CFPS of BAK and VDAC resulted in the majority of full-length species being expressed. In the case of CXCR4/CD4 expression specifically however, relatively low levels of its full-length form was produced and a large amount of truncated species occurred. The formation of truncated species is wasteful process since it deprives the the CFPS reaction of valuable resources which would otherwise be used for full-length protein expression. With the sequence of the encoding DNA being crucial in the determination of complete, or incomplete, protein expression in CFPS, a systematic study in which the coding sequences of CXCR4/CD4 are modified to avoid codons and secondary mRNA structures which interrupt protein translation could be considered for future work.

The optimisation of CFPS expression and purification of SDF1-α and 2dCD4 is also described in chapter three. As opposed to the reference BL21 lysate, the use of SHuffle lysate for the expression of SDF1-α and 2dCD4 was found to be critical in enabling the formation of disulphide bridges, and thus protein solubility and correct protein folding. In contrast, when these are expressed in *E. coli* cells, they occur as non-functional and insoluble proteins (Picciocchi *et al.*, 2014, Murphy *et al.*, 2007), which require refolding. Refolding is however a laborious and potentially wasteful step, which requires high degree of specific optimisation in order to give high yields of correctly folded, soluble and functional protein. In this case, the optimised CFPS system can be used advantageously for the almost immediate production of soluble and functional proteins, which can be subsequently purified using standard methods.

Additionally, CFPS allowed the production of deuterated SDF1-α and 2dCD4 with relative ease and speed. CFPS was found here to be particularly advantageous over classical *E. coli* expression for the production of deuterated proteins. The speed and ease of production of either unlabelled or labelled proteins are critical aspect in the context of both industry and academia where tight schedules and deadlines need to be met. However, it was noted CFPS did not allow the expression of fully perdeuterated protein. A maximum 77 % deuteration level for non-exchangeable hydrogen was obtained for deuterated SDF1-α expressed using CFPS. With CFPS being carried out in $H_2O$-based conditions, SDF1-α undergoes significant back-protonation of $^{2}H$ to $^{1}H$, resulting in the relatively lower deuteration levels (Etezady-Esfarjani *et al.*, 2007b, Tonelli *et al.*, 2011). Despite the high cost of the deuterated amino acids, CFPS can therefore be an effective and alternative system for the expression of

deuterated proteins. In particular, the ease of expression of deuterated 2dCD4 is an important technical development since it is a critical component of study in the field of HIV research. As such, the work presented here serves to demonstrate the application of CFPS as an accessible tool for routine expression of not only membrane proteins, but proteins, labelled or unlabelled, in general.

Following the initial small-scale optimisation, for the purposes of structural biology and therefore the requirement of large amounts of proteins, CFPS was used for the large-scale production of both unlabelled and labelled SDF1-α and 2dCD4 (100 mL batches). Throughout the optimisation studies and scaling up, standard operating protocols (SOPS) were developed, which in the context of industry, can serve to standardise protocols and ensure reproducibility. This is critical for Synthelis where developments towards reliable and routine scaling-up of production are required.

In chapter four, CFPS-produced proteins were biophysically and functionally characterised using an array of methods. Functional assessments using ELISAs and SPR provided indications on the structural integrity of CFPS-generated full-length CD4- and CXCR4-proteoliposomes. Mass-spectrometry of SDF1-α served to establish that the CFPS allowed for formation of fully translated proteins with disulphide bonds. Additionally, mass-spectrometry allowed the confirmation of the deuteration levels of SDF1-α. Chemotaxis assays of hydrogenated and deuterated SDF1-α served to compare and confirm the functionality of these proteins. Particularly, both hydrogenated and deuterated forms of SDF1-α were found to be equivalent in functionality.

As well as providing important scientific insights to their biophysical and functional state, these techniques can be proposed as routine methods for the quality assurance of CFPS-produced proteins in the industrial context. ELISA in particular, is a relatively cheap method, but a very versatile tool which could be implemented in the context of small and medium companies. In line with quality control concerns, the application of these methods would enable the identification of functional deviations in CFPS expressed proteins. In this context such functional characterisation would provide an added value and reinforce customer confidence in CFPS products.

Also described in chapter four, methods for characterising CFPS membrane proteins in proteoliposomes despite the presence of contaminating lysate proteins, were assessed. The NanoSight$^{TM}$ particle tracker was found to be a relatively straightforward visual tool which allowed the determination of the size distribution and homogeneity of liposomes/CXCR4-proteoliposomes before, during and after expression/purification. These results can be used for assessing, improving and standardising the quality

of CFPS reactions, and the subsequent purification and resuspension buffers. In the industrial context, the device can be a useful bench-side tool which could be introduced as part of the battery of quality control assays. In this context, the NanoSight™ can assist in ensuring the state of proteoliposome and identifying compromised batches where proteoliposome are either aggregated or destroyed.

Additionally, the study carried out here with the NanoSight™ particle tracker serves as a reference for future work which will involve the use of the NanoSight™ particle tracker coupled with multiple wavelength lasers and fluorescence detector system. By placing specific fluorescent tags or antibodies on the CXCR4, the upgraded device could be used for the determination of the actual number of correctly inserted and orientated CXCR4 molecules present in the proteoliposomes. Achieving this goal, would answer the ever so critical concern of customers of the amount of functional CXCR4 molecules embedded in the proteoliposomes. Further pushing the limits of the NanoSight™ particle tracker, the upgraded device could subsequently be used for simultaneously detecting the binding of SDF1-α to CXCR4 after each of these have been specifically labelled with a fluorescent tag or antibody. The ability to visually observe the interaction between SDF1-α to CXCR4 will further confirm the functionality of both of these CFPS-produced proteins.

In further work described in chapter four, the proteoliposomes of CXCR4 and CD4 were assessed by TEM. Despite the destructive nature of the experiment for proteoliposomes, TEM serves as a stepping stone towards cryo-EM. Cryo-EM as a tool for structural characterisation benefits from the advantage in its ability to study samples with minimal specific preparation and in their native state. Cryo-EM, having been used for the study of isolated membrane proteins, membrane proteins embedded in the membrane and the dynamics of the membrane (Goldie *et al.*, 2014, Abe and Fujiyoshi, 2016, Mitsuoka and Gerle, 2016, Rawson *et al.*, 2016), makes a very exciting prospect for the study of CFPS membrane proteins embedded in proteoliposomes.

In chapter five, CFPS-generated SDF1-α was successfully crystallised and subject to X-ray diffraction with the goal of demonstrating that the CFPS product has the same molecular structure as the classically produced protein. The resulting high-resolution crystal structure of CFPS-generated SDF1-α, showed strong similarity to previous published structures and demonstrates a high degree of structural integrity. This result is important for Synthelis and strongly suggests that proteins generated by CFPS are structurally equivalent to classical *in vivo* methods.

As described in chapter six, with the goal of further demonstrating the applicability of CFPS for the production of labelled protein for structural studies, CFPS-generated deuterated 2dCD4, complexed to viral GP120 was analysed using small-angle neutron scattering (SANS). However, the low SANS signals were a main limitation in the interpretation of the data and models. The experiment however showed that CFPS-produced deuterated 2dCD4 could be effectively matched-out. As a preliminary study, the experiment serves as the basis for setting out the conditions required to enable structural interpretation. Namely, the use of monomeric GP120 rather than dimeric GP120 can help to improve the signal of 2dCD4. Unlike with dimeric GP120 where the sample is likely to consist of a heterogeneous mix of unbound, singly-bound and doubly-bound 2dCD4 to GP120, the use of monomeric GP120 will enable the production of a highly homogenous solution consisting of monomeric GP120 bound to a single 2dCD4 molecule. Nevertheless, as a continuation of previous experiments carried out on deuterated 2dCD4 and SANS (Channell, 2018), the work here provides a proof-of-concept for the CFPS-production of deuterated protein for structural studies. The SANS study was essential in understanding the challenges that would limit structural interpretation. Namely, the flexible loops and glycans of GP120 (37 % by weight of GP120), will pose a challenge in accurately defining the models. Techniques which allow for selective labelling of either the loops of GP120 or the glycan could be considered in order to pin-point these respective residues and moieties during SANS.

Finally, in chapter seven, a feasibility study was made for the development of an alternative and novel SANS method for structurally characterising CFPS produced proteins despite the challenge of significant contamination arising from lysate proteins Using hydrogenated MBP, BSA and globulins as model proteins, it was found that even at high deuterated lysate concentrations, the lysate could effectively be matched-out against the background. Similarly, hydrogenated lysate could be effectively matched-out when studying deuterated MBP. Initially, it was suggested that the very high levels of hydrogen incoherent scattering from the hydrogenated lysate would conceal the signal from the deuterated protein. Hence, the second strategy of using a hydrogenated protein within deuterated lysate was implemented, where it was expected that the levels of hydrogen incoherent scattering against which the coherent SANS signal of the target protein, would be much lower. In either case however, it was shown that structural information of the model proteins of concern was still conserved and could be interpreted to enable the determination of structural parameters and the generation of low-resolution models.

However, challenges remain to develop the technique. Firstly, the typical CFPS reaction does not generate as high amounts of proteins as were considered here. However, with the use of proteoliposomes which have undergone partial purification, significantly higher membrane protein concentrations can be achieved whilst simultaneously reducing the overall amount of contaminating lysate. To be able to study proteoliposomes however, it is necessary to produce liposomes with matched-out lipids, such that both the lysate and liposome are simultaneously matched-out during SANS, leaving a strong membrane protein signal. The second challenge that needs to be addressed is the inability to generate hydrogenated CFPS proteins when using deuterated lysate (chapter two). For now, it is only possible to produce deuterated CFPS proteins while using the hydrogenated lysate. Future work to identify the conditions in which the deuterated lysate can be made active for the production of hydrogenated proteins would involve supplementing the CFPS reaction with the functional components (e.g. polymerase). Thirdly, unpurified proteins produced using CFPS (and indeed any expression technique) may occur as a heterogeneous mix, comprising monomers, oligomers and truncated species resulting in a non-ideal sample where the interpretability of resulting SANS data is challenging.

Nevertheless, as a stand-alone experiment whereby it is possible to obtain structural insights into proteins despite in the presence of an over-abundance of heterogeneous biomolecules, this feasibility study can find further application in the field of SANS and structural biology. Namely, the study of proteins under crowded conditions, which are close to the native state of proteins in cells, is therefore a very interesting prospect. Additionally, the technique can allow the structural analysis of proteins which are particularly unstable and prone to deterioration upon purification. Finally, these studies have evoked the intriguing prospect of being able to use CFPS to observe and study *in situ* the translation of proteins.

As a CIFRE funded project, this project was set in the context of collaboration between industry and academia. Synthelis (France), which originated as a spin-off of the University of Grenoble Alps (UGA, France), is a small biotech company that specialises in CFPS expression of membrane and other difficult-to-express proteins using patented technology. The main aim of the project was to demonstrate the structural integrity of CFPS products, as being equivalent to *in vivo* methods, using advanced structural tools. For Synthelis, the purpose of the task was to reinforce the customer confidence, and add value to CFPS products. To achieve this goal, extensive collaboration, comprising the use of resources and facilities with the academic partners the Institute Laue-Langevin (ILL, France) and the European Synchrotron Radiation Facility (ESRF, France) was established. Additionally, international collaboration with the HIV Pathogenesis Research Laboratory (HIVRPL) from the University of Witwatersrand (South

Africa) was made. Over the course of the three years of collaboration, lessons learnt with the development of experimental methods, and relationships built between Synthelis and the academic partners, have led to the foundation of industrial-academic partnerships. Synthelis and ESRF are building, as a customer service, a CFPS protein production and X-ray crystallography gene-to-structure pipeline.

This project has taken the first small but significant steps towards the generalised use of CFPS methods as part of the armoury of protein production techniques available to integrated structural biologists. A number of key advantages have been identified including the prospect of protein structure determination within complex and relevant biologically relevant environments, and where isotope labelling used in conjunction with neutron scattering approaches can provide completely novel approaches to structure determination. Finally, this project will serve as the basis for ultimately, the production of membrane protein by CFPS for structural characterisation.

## REFERENCES


ABE, K. & FUJIYOSHI, Y. 2016. Cryo-electron microscopy for structure analyses of membrane proteins in the lipid bilayer. *Current opinion in structural biology,* 39**,** 71-78.

ACHARYA, P., TOLBERT, W. D., GOHAIN, N., WU, X., YU, L., LIU, T., HUANG, W., HUANG, C. C., KWON, Y. D., LOUDER, R. K., LUONGO, T. S., MCLELLAN, J. S., PANCERA, M., YANG, Y., ZHANG, B., FLINKO, R., FOULKE, J. S., JR., SAJADI, M. M., KAMIN-LEWIS, R., ROBINSON, J. E., MARTIN, L., KWONG, P. D., GUAN, Y., DEVICO, A. L., LEWIS, G. K. & PAZGIER, M. 2014. Structural definition of an antibody-dependent cellular cytotoxicity response implicated in reduced risk for HIV-1 infection. *Journal of virology,* 88**,** 12895-906.

AKAMA, S., YAMAMURA, M. & KIGAWA, T. 2012. A multiphysics model of in vitro transcription coupling enzymatic reaction and precipitation formation. *Biophysical journal,* 102**,** 221-30.

AKASHI, K., MIYATA, H., ITOH, H. & KINOSITA, K., JR. 1998. Formation of giant liposomes promoted by divalent cations: critical role of electrostatic repulsion. *Biophysical journal,* 74**,** 2973-82.

ALMARAZ, R. T. & LI, Y. 2017. Labeling glycans on living cells by a chemoenzymatic glycoengineering approach. *Biology open,* 6**,** 923-927.

ANASTASINA, M., TERENIN, I., BUTCHER, S. J. & KAINOV, D. E. 2014. A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *BioTechniques,* 56**,** 36-9.

ANDERSON, M. J., STARK, J. C., HODGMAN, C. E. & JEWETT, M. C. 2015. Energizing eukaryotic cell-free protein synthesis with glucose metabolism. *FEBS letters,* 589**,** 1723-1727.

APPOLAIRE, A., GIRARD, E., COLOMBO, M., DURA, M. A., MOULIN, M., HARTLEIN, M., FRANZETTI, B. & GABEL, F. 2014. Small-angle neutron scattering reveals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta crystallographica. Section D, Biological crystallography,* 70**,** 2983-93.

ARECHAGA, I., MIROUX, B., KARRASCH, S., HUIJBREGTS, R., DE KRUIJFF, B., RUNSWICK, M. J. & WALKER, J. E. 2000. Characterisation of new intracellular membranes in Escherichia coli accompanying large scale over-production of the b subunit of F(1)F(o) ATP synthase. *FEBS letters,* 482**,** 215-9.

ARUMUGAM, T. U., ITO, D., TAKASHIMA, E., TACHIBANA, M., ISHINO, T., TORII, M. & TSUBOI, T. 2014. Application of wheat germ cell-free protein expression system for novel malaria vaccine candidate discovery. *Expert review of vaccines,* 13**,** 75-85.

ASHISH, GARG, R., ANGUITA, J. & KRUEGER, J. K. 2006. Binding of full-length HIV-1 gp120 to CD4 induces structural reorientation around the gp120 core. *Biophysical journal,* 91**,** L69-71.

BANEYX, F. & MUJACIC, M. 2004. Recombinant protein folding and misfolding in Escherichia coli. *Nature biotechnology,* 22**,** 1399-408.

BANK, A. & MARKS, P. A. 1966. Protein synthesis in a cell free human reticulocyte system: ribosome function in thalassemia. *The Journal of clinical investigation,* 45**,** 330-6.

BARIBAUD, F., EDWARDS, T. G., SHARRON, M., BRELOT, A., HEVEKER, N., PRICE, K., MORTARI, F., ALIZON, M., TSANG, M. & DOMS, R. W. 2001. Antigenically distinct conformations of CXCR4. *Journal of virology,* 75**,** 8957-67.

BAZZACCO, P., BILLON-DENIS, E., SHARMA, K. S., CATOIRE, L. J., MARY, S., LE BON, C., POINT, E., BANERES, J. L., DURAND, G., ZITO, F., PUCCI, B. & POPOT, J. L. 2012. Nonionic homopolymeric amphipols: application to membrane protein folding, cell-free synthesis, and solution nuclear magnetic resonance. *Biochemistry,* 51**,** 1416-30.

BJARNADOTTIR, T. K., GLORIAM, D. E., HELLSTRAND, S. H., KRISTIANSSON, H., FREDRIKSSON, R. & SCHIOTH, H. B. 2006. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics,* 88**,** 263-73.

BLACHLY-DYSON, E. & FORTE, M. 2001. VDAC channels. *IUBMB life,* 52**,** 113-8.

BLEUL, C. C., FUHLBRIGGE, R. C., CASASNOVAS, J. M., AIUTI, A. & SPRINGER, T. A. 1996. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *The Journal of experimental medicine,* 184**,** 1101-9.

BONINCONTRO, A., BRIGANTI, G., GIANSANTI, A., PEDONE, F. & RISULEO, G. 1993. Effects of magnesium ions on ribosomes: a fluorescence study. *Biochimica et biophysica acta,* 1174**,** 27-30.

BONSIGNORI, M., HWANG, K. K., CHEN, X., TSAO, C. Y., MORRIS, L., GRAY, E., MARSHALL, D. J., CRUMP, J. A., KAPIGA, S. H., SAM, N. E., SINANGIL, F., PANCERA, M., YONGPING, Y., ZHANG, B., ZHU, J., KWONG, P. D., O'DELL, S., MASCOLA, J. R., WU, L., NABEL, G. J., PHOGAT, S., SEAMAN, M. S., WHITESIDES, J. F., MOODY, M. A., KELSOE, G., YANG, X., SODROSKI, J., SHAW, G. M., MONTEFIORI, D. C., KEPLER, T. B., TOMARAS, G. D., ALAM, S. M., LIAO, H. X. & HAYNES, B. F. 2011. Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *Journal of virology,* 85**,** 9998-10009.

BOOS, W. & SHUMAN, H. 1998. Maltose/maltodextrin system of Escherichia coli: transport, metabolism, and regulation. *Microbiology and molecular biology reviews : MMBR,* 62**,** 204-29.

BORKOWSKI, O., BRICIO, C., MURGIANO, M., ROTHSCHILD-MANCINELLI, B., STAN, G. B. & ELLIS, T. 2018. Cell-free prediction of protein expression costs for growing cells. *Nature communications,* 9**,** 1457.

BOWLER, M. W., NURIZZO, D., BARRETT, R., BETEVA, A., BODIN, M., CASEROTTO, H., DELAGENIERE, S., DOBIAS, F., FLOT, D., GIRAUD, T., GUICHARD, N., GUIJARRO, M., LENTINI, M., LEONARD, G. A., MCSWEENEY, S., OSKARSSON, M., SCHMIDT, W., SNIGIREV, A., VON STETTEN, D., SURR, J., SVENSSON, O., THEVENEAU, P. & MUELLER-DIECKMANN, C. 2015. MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of synchrotron radiation,* 22**,** 1540-7.

BRADY, R. L., DODSON, E. J., DODSON, G. G., LANGE, G., DAVIS, S. J., WILLIAMS, A. F. & BARCLAY, A. N. 1993. Crystal structure of domains 3 and 4 of rat CD4: relation to the NH2-terminal domains. *Science,* 260**,** 979-83.

BRAUN, M. K., GRIMALDO, M., ROOSEN-RUNGE, F., HOFFMANN, I., CZAKKEL, O., SZTUCKI, M., ZHANG, F., SCHREIBER, F. & SEYDEL, T. 2017. Crowding-Controlled Cluster Size in Concentrated Aqueous Protein Solutions: Structure, Self- and Collective Diffusion. *The journal of physical chemistry letters,* 8**,** 2590-2596.

BRAUN, P. & LABAER, J. 2004. High throughput protein production for functional proteomics. *Drug discovery today,* 9**,** S1-7.

BROCKHAUSER, S., SVENSSON, O., BOWLER, M. W., NANAO, M., GORDON, E., LEAL, R. M., POPOV, A., GERRING, M., MCCARTHY, A. A. & GOTZ, A. 2012. The use of workflows in the design and implementation of complex experiments in macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography,* 68**,** 975-84.

BRODEL, A. K., SONNABEND, A. & KUBICK, S. 2014. Cell-free protein expression based on extracts from CHO cells. *Biotechnology and bioengineering,* 111**,** 25-36.

BRODEL, A. K., WUSTENHAGEN, D. A. & KUBICK, S. 2015. Cell-free protein synthesis systems derived from cultured mammalian cells. *Methods in molecular biology,* 1261**,** 129-40.

BROMLEY, S. K., BURACK, W. R., JOHNSON, K. G., SOMERSALO, K., SIMS, T. N., SUMEN, C., DAVIS, M. M., SHAW, A. S., ALLEN, P. M. & DUSTIN, M. L. 2001. The immunological synapse. *Annual review of immunology,* 19**,** 375-96.

BURKLY, L. C., OLSON, D., SHAPIRO, R., WINKLER, G., ROSA, J. J., THOMAS, D. W., WILLIAMS, C. & CHISHOLM, P. 1992. Inhibition of HIV infection by a novel CD4 domain 2-specific monoclonal antibody. Dissecting the basis for its inhibitory effect on HIV-induced cell fusion. *Journal of immunology,* 149**,** 1779-87.

CALHOUN, K. A. & SWARTZ, J. R. 2005. Energizing cell-free protein synthesis with glucose metabolism. *Biotechnology and bioengineering,* 90**,** 606-13.
CALLOW, P., SUKHODUB, A., TAYLOR, J. E. & KNEALE, G. G. 2007. Shape and subunit organisation of the DNA methyltransferase M.AhdI by small-angle neutron scattering. *Journal of molecular biology,* 369**,** 177-85.
CARLSON, E. D., GAN, R., HODGMAN, C. E. & JEWETT, M. C. 2012a. Cell-free protein synthesis: applications come of age. *Biotechnology advances,* 30**,** 1185-94.
CARLSON, E. D., GAN, R., HODGMAN, C. E. & JEWETT, M. C. 2012b. Cell-free protein synthesis: applications come of age. *Biotechnol Adv,* 30**,** 1185-94.
CARSON, J. L. 2014. Fundamental technical elements of freeze-fracture/freeze-etch in biological electron microscopy. *Journal of visualized experiments : JoVE***,** 51694.
CASCHERA, F. & NOIREAUX, V. 2014. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie,* 99**,** 162-8.
CASCHERA, F. & NOIREAUX, V. 2015. A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metabolic engineering,* 27**,** 29-37.
CENATIEMPO, Y. 1986. Prokaryotic gene expression in vitro: transcription-translation coupled systems. *Biochimie,* 68**,** 505-15.
CERUTTI, N., KILLICK, M., JUGNARAIN, V., PAPATHANASOPOULOS, M. & CAPOVILLA, A. 2014a. Disulfide reduction in CD4 domain 1 or 2 is essential for interaction with HIV glycoprotein 120 (gp120), which impairs thioredoxin-driven CD4 dimerization. *The Journal of biological chemistry,* 289**,** 10455-65.
CERUTTI, N., KILLICK, M., JUGNARAIN, V., PAPATHANASOPOULOS, M. & CAPOVILLA, A. 2014b. Disulfide reduction in CD4 Domain 1 or 2 is essential for interaction with HIV gp120, which impairs Thioredoxin-driven CD4 dimerization. *The Journal of biological chemistry*.
CERUTTI, N., MENDELOW, B. V., NAPIER, G. B., PAPATHANASOPOULOS, M. A., KILLICK, M., KHATI, M., STEVENS, W. & CAPOVILLA, A. 2010. Stabilization of HIV-1 gp120-CD4 receptor complex through targeted interchain disulfide exchange. *The Journal of biological chemistry,* 285**,** 25743-52.
CHADLI, M., REBAUD, S., MANITI, O., TILLIER, B., CORTES, S. & GIRARD-EGROT, A. 2017. New Tethered Phospholipid Bilayers Integrating Functional G-Protein-Coupled Receptor Membrane Proteins. *Langmuir : the ACS journal of surfaces and colloids,* 33**,** 10385-10401.
CHANNELL, J. A. 2018. *Biophysical studies of dynamic CD4 changes implicated in HIV-1 infection.* Doctor of Philosophy, Keele University.
CHEN, B., VOGAN, E. M., GONG, H., SKEHEL, J. J., WILEY, D. C. & HARRISON, S. C. 2005. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature,* 433**,** 834-41.
CHEN, H. Z. & ZUBAY, G. 1983. Prokaryotic coupled transcription-translation. *Methods in enzymology,* 101**,** 674-90.
CHEN, L., KWON, Y. D., ZHOU, T., WU, X., O'DELL, S., CAVACINI, L., HESSELL, A. J., PANCERA, M., TANG, M., XU, L., YANG, Z. Y., ZHANG, M. Y., ARTHOS, J., BURTON, D. R., DIMITROV, D. S., NABEL, G. J., POSNER, M. R., SODROSKI, J., WYATT, R., MASCOLA, J. R. & KWONG, P. D. 2009. Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science,* 326**,** 1123-7.
CHENG, E. H., SHEIKO, T. V., FISHER, J. K., CRAIGEN, W. J. & KORSMEYER, S. J. 2003. VDAC2 inhibits BAK activation and mitochondrial apoptosis. *Science,* 301**,** 513-7.
CHEREZOV, V., ROSENBAUM, D. M., HANSON, M. A., RASMUSSEN, S. G., THIAN, F. S., KOBILKA, T. S., CHOI, H. J., KUHN, P., WEIS, W. I., KOBILKA, B. K. & STEVENS, R. C. 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science,* 318**,** 1258-65.
CHESTERS, J. K. 1968. Cell-free protein synthesis by rumen protozoa. *The Journal of protozoology,* 15**,** 509-12.

CHI, H., WANG, X., LI, J., REN, H. & HUANG, F. 2015. Folding of newly translated membrane protein CCR5 is assisted by the chaperonin GroEL-GroES. *Scientific reports,* 5**,** 17037.
CHI, H., WANG, X., LI, J., REN, H. & HUANG, F. 2016. Chaperonin-enhanced Escherichia coli cell-free expression of functional CXCR4. *Journal of biotechnology,* 231**,** 193-200.
CIONI, P. & STRAMBINI, G. B. 2002. Effect of heavy water on protein flexibility. *Biophysical journal,* 82**,** 3246-53.
CLARK-LEWIS, I., KIM, K. S., RAJARATHNAM, K., GONG, J. H., DEWALD, B., MOSER, B., BAGGIOLINI, M. & SYKES, B. D. 1995. Structure-activity relationships of chemokines. *Journal of leukocyte biology,* 57**,** 703-11.
COMPTON, E. L., PAGE, K., FINDLAY, H. E., HAERTLEIN, M., MOULIN, M., ZACHARIAE, U., NORMAN, D. G., GABEL, F. & JAVELLE, A. 2014. Conserved structure and domain organization among bacterial Slc26 transporters. *The Biochemical journal,* 463**,** 297-307.
CONG, X., TOPIN, J. & GOLEBIOWSKI, J. 2017. Class A GPCRs: Structure, Function, Modeling and Structure-based Ligand Design. *Current pharmaceutical design,* 23**,** 4390-4409.
CONNELL, B. J., SADIR, R., BALEUX, F., LAGURI, C., KLEMAN, J. P., LUO, L., ARENZANA-SEISDEDOS, F. & LORTAT-JACOB, H. 2016. Heparan sulfate differentially controls CXCL12alpha- and CXCL12gamma-mediated cell migration through differential presentation to their receptor CXCR4. *Science signaling,* 9**,** ra107.
CONNOLLY, T. & GILMORE, R. 1986. Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein. *The Journal of cell biology,* 103**,** 2253-61.
COOKE, R. M., BROWN, A. J., MARSHALL, F. H. & MASON, J. S. 2015. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. *Drug discovery today,* 20**,** 1355-64.
CORIN, K., BAASKE, P., RAVEL, D. B., SONG, J., BROWN, E., WANG, X., GEISSLER, S., WIENKEN, C. J., JERABEK-WILLEMSEN, M., DUHR, S., BRAUN, D. & ZHANG, S. 2011. A robust and rapid method of producing soluble, stable, and functional G-protein coupled receptors. *PloS one,* 6**,** e23036.
CORTES, S., BARETTE, C., BEROUD, R., DE WAARD, M. & SCHAACK, B. 2018. Functional characterization of cell-free expressed Kv1.3 channel using a voltage-sensitive fluorescent dye. *Protein expression and purification,* 145**,** 94-99.
COSTELLO, M. J., VIITANEN, P., CARRASCO, N., FOSTER, D. L. & KABACK, H. R. 1984. Morphology of proteoliposomes reconstituted with purified lac carrier protein from Escherichia coli. *The Journal of biological chemistry,* 259**,** 15579-86.
COWAN, J. A. 2002. Structural and catalytic chemistry of magnesium-dependent enzymes. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine,* 15**,** 225-35.
CRAIG, D., HOWELL, M. T., GIBBS, C. L., HUNT, T. & JACKSON, R. J. 1992. Plasmid cDNA-directed protein synthesis in a coupled eukaryotic in vitro transcription-translation system. *Nucleic acids research,* 20**,** 4987-95.
CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561-3.
DA VELA, S., ROOSEN-RUNGE, F., SKODA, M. W. A., JACOBS, R. M. J., SEYDEL, T., FRIELINGHAUS, H., SZTUCKI, M., SCHWEINS, R., ZHANG, F. & SCHREIBER, F. 2017. Effective Interactions and Colloidal Stability of Bovine gamma-Globulin in Solution. *The journal of physical chemistry. B,* 121**,** 5759-5769.
DANPING ZHANG, P. W., LIMEI FAN, JIAZHANG LIAN, LEI HUANGA, JIN CAI, ZHINAN XUA, 2010. High-level soluble expression of hIGF-1 fusion protein in recombinant Escherichia coli. *Process Biochemistry,* 45**,** 1401–1405.
DEALWIS, C., FERNANDEZ, E. J., THOMPSON, D. A., SIMON, R. J., SIANI, M. A. & LOLIS, E. 1998. Crystal structure of chemically synthesized [N33A] stromal cell-derived factor 1alpha, a potent ligand

for the HIV-1 "fusin" coreceptor. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 6941-6.

DENIAUD, A., LIGUORI, L., BLESNEAC, I., LENORMAND, J. L. & PEBAY-PEYROULA, E. 2010. Crystallization of the membrane protein hVDAC1 produced in cell-free system. *Biochimica et biophysica acta,* 1798**,** 1540-6.

DENISOVA, G., RAVIV, D., MONDOR, I., SATTENTAU, Q. J. & GERSHONI, J. M. 1997. Conformational transitions in CD4 due to complexation with HIV envelope glycoprotein gp120. *Journal of immunology,* 158**,** 1157-64.

DEVRIES, J. K. & ZUBAY, G. 1967. DNA-directed peptide synthesis. II. The synthesis of the alpha-fragment of the enzyme beta-galactosidase. *Proceedings of the National Academy of Sciences of the United States of America,* 57**,** 1010-2.

DIEDERICHS, J. E. 1996. Plasma protein adsorption patterns on liposomes: establishment of analytical procedure. *Electrophoresis,* 17**,** 607-11.

DIMASI, N., FLOT, D., DUPEUX, F. & MARQUEZ, J. A. 2007. Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta crystallographica. Section F, Structural biology and crystallization communications,* 63**,** 204-8.

DIRCKX, L., LINDEMANN, D., ETTE, R., MANZONI, C., MORITZ, D. & MOUS, J. 1990. Mutation of conserved N-glycosylation sites around the CD4-binding site of human immunodeficiency virus type 1 GP120 affects viral infectivity. *Virus research,* 18**,** 9-20.

DISKIN, R., MARCOVECCHIO, P. M. & BJORKMAN, P. J. 2010. Structure of a clade C HIV-1 gp120 bound to CD4 and CD4-induced antibody reveals anti-CD4 polyreactivity. *Nature structural & molecular biology,* 17**,** 608-13.

DOMANSKA, U. M., KRUIZINGA, R. C., NAGENGAST, W. B., TIMMER-BOSSCHA, H., HULS, G., DE VRIES, E. G. & WALENKAMP, A. M. 2013. A review on CXCR4/CXCL12 axis in oncology: no place to hide. *European journal of cancer,* 49**,** 219-30.

DONDAPATI, S. K., WUSTENHAGEN, D. A., STRAUCH, E. & KUBICK, S. 2018. Cell-free production of pore forming toxins: Functional analysis of thermostable direct hemolysin from Vibrio parahaemolyticus. *Engineering in life sciences,* 18**,** 140-148.

DOORES, K. J., BONOMELLI, C., HARVEY, D. J., VASILJEVIC, S., DWEK, R. A., BURTON, D. R., CRISPIN, M. & SCANLAN, C. N. 2010. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 13800-5.

DORR, J. M., KOORENGEVEL, M. C., SCHAFER, M., PROKOFYEV, A. V., SCHEIDELAAR, S., VAN DER CRUIJSEN, E. A., DAFFORN, T. R., BALDUS, M. & KILLIAN, J. A. 2014. Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K+ channel: the power of native nanodiscs. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 18607-12.

DUNNE, O., WEIDENHAUPT, M., CALLOW, P., MARTEL, A., MOULIN, M., PERKINS, S. J., HAERTLEIN, M. & FORSYTH, V. T. 2017a. Matchout deuterium labelling of proteins for small-angle neutron scattering studies using prokaryotic and eukaryotic expression systems and high cell-density cultures. *Eur Biophys J,* 46**,** 425-432.

DUNNE, O., WEIDENHAUPT, M., CALLOW, P., MARTEL, A., MOULIN, M., PERKINS, S. J., HAERTLEIN, M. & FORSYTH, V. T. 2017b. Matchout deuterium labelling of proteins for small-angle neutron scattering studies using prokaryotic and eukaryotic expression systems and high cell-density cultures. *European biophysics journal : EBJ,* 46**,** 425-432.

DUONG-LY, K. C. & GABELLI, S. B. 2014. Salting out of proteins using ammonium sulfate precipitation. *Methods in enzymology,* 541**,** 85-94.

DURAND, D., VIVES, C., CANNELLA, D., PEREZ, J., PEBAY-PEYROULA, E., VACHETTE, P. & FIESCHI, F. 2010. NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology,* 169**,** 45-53.

EITEMAN, M. A. & ALTMAN, E. 2006. Overcoming acetate in Escherichia coli recombinant protein fermentations. *Trends in biotechnology,* 24**,** 530-6.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography,* 60**,** 2126-32.

ENDO, Y. & SAWASAKI, T. 2004. High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *Journal of structural and functional genomics,* 5**,** 45-57.

ENDOH, T., KANAI, T. & IMANAKA, T. 2007. A highly productive system for cell-free protein synthesis using a lysate of the hyperthermophilic archaeon, Thermococcus kodakaraensis. *Applied microbiology and biotechnology,* 74**,** 1153-61.

ENGELHARDT, B. & RANSOHOFF, R. M. 2012. Capture, crawl, cross: the T cell code to breach the blood-brain barriers. *Trends in immunology,* 33**,** 579-89.

ETEZADY-ESFARJANI, T., HILLER, S., VILLALBA, C. & WUTHRICH, K. 2007a. Cell-free protein synthesis of perdeuterated proteins for NMR studies. *Journal of biomolecular NMR,* 39**,** 229-38.

ETEZADY-ESFARJANI, T., HILLER, S., VILLALBA, C. & WUTHRICH, K. 2007b. Cell-free protein synthesis of perdeuterated proteins for NMR studies. *J Biomol NMR,* 39**,** 229-38.

FAILMEZGER, J., RAUTER, M., NITSCHEL, R., KRAML, M. & SIEMANN-HERZBERG, M. 2017. Cell-free protein synthesis from non-growing, stressed Escherichia coli. *Scientific reports,* 7**,** 16524.

FAVAUDON, V. & POCHON, F. 1976. Magnesium dependence of the association kinetics of Escherichia coli ribosomal subunits. *Biochemistry,* 15**,** 3903-12.

FENZ, S. F., SACHSE, R., SCHMIDT, T. & KUBICK, S. 2014. Cell-free synthesis of membrane proteins: tailored cell models out of microsomes. *Biochimica et biophysica acta,* 1838**,** 1382-8.

FINZI, A., PACHECO, B., ZENG, X., KWON, Y. D., KWONG, P. D. & SODROSKI, J. 2010. Conformational characterization of aberrant disulfide-linked HIV-1 gp120 dimers secreted from overexpressing cells. *Journal of virological methods,* 168**,** 155-61.

FRANKEN, L. E., BOEKEMA, E. J. & STUART, M. C. A. 2017. Transmission Electron Microscopy as a Tool for the Characterization of Soft Materials: Application and Interpretation. *Advanced science,* 4**,** 1600476.

FREEMAN, M. M., SEAMAN, M. S., RITS-VOLLOCH, S., HONG, X., KAO, C. Y., HO, D. D. & CHEN, B. 2010. Crystal structure of HIV-1 primary receptor CD4 in complex with a potent antiviral antibody. *Structure,* 18**,** 1632-41.

FUJIKI, Y., HUBBARD, A. L., FOWLER, S. & LAZAROW, P. B. 1982. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *The Journal of cell biology,* 93**,** 97-102.

GALE, E. F. & FOLKES, J. P. 1954. Effect of nucleic acids on protein synthesis and amino-acid incorporation in disrupted staphylococcal cells. *Nature,* 173**,** 1223-7.

GALLIS, B. M., MCDONNELL, J. P., HOPPER, J. E. & YOUNG, E. T. 1975. Translation of poly(riboadenylic acid)-enriched messenger RNAs from the yeast, Saccharomyces cerevisiae, in heterologous cell-free systems. *Biochemistry,* 14**,** 1038-46.

GALLUZZI, L. & KROEMER, G. 2007. Mitochondrial apoptosis without VDAC. *Nature cell biology,* 9**,** 487-9.

GARAMELLA, J., MARSHALL, R., RUSTAD, M. & NOIREAUX, V. 2016. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS synthetic biology,* 5**,** 344-55.

GARCES, F., SOK, D., KONG, L., MCBRIDE, R., KIM, H. J., SAYE-FRANCISCO, K. F., JULIEN, J. P., HUA, Y., CUPO, A., MOORE, J. P., PAULSON, J. C., WARD, A. B., BURTON, D. R. & WILSON, I. A. 2014. Structural evolution of glycan recognition by a family of potent HIV antibodies. *Cell,* 159**,** 69-79.

GARCIA, B. A., HAKE, S. B., DIAZ, R. L., KAUER, M., MORRIS, S. A., RECHT, J., SHABANOWITZ, J., MISHRA, N., STRAHL, B. D., ALLIS, C. D. & HUNT, D. F. 2007. Organismal differences in post-translational modifications in histones H3 and H4. *The Journal of biological chemistry,* 282**,** 7641-55.

GARLICK, R. L., KIRSCHNER, R. J., ECKENRODE, F. M., TARPLEY, W. G. & TOMICH, C. S. 1990. Escherichia coli expression, purification, and biological activity of a truncated soluble CD4. *AIDS research and human retroviruses,* 6**,** 465-79.

GOERKE, A. R. & SWARTZ, J. R. 2008. Development of cell-free protein synthesis platforms for disulfide bonded proteins. *Biotechnology and bioengineering,* 99**,** 351-67.

GOLDIE, K. N., ABEYRATHNE, P., KEBBEL, F., CHAMI, M., RINGLER, P. & STAHLBERG, H. 2014. Cryo-electron microscopy of membrane proteins. *Methods in molecular biology,* 1117**,** 325-41.

GOPING, I. S., GROSS, A., LAVOIE, J. N., NGUYEN, M., JEMMERSON, R., ROTH, K., KORSMEYER, S. J. & SHORE, G. C. 1998. Regulated targeting of BAX to mitochondria. *The Journal of cell biology,* 143**,** 207-15.

GOREN, M. A. & FOX, B. G. 2008. Wheat germ cell-free translation, purification, and assembly of a functional human stearoyl-CoA desaturase complex. *Protein expression and purification,* 62**,** 171-8.

GOZANSKY, E. K., LOUIS, J. M., CAFFREY, M. & CLORE, G. M. 2005. Mapping the binding of the N-terminal extracellular tail of the CXCR4 receptor to stromal cell-derived factor-1alpha. *Journal of molecular biology,* 345**,** 651-8.

GRIMALDO, M., LOPEZ, H., BECK, C., ROOSEN-RUNGE, F., MOULIN, M., DEVOS, J. M., LAUX, V., HARTLEIN, M., DA VELA, S., SCHWEINS, R., MARIANI, A., ZHANG, F., BARRAT, J. L., OETTEL, M., FORSYTH, V. T., SEYDEL, T. & SCHREIBER, F. 2019. Protein Short-Time Diffusion in a Naturally Crowded Environment. *The journal of physical chemistry letters***,** 1709-1715.

GRISSHAMMER, R. 2006. Understanding recombinant expression of membrane proteins. *Current opinion in biotechnology,* 17**,** 337-40.

GUIGNARD, L., OZAWA, K., PURSGLOVE, S. E., OTTING, G. & DIXON, N. E. 2002. NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. *FEBS letters,* 524**,** 159-62.

GUINIER, A. & FOURMET, G. 1955. Small-Angle X-ray scattering. *New York: John Wiley and Sons.*

GUTTMAN, M., KAHN, M., GARCIA, N. K., HU, S. L. & LEE, K. K. 2012. Solution structure, conformational dynamics, and CD4-induced activation in full-length, glycosylated, monomeric HIV gp120. *Journal of virology,* 86**,** 8750-64.

GUYON, A. 2014. CXCL12 chemokine and its receptors as major players in the interactions between immune and nervous systems. *Frontiers in cellular neuroscience,* 8**,** 65.

HABERSTOCK, S., ROOS, C., HOEVELS, Y., DOTSCH, V., SCHNAPP, G., PAUTSCH, A. & BERNHARD, F. 2012. A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. *Protein expression and purification,* 82**,** 308-16.

HAERTLEIN, M., MOULIN, M., DEVOS, J. M., LAUX, V., DUNNE, O. & FORSYTH, V. T. 2016. Biomolecular Deuteration for Neutron Structural Biology and Dynamics. *Methods in enzymology,* 566**,** 113-57.

HARRIS, L. J., LARSON, S. B., HASEL, K. W. & MCPHERSON, A. 1997. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry,* 36**,** 1581-97.

HARTWIG, A. 2001. Role of magnesium in genomic stability. *Mutation research,* 475**,** 113-21.

HAUSER, A. S., ATTWOOD, M. M., RASK-ANDERSEN, M., SCHIOTH, H. B. & GLORIAM, D. E. 2017. Trends in GPCR drug discovery: new agents, targets and indications. *Nature reviews. Drug discovery,* 16**,** 829-842.

HE, M. & TAUSSIG, M. J. 2007. Rapid discovery of protein interactions by cell-free protein technologies. *Biochemical Society transactions,* 35**,** 962-5.

HENDRIX, C. W., FLEXNER, C., MACFARLAND, R. T., GIANDOMENICO, C., FUCHS, E. J., REDPATH, E., BRIDGER, G. & HENSON, G. W. 2000. Pharmacokinetics and safety of AMD-3100, a novel antagonist of the CXCR-4 chemokine receptor, in human volunteers. *Antimicrobial agents and chemotherapy,* 44**,** 1667-73.

HICKEY, K. D. & BUHR, M. M. 2011. Lipid bilayer composition affects transmembrane protein orientation and function. *Journal of lipids,* 2011**,** 208457.

HODGMAN, C. E. & JEWETT, M. C. 2013. Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis. *Biotechnology and bioengineering,* 110**,** 2643-54.

HODGMAN, C. E. & JEWETT, M. C. 2014. Characterizing IGR IRES-mediated translation initiation for use in yeast cell-free protein synthesis. *New biotechnology,* 31**,** 499-505.

HOHLEFELDER, L. S., STOGBAUER, T., OPITZ, M., BAYERL, T. M. & RADLER, J. O. 2013. Heavy water reduces GFP expression in prokaryotic cell-free assays at the translation level while stimulating its transcription. *Biomed Res Int,* 2013**,** 592745.

HOPPE, S. M., SASAKI, D. Y., KINGHORN, A. N. & HATTAR, K. 2013. In-situ transmission electron microscopy of liposomes in an aqueous environment. *Langmuir : the ACS journal of surfaces and colloids,* 29**,** 9958-61.

HOSHINO, S. & SUGIYAMA, S. 1974. Protein synthesis in a cell-free preparation from rumen Protozoa. *Comparative biochemistry and physiology. B, Comparative biochemistry,* 48**,** 39-45.

HOUSTON, J. E., BRANDL, G., DROCHNER, M., KEMMERLING, G., ENGELS, R., PAPAGIANNOPOULOS, A., SARTER, M., STADLER, A. & RADULESCU, A. 2018. The high-intensity option of the SANS diffractometer KWS-2 at JCNS - characterization and performance of the new multi-megahertz detection system. *Journal of applied crystallography,* 51**,** 323-336.

HUNTER, A. R., FARRELL, P. J., JACKSON, R. J. & HUNT, T. 1977. The role of polyamines in cell-free protein synthesis in the wheat-germ system. *European journal of biochemistry,* 75**,** 149-57.

HUSSEY, R. E., RICHARDSON, N. E., KOWALSKI, M., BROWN, N. R., CHANG, H. C., SILICIANO, R. F., DORFMAN, T., WALKER, B., SODROSKI, J. & REINHERZ, E. L. 1988. A soluble CD4 protein selectively inhibits HIV replication and syncytium formation. *Nature,* 331**,** 78-81.

INCARDONA, M. F., BOURENKOV, G. P., LEVIK, K., PIERITZ, R. A., POPOV, A. N. & SVENSSON, O. 2009. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of synchrotron radiation,* 16**,** 872-9.

ISHIHARA, G., GOTO, M., SAEKI, M., ITO, K., HORI, T., KIGAWA, T., SHIROUZU, M. & YOKOYAMA, S. 2005. Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Protein expression and purification,* 41**,** 27-37.

JENNINGS, M. J., BARRIOS, A. F. & TAN, S. 2016. Elimination of truncated recombinant protein expressed in Escherichia coli by removing cryptic translation initiation site. *Protein expression and purification,* 121**,** 17-21.

JIN, D. K., SHIDO, K., KOPP, H. G., PETIT, I., SHMELKOV, S. V., YOUNG, L. M., HOOPER, A. T., AMANO, H., AVECILLA, S. T., HEISSIG, B., HATTORI, K., ZHANG, F., HICKLIN, D. J., WU, Y., ZHU, Z., DUNN, A., SALARI, H., WERB, Z., HACKETT, N. R., CRYSTAL, R. G., LYDEN, D. & RAFII, S. 2006. Cytokine-mediated deployment of SDF-1 induces revascularization through recruitment of CXCR4+ hemangiocytes. *Nature medicine,* 12**,** 557-67.

KAINOSHO, M., TORIZAWA, T., IWASHITA, Y., TERAUCHI, T., MEI ONO, A. & GUNTERT, P. 2006. Optimal isotope labelling for NMR protein structure determinations. *Nature,* 440**,** 52-7.

KAISER, L., GRAVELAND-BIKKER, J., STEUERWALD, D., VANBERGHEM, M., HERLIHY, K. & ZHANG, S. 2008. Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 15726-31.

KANG, S. H., KIM, D. M., KIM, H. J., JUN, S. Y. & LEE, K. Y. 2005. Cell-free production of aggregation-prone proteins in soluble and active forms. *Biotechnology progress,* 21**,** 1412-9.

KATRITCH, V., CHEREZOV, V. & STEVENS, R. C. 2013. Structure-function of the G protein-coupled receptor superfamily. *Annual review of pharmacology and toxicology,* 53**,** 531-56.

KAWASAKI, T., GOUDA, M. D., SAWASAKI, T., TAKAI, K. & ENDO, Y. 2003. Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *European journal of biochemistry,* 270**,** 4780-6.

KERN, J. A. & DAVIS, R. H. 1997. Application of solution equilibrium analysis to in vitro RNA transcription. *Biotechnology progress,* 13**,** 747-56.

KHAN, K. H. 2013. Gene expression in Mammalian cells and its applications. *Advanced pharmaceutical bulletin,* 3**,** 257-63.

KIGAWA, T., MUTO, Y. & YOKOYAMA, S. 1995. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *Journal of biomolecular NMR,* 6**,** 129-34.

KIGAWA, T., YABUKI, T., MATSUDA, N., MATSUDA, T., NAKAJIMA, R., TANAKA, A. & YOKOYAMA, S. 2004. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *Journal of structural and functional genomics,* 5**,** 63-8.

KIGAWA, T., YABUKI, T., YOSHIDA, Y., TSUTSUI, M., ITO, Y., SHIBATA, T. & YOKOYAMA, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS letters,* 442**,** 15-9.

KIM, D. M. & CHOI, C. Y. 1996. A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. *Biotechnology progress,* 12**,** 645-9.

KIM, D. M. & SWARTZ, J. R. 1999. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnology and bioengineering,* 66**,** 180-8.

KIM, D. M. & SWARTZ, J. R. 2000. Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnology progress,* 16**,** 385-90.

KIM, D. M. & SWARTZ, J. R. 2001. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnology and bioengineering,* 74**,** 309-16.

KIM, D. M. & SWARTZ, J. R. 2004. Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli. *Biotechnology and bioengineering,* 85**,** 122-9.

KIM, T. W., KEUM, J. W., OH, I. S., CHOI, C. Y., KIM, H. C. & KIM, D. M. 2007. An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source. *Journal of biotechnology,* 130**,** 389-93.

KIM, T. W., KEUM, J. W., OH, I. S., CHOI, C. Y., PARK, C. G. & KIM, D. M. 2006a. Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of biotechnology,* 126**,** 554-61.

KIM, T. W., KIM, D. M. & CHOI, C. Y. 2006b. Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system. *Journal of biotechnology,* 124**,** 373-80.

KLAMMT, C., LOHR, F., SCHAFER, B., HAASE, W., DOTSCH, V., RUTERJANS, H., GLAUBITZ, C. & BERNHARD, F. 2004a. High level cell-free expression and specific labeling of integral membrane proteins. *European journal of biochemistry / FEBS,* 271**,** 568-80.

KLAMMT, C., LOHR, F., SCHAFER, B., HAASE, W., DOTSCH, V., RUTERJANS, H., GLAUBITZ, C. & BERNHARD, F. 2004b. High level cell-free expression and specific labeling of integral membrane proteins. *European journal of biochemistry,* 271**,** 568-80.

KLAMMT, C., SCHWARZ, D., FENDLER, K., HAASE, W., DOTSCH, V. & BERNHARD, F. 2005. Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *The FEBS journal,* 272**,** 6024-38.

KLINE, S. R. 2006. Reduction and analysis of SANS and USANS data using IGOR Pro. *J. Appl. Crystallogr,* 39**,** 895–900.
KOBAYASHI, T., MACHIDA, K. & IMATAKA, H. 2014. Human cell extract-derived cell-free systems for virus synthesis. *Methods in molecular biology,* 1118**,** 149-56.
KOBAYASHI, T., MIKAMI, S., YOKOYAMA, S. & IMATAKA, H. 2007. An improved cell-free system for picornavirus synthesis. *Journal of virological methods,* 142**,** 182-8.
KOCH, A. L. 1968. Theory of the angular dependence of light scattered by bacteria and similar-sized biological objects. *Journal of theoretical biology,* 18**,** 133-56.
KOCH, A. L. 1970. Turbidity measurements of bacterial cultures in some available commercial instruments. *Analytical biochemistry,* 38**,** 252-9.
KOFUKU, Y., YOSHIURA, C., UEDA, T., TERASAWA, H., HIRAI, T., TOMINAGA, S., HIROSE, M., MAEDA, Y., TAKAHASHI, H., TERASHIMA, Y., MATSUSHIMA, K. & SHIMADA, I. 2009. Structural basis of the interaction between chemokine stromal cell-derived factor-1/CXCL12 and its G-protein-coupled receptor CXCR4. *The Journal of biological chemistry,* 284**,** 35240-50.
KOLB, V. A., MAKEYEV, E. V. & SPIRIN, A. S. 2000. Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. *The Journal of biological chemistry,* 275**,** 16597-601.
KONAREV, P., VOLKOV, V., SOKOLOVA, A., KOCH, M., AND SVERGUN, D. 2003. PRIMUS: a windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallography,* 36**,** 1277–1282.
KONG, L., WILSON, I. A. & KWONG, P. D. 2015. Crystal structure of a fully glycosylated HIV-1 gp120 core reveals a stabilizing role for the glycan at Asn262. *Proteins,* 83**,** 590-6.
KOPPEL, D. A., KINNALLY, K. W., MASTERS, P., FORTE, M., BLACHLY-DYSON, E. & MANNELLA, C. A. 1998. Bacterial expression and characterization of the mitochondrial outer membrane channel. Effects of n-terminal modifications. *The Journal of biological chemistry,* 273**,** 13794-800.
KOVTUN, O., MUREEV, S., JUNG, W., KUBALA, M. H., JOHNSTON, W. & ALEXANDROV, K. 2011. Leishmania cell-free protein expression system. *Methods,* 55**,** 58-64.
KRAMER, G., BOEHRINGER, D., BAN, N. & BUKAU, B. 2009. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature structural & molecular biology,* 16**,** 589-97.
KRATKY, O. & POROD, G. 1949. Diffuse small-angle scattering of X-rays in colloid systems. *Journal of colloid science,* 4**,** 35-70.
KRIEG, P. A. & MELTON, D. A. 1987. In vitro RNA synthesis with SP6 RNA polymerase. *Methods in enzymology,* 155**,** 397-415.
KRINSKY, N., KADURI, M., SHAINSKY-ROITMAN, J., GOLDFEDER, M., IVANIR, E., BENHAR, I., SHOHAM, Y. & SCHROEDER, A. 2016. A Simple and Rapid Method for Preparing a Cell-Free Bacterial Lysate for Protein Synthesis. *PloS one,* 11**,** e0165137.
KUCIA, M., JANKOWSKI, K., RECA, R., WYSOCZYNSKI, M., BANDURA, L., ALLENDORF, D. J., ZHANG, J., RATAJCZAK, J. & RATAJCZAK, M. Z. 2004. CXCR4-SDF-1 signalling, locomotion, chemotaxis and adhesion. *Journal of molecular histology,* 35**,** 233-45.
KWON, Y. C. & JEWETT, M. C. 2015. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific reports,* 5**,** 8663.
KWONG, P. D., RYU, S. E., HENDRICKSON, W. A., AXEL, R., SWEET, R. M., FOLENA-WASSERMAN, G., HENSLEY, P. & SWEET, R. W. 1990. Molecular characteristics of recombinant human CD4 as deduced from polymorphic crystals. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 6423-7.
KWONG, P. D., WYATT, R., ROBINSON, J., SWEET, R. W., SODROSKI, J. & HENDRICKSON, W. A. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature,* 393**,** 648-59.

LAGUERRE, A., LOHR, F., HENRICH, E., HOFFMANN, B., ABDUL-MANAN, N., CONNOLLY, P. J., PEROZO, E., MOORE, J. M., BERNHARD, F. & DOTSCH, V. 2016. From Nanodiscs to Isotropic Bicelles: A Procedure for Solution Nuclear Magnetic Resonance Studies of Detergent-Sensitive Integral Membrane Proteins. *Structure,* 24**,** 1830-1841.

LAMBORG, M. R. & ZAMECNIK, P. C. 1960. Amino acid incorporation into protein by extracts of E. coli. *Biochimica et biophysica acta,* 42**,** 206-11.

LAURSEN, B. S., SORENSEN, H. P., MORTENSEN, K. K. & SPERLING-PETERSEN, H. U. 2005. Initiation of protein synthesis in bacteria. *Microbiology and molecular biology reviews : MMBR,* 69**,** 101-23.

LAUX, V., CALLOW, P., SVERGUN, D. I., TIMMINS, P. A., FORSYTH, V. T. & HAERTLEIN, M. 2008. Selective deuteration of tryptophan and methionine residues in maltose binding protein: a model system for neutron scattering. *European biophysics journal : EBJ,* 37**,** 815-22.

LE MAIRE, M., CHAMPEIL, P. & MOLLER, J. V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et biophysica acta,* 1508**,** 86-111.

LEE, B., SHARRON, M., BLANPAIN, C., DORANZ, B. J., VAKILI, J., SETOH, P., BERG, E., LIU, G., GUY, H. R., DURELL, S. R., PARMENTIER, M., CHANG, C. N., PRICE, K., TSANG, M. & DOMS, R. W. 1999. Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *The Journal of biological chemistry,* 274**,** 9617-26.

LEMASTERS, J. J. & HOLMUHAMEDOV, E. 2006. Voltage-dependent anion channel (VDAC) as mitochondrial governator--thinking outside the box. *Biochimica et biophysica acta,* 1762**,** 181-90.

LEVI, V. & GONZALEZ FLECHA, F. L. 2002. Reversible fast-dimerization of bovine serum albumin detected by fluorescence resonance energy transfer. *Biochimica et biophysica acta,* 1599**,** 141-8.

LI, G. W., OH, E. & WEISSMAN, J. S. 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature,* 484**,** 538-41.

LI, H., CHIEN, P. C., JR., TUEN, M., VISCIANO, M. L., COHEN, S., BLAIS, S., XU, C. F., ZHANG, H. T. & HIOE, C. E. 2008. Identification of an N-linked glycosylation in the C4 region of HIV-1 envelope gp120 that is critical for recognition of neighboring CD4 T cell epitopes. *Journal of immunology,* 180**,** 4011-21.

LIGUORI, L. & LENORMAND, J. L. 2009. Production of recombinant proteoliposomes for therapeutic uses. *Methods in enzymology,* 465**,** 209-23.

LIGUORI, L., MARQUES, B. & LENORMAND, J. L. 2008. A bacterial cell-free expression system to produce membrane proteins and proteoliposomes: from cDNA to functional assay. *Current protocols in protein science,* Chapter 5**,** Unit 5 22.

LIGUORI, L., MARQUES, B., VILLEGAS-MENDEZ, A., ROTHE, R. & LENORMAND, J. L. 2007. Production of membrane proteins using cell-free expression systems. *Expert review of proteomics,* 4**,** 79-90.

LIGUORI, L., PASTORINO, F., ROUSSET, X., ALFANO, S., CORTES, S., EMIONITE, L., DAGA, A., PONZONI, M. & LENORMAND, J. L. 2015. Anti-Tumor Effects of Bak-Proteoliposomes against Glioblastoma. *Molecules,* 20**,** 15893-909.

LITTLEFIELD, J. W., KELLER, E. B., GROSS, J. & ZAMECNIK, P. C. 1955. Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *The Journal of biological chemistry,* 217**,** 111-23.

LIU, D. V., ZAWADA, J. F. & SWARTZ, J. R. 2005. Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnology progress,* 21**,** 460-5.

LIU, J., BARTESAGHI, A., BORGNIA, M. J., SAPIRO, G. & SUBRAMANIAM, S. 2008. Molecular architecture of native HIV-1 gp120 trimers. *Nature,* 455**,** 109-13.

LOBANOV, M., BOGATYREVA, N. S. & GALZITSKAIA, O. V. 2008. [Radius of gyration is indicator of compactness of protein structure]. *Molekuliarnaia biologiia,* 42**,** 701-6.

LOBSTEIN, J., EMRICH, C. A., JEANS, C., FAULKNER, M., RIGGS, P. & BERKMEN, M. 2012. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial cell factories,* 11**,** 56.

LOETSCHER, P., GONG, J. H., DEWALD, B., BAGGIOLINI, M. & CLARK-LEWIS, I. 1998. N-terminal peptides of stromal cell-derived factor-1 with CXC chemokine receptor 4 agonist and antagonist activities. *J Biol Chem,* 273**,** 22279-83.

LOETSCHER, P., MOSER, B. & BAGGIOLINI, M. 2000. Chemokines and their receptors in lymphocyte traffic and HIV infection. *Advances in immunology,* 74**,** 127-80.

LONG, A. R., O'BRIEN, C. C. & ALDER, N. N. 2012. The cell-free integration of a polytopic mitochondrial membrane protein into liposomes occurs cotranslationally and in a lipid-dependent manner. *PloS one,* 7**,** e46332.

LYUKMANOVA, E. N., SHENKAREV, Z. O., KHABIBULLINA, N. F., KOPEINA, G. S., SHULEPKO, M. A., PARAMONOV, A. S., MINEEV, K. S., TIKHONOV, R. V., SHINGAROVA, L. N., PETROVSKAYA, L. E., DOLGIKH, D. A., ARSENIEV, A. S. & KIRPICHNIKOV, M. P. 2012. Lipid-protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: comparison with detergent micelles, bicelles and liposomes. *Biochimica et biophysica acta,* 1818**,** 349-58.

MADIN, K., SAWASAKI, T., OGASAWARA, T. & ENDO, Y. 2000. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 559-64.

MAJOREK, K. A., POREBSKI, P. J., DAYAL, A., ZIMMERMAN, M. D., JABLONSKA, K., STEWART, A. J., CHRUSZCZ, M. & MINOR, W. 2012. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Molecular immunology,* 52**,** 174-82.

MARCUS, A. & FEELEY, J. 1966. Ribosome activation and polysome formation in vitro: requirement for ATP. *Proceedings of the National Academy of Sciences of the United States of America,* 56**,** 1770-7.

MARIC, S., SKAR-GISLINGE, N., MIDTGAARD, S., THYGESEN, M. B., SCHILLER, J., FRIELINGHAUS, H., MOULIN, M., HAERTLEIN, M., FORSYTH, V. T., POMORSKI, T. G. & ARLETH, L. 2014. Stealth carriers for low-resolution structure determination of membrane proteins in solution. *Acta crystallographica. Section D, Biological crystallography,* 70**,** 317-28.

MARISCH, K., BAYER, K., SCHARL, T., MAIRHOFER, J., KREMPL, P. M., HUMMEL, K., RAZZAZI-FAZELI, E. & STRIEDNER, G. 2013. A comparative analysis of industrial Escherichia coli K-12 and B strains in high-glucose batch cultivations on process-, transcriptome- and proteome level. *PloS one,* 8**,** e70516.

MATSUDA, T., KOSHIBA, S., TOCHIO, N., SEKI, E., IWASAKI, N., YABUKI, T., INOUE, M., YOKOYAMA, S. & KIGAWA, T. 2007. Improving cell-free protein synthesis for stable-isotope labeling. *Journal of biomolecular NMR,* 37**,** 225-9.

MATSUMOTO, K., TOMIKAWA, C., TOYOOKA, T., OCHI, A., TAKANO, Y., TAKAYANAGI, N., ABE, M., ENDO, Y. & HORI, H. 2008. Production of yeast tRNA (m(7)G46) methyltransferase (Trm8-Trm82 complex) in a wheat germ cell-free translation system. *Journal of biotechnology,* 133**,** 453-60.

MATTHAEI, H. & NIRENBERG, M. W. 1961. The dependence of cell-free protein synthesis in E. coli upon RNA prepared from ribosomes. *Biochemical and biophysical research communications,* 4**,** 404-8.

MCCOMMIS, K. S. & BAINES, C. P. 2012. The role of VDAC in cell death: friend or foe? *Biochimica et biophysica acta,* 1818**,** 1444-50.

MCLELLAN, J. S., PANCERA, M., CARRICO, C., GORMAN, J., JULIEN, J. P., KHAYAT, R., LOUDER, R., PEJCHAL, R., SASTRY, M., DAI, K., O'DELL, S., PATEL, N., SHAHZAD-UL-HUSSAN, S., YANG, Y., ZHANG, B., ZHOU, T., ZHU, J., BOYINGTON, J. C., CHUANG, G. Y., DIWANJI, D., GEORGIEV, I., KWON, Y. D., LEE, D., LOUDER, M. K., MOQUIN, S., SCHMIDT, S. D., YANG, Z. Y., BONSIGNORI, M.,

CRUMP, J. A., KAPIGA, S. H., SAM, N. E., HAYNES, B. F., BURTON, D. R., KOFF, W. C., WALKER, L. M., PHOGAT, S., WYATT, R., ORWENYO, J., WANG, L. X., ARTHOS, J., BEWLEY, C. A., MASCOLA, J. R., NABEL, G. J., SCHIEF, W. R., WARD, A. B., WILSON, I. A. & KWONG, P. D. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature,* 480**,** 336-43.

MERTENS, H. D. T. & SVERGUN, D. I. 2017. Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Archives of biochemistry and biophysics,* 628**,** 33-41.

MICHEL, E. & WUTHRICH, K. 2012. Cell-free expression of disulfide-containing eukaryotic proteins for structural biology. *The FEBS journal,* 279**,** 3176-84.

MIKAMI, S., KOBAYASHI, T. & IMATAKA, H. 2010. Cell-free protein synthesis systems with extracts from cultured human cells. *Methods in molecular biology,* 607**,** 43-52.

MIKAMI, S., KOBAYASHI, T., MASUTANI, M., YOKOYAMA, S. & IMATAKA, H. 2008. A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins. *Protein expression and purification,* 62**,** 190-8.

MIROUX, B. & WALKER, J. E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology,* 260**,** 289-98.

MITSUOKA, K. & GERLE, C. 2016. [Recent advances of single particle analysis by cryo-electron microscopy and its application to membrane proteins]. *Seikagaku. The Journal of Japanese Biochemical Society,* 88**,** 532-6.

MOEPPS, B., BRAUN, M., KNOPFLE, K., DILLINGER, K., KNOCHEL, W. & GIERSCHIK, P. 2000. Characterization of a Xenopus laevis CXC chemokine receptor 4: implications for hematopoietic cell development in the vertebrate embryo. *European journal of immunology,* 30**,** 2924-34.

MOLDOVEANU, T., LIU, Q., TOCILJ, A., WATSON, M., SHORE, G. & GEHRING, K. 2006. The X-ray structure of a BAK homodimer reveals an inhibitory zinc binding site. *Molecular cell,* 24**,** 677-688.

MONACO, S., GORDON, E., BOWLER, M. W., DELAGENIERE, S., GUIJARRO, M., SPRUCE, D., SVENSSON, O., MCSWEENEY, S. M., MCCARTHY, A. A., LEONARD, G. & NANAO, M. H. 2013. Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *Journal of applied crystallography,* 46**,** 804-810.

MOORE, J. M., PATAPOFF, T. W. & CROMWELL, M. E. 1999. Kinetics and thermodynamics of dimer formation and dissociation for a recombinant humanized monoclonal antibody to vascular endothelial growth factor. *Biochemistry,* 38**,** 13960-7.

MOORE, J. P., SATTENTAU, Q. J., KLASSE, P. J. & BURKLY, L. C. 1992. A monoclonal antibody to CD4 domain 2 blocks soluble CD4-induced conformational changes in the envelope glycoproteins of human immunodeficiency virus type 1 (HIV-1) and HIV-1 infection of CD4+ cells. *Journal of virology,* 66**,** 4784-93.

MULLER, A., HOMEY, B., SOTO, H., GE, N., CATRON, D., BUCHANAN, M. E., MCCLANAHAN, T., MURPHY, E., YUAN, W., WAGNER, S. N., BARRERA, J. L., MOHAR, A., VERASTEGUI, E. & ZLOTNIK, A. 2001. Involvement of chemokine receptors in breast cancer metastasis. *Nature,* 410**,** 50-6.

MURAKAMI, T. & YAMAMOTO, N. 2010. Role of CXCR4 in HIV infection and its potential as a therapeutic target. *Future microbiology,* 5**,** 1025-39.

MURPHY, J. W., CHO, Y., SACHPATZIDIS, A., FAN, C., HODSDON, M. E. & LOLIS, E. 2007. Structural and functional basis of CXCL12 (stromal cell-derived factor-1 alpha) binding to heparin. *The Journal of biological chemistry,* 282**,** 10018-27.

MURPHY, J. W., YUAN, H., KONG, Y., XIONG, Y. & LOLIS, E. J. 2010. Heterologous quaternary structure of CXCL12 and its relationship to the CC chemokine family. *Proteins,* 78**,** 1331-7.

MURSHUDOV, G. N., SKUBAK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography,* 67**,** 355-67.

MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography,* 53**,** 240-55.

MYSZKA, D. G., SWEET, R. W., HENSLEY, P., BRIGHAM-BURKE, M., KWONG, P. D., HENDRICKSON, W. A., WYATT, R., SODROSKI, J. & DOYLE, M. L. 2000. Energetics of the HIV gp120-CD4 binding reaction. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 9026-31.

NAGASAWA, T., HIROTA, S., TACHIBANA, K., TAKAKURA, N., NISHIKAWA, S., KITAMURA, Y., YOSHIDA, N., KIKUTANI, H. & KISHIMOTO, T. 1996. Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature,* 382**,** 635-8.

NEERATHILINGAM, M., GREENE, L. H., COLEBROOKE, S. A., CAMPBELL, I. D. & STAUNTON, D. 2005. Quantitation of protein expression in a cell-free system: Efficient detection of yields and 19F NMR to identify folded protein. *Journal of biomolecular NMR,* 31**,** 11-9.

NEVES, P., LOPES, S. C., SOUSA, I., GARCIA, S., EATON, P. & GAMEIRO, P. 2009. Characterization of membrane protein reconstitution in LUVs of different lipid composition by fluorescence anisotropy. *Journal of pharmaceutical and biomedical analysis,* 49**,** 276-81.

NGUYEN, T. A., LIEU, S. S. & CHANG, G. 2010. An Escherichia coli-based cell-free system for large-scale production of functional mammalian membrane proteins suitable for X-ray crystallography. *Journal of molecular microbiology and biotechnology,* 18**,** 85-91.

NIEMANN, H. H., PETOUKHOV, M. V., HARTLEIN, M., MOULIN, M., GHERARDI, E., TIMMINS, P., HEINZ, D. W. & SVERGUN, D. I. 2008. X-ray and neutron small-angle scattering analysis of the complex formed by the Met receptor and the Listeria monocytogenes invasion protein InlB. *Journal of molecular biology,* 377**,** 489-500.

NIRENBERG, M. W. & MATTHAEI, J. H. 1961. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America,* 47**,** 1588-602.

NIWA, T., SASAKI, Y., UEMURA, E., NAKAMURA, S., AKIYAMA, M., ANDO, M., SAWADA, S., MUKAI, S. A., UEDA, T., TAGUCHI, H. & AKIYOSHI, K. 2015. Comprehensive study of liposome-assisted synthesis of membrane proteins using a reconstituted cell-free translation system. *Scientific reports,* 5**,** 18025.

NOZAWA, A., OGASAWARA, T., MATSUNAGA, S., IWASAKI, T., SAWASAKI, T. & ENDO, Y. 2011. Production and partial purification of membrane proteins using a liposome-supplemented wheat cell-free translation system. *BMC biotechnology,* 11**,** 35.

O'CONNELL, R. J., KIM, J. H. & EXCLER, J. L. 2014. The HIV-1 gp120 V1V2 loop: structure, function and importance for vaccine development. *Expert review of vaccines,* 13**,** 1489-500.

OHNISHI, Y., SENDA, T., NANDHAGOPAL, N., SUGIMOTO, K., SHIODA, T., NAGAL, Y. & MITSUI, Y. 2000. Crystal structure of recombinant native SDF-1alpha with additional mutagenesis studies: an attempt at a more comprehensive interpretation of accumulated structure-activity relationship data. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research,* 20**,** 691-700.

OLDHAM, M. L. & CHEN, J. 2011. Snapshots of the maltose transporter during ATP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 15152-6.

OLDHAM, M. L., KHARE, D., QUIOCHO, F. A., DAVIDSON, A. L. & CHEN, J. 2007. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature,* 450**,** 515-21.

OPEL, K. L., CHUNG, D. & MCCORD, B. R. 2010. A study of PCR inhibition mechanisms using real time PCR. *Journal of forensic sciences,* 55**,** 25-33.

ORIMO, A., GUPTA, P. B., SGROI, D. C., ARENZANA-SEISDEDOS, F., DELAUNAY, T., NAEEM, R., CAREY, V. J., RICHARDSON, A. L. & WEINBERG, R. A. 2005. Stromal fibroblasts present in invasive human

breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell,* 121**,** 335-48.

OVERINGTON, J. P., AL-LAZIKANI, B. & HOPKINS, A. L. 2006. How many drug targets are there? *Nature reviews. Drug discovery,* 5**,** 993-6.

OZOROWSKI, G., PALLESEN, J., DE VAL, N., LYUMKIS, D., COTTRELL, C. A., TORRES, J. L., COPPS, J., STANFIELD, R. L., CUPO, A., PUGACH, P., MOORE, J. P., WILSON, I. A. & WARD, A. B. 2017. Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. *Nature,* 547**,** 360-363.

PANICO, M., BOUCHE, L., BINET, D., O'CONNOR, M. J., RAHMAN, D., PANG, P. C., CANIS, K., NORTH, S. J., DESROSIERS, R. C., CHERTOVA, E., KEELE, B. F., BESS, J. W., JR., LIFSON, J. D., HASLAM, S. M., DELL, A. & MORRIS, H. R. 2016. Mapping the complete glycoproteome of virion-derived HIV-1 gp120 provides insights into broadly neutralizing antibody binding. *Scientific reports,* 6**,** 32956.

PAPAGEORGIOU, F., PIPPA, N., NAZIRIS, N. & DEMETZOS, C. 2018. Physicochemical Study of the Protein-Liposome Interactions: Influence of Liposome Composition and Concentration on Protein Binding. *Journal of liposome research***,** 1-29.

PAREEK, V., BHARGAVA, A., BHANOT, V., GUPTA, R., JAIN, N. & PANWAR, J. 2018. Formation and Characterization of Protein Corona Around Nanoparticles: A Review. *Journal of nanoscience and nanotechnology,* 18**,** 6653-6670.

PATRA, S. K., ALONSO, A. & GONI, F. M. 1998. Detergent solubilisation of phospholipid bilayers in the gel state: the role of polar and hydrophobic forces. *Biochimica et biophysica acta,* 1373**,** 112-8.

PEGU, A., YANG, Z. Y., BOYINGTON, J. C., WU, L., KO, S. Y., SCHMIDT, S. D., MCKEE, K., KONG, W. P., SHI, W., CHEN, X., TODD, J. P., LETVIN, N. L., HUANG, J., NASON, M. C., HOXIE, J. A., KWONG, P. D., CONNORS, M., RAO, S. S., MASCOLA, J. R. & NABEL, G. J. 2014. Neutralizing antibodies to HIV-1 envelope protect more effectively in vivo than those to the CD4 receptor. *Science translational medicine,* 6**,** 243ra88.

PELHAM, H. R. & JACKSON, R. J. 1976. An efficient mRNA-dependent translation system from reticulocyte lysates. *European journal of biochemistry,* 67**,** 247-56.

PETOUKHOV, M. V., FRANKE, D., SHKUMATOV, A. V., TRIA, G., KIKHNEY, A. G., GAJDA, M., GORBA, C., MERTENS, H. D., KONAREV, P. V. & SVERGUN, D. I. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of applied crystallography,* 45**,** 342-350.

PETOUKHOV, M. V. & SVERGUN, D. I. 2005. Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys J,* 89**,** 1237-1250.

PHUE, J. N. & SHILOACH, J. 2004. Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in E. coli B (BL21) and E. coli K (JM109). *Journal of biotechnology,* 109**,** 21-30.

PICCIOCCHI, A., SIAUCIUNAITEE-GAUBARD, L., PETIT-HARTLEIN, I., SADIR, R., REVILLOUD, J., CARO, L., VIVAUDOU, M., FIESCHI, F., MOREAU, C. & VIVES, C. 2014. C-terminal engineering of CXCL12 and CCL5 chemokines: functional characterization by electrophysiological recordings. *PloS one,* 9**,** e87394.

PLATH, F., RINGLER, P., GRAFF-MEYER, A., STAHLBERG, H., LAUER, M. E., RUFER, A. C., GRAEWERT, M. A., SVERGUN, D., GELLERMANN, G., FINKLER, C., STRACKE, J. O., KOULOV, A. & SCHNAIBLE, V. 2016. Characterization of mAb dimers reveals predominant dimer forms common in therapeutic mAbs. *mAbs,* 8**,** 928-40.

PLOTKIN, J. B. & KUDLA, G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics,* 12**,** 32-42.

POTTERTON, E., BRIGGS, P., TURKENBURG, M. & DODSON, E. 2003. A graphical user interface to the CCP4 program suite. *Acta Crystallographica. ,* D59**,** 1131-1137.

PRATT 1984. Coupled Transcription-Translation in Prokaryotic Cell-Free Systems. *Transcription and translation: a practical approach***,** 179-209.
PUTNAM, C. D. 2016. Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. *Journal of applied crystallography,* 49**,** 1412-1419.
QIN, L., KUFAREVA, I., HOLDEN, L. G., WANG, C., ZHENG, Y., ZHAO, C., FENALTI, G., WU, H., HAN, G. W., CHEREZOV, V., ABAGYAN, R., STEVENS, R. C. & HANDEL, T. M. 2015. Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science,* 347**,** 1117-22.
RADULESCU, A., SZEKELY, N. K., APPAVOU, M. S., PIPICH, V., KOHNKE, T., OSSOVYI, V., STARINGER, S., SCHNEIDER, G. J., AMANN, M., ZHANG-HAAGEN, B., BRANDL, G., DROCHNER, M., ENGELS, R., HANSLIK, R. & KEMMERLING, G. 2016. Studying Soft-matter and Biological Systems over a Wide Length-scale from Nanometer and Micrometer Sizes at the Small-angle Neutron Diffractometer KWS-2. *Journal of visualized experiments : JoVE*.
RASHEED, M., BETTADAPURA, R. & BAJAJ, C. 2015. Computational Refinement and Validation Protocol for Proteins with Large Variable Regions Applied to Model HIV Env Spike in CD4 and 17b Bound State. *Structure,* 23**,** 1138-49.
RAWSON, S., DAVIES, S., LIPPIAT, J. D. & MUENCH, S. P. 2016. The changing landscape of membrane protein structural biology through developments in electron microscopy. *Molecular membrane biology,* 33**,** 12-22.
RECEVEUR-BRECHOT, V. & DURAND, D. 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Current protein & peptide science,* 13**,** 55-75.
REED, J. C. 2006. Proapoptotic multidomain Bcl-2/Bax-family proteins: mechanisms, physiological roles, and therapeutic opportunities. *Cell death and differentiation,* 13**,** 1378-86.
REID, D. W. & NICCHITTA, C. V. 2012. Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *The Journal of biological chemistry,* 287**,** 5518-27.
RITS-VOLLOCH, S., FREY, G., HARRISON, S. C. & CHEN, B. 2006. Restraining the conformation of HIV-1 gp120 by removing a flexible loop. *The EMBO journal,* 25**,** 5026-35.
RIZZUTO, C. D., WYATT, R., HERNANDEZ-RAMOS, N., SUN, Y., KWONG, P. D., HENDRICKSON, W. A. & SODROSKI, J. 1998. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science,* 280**,** 1949-53.
ROBERTS, B. E. & PATERSON, B. M. 1973. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. *Proceedings of the National Academy of Sciences of the United States of America,* 70**,** 2330-4.
ROOS, C., KAI, L., HABERSTOCK, S., PROVERBIO, D., GHOSHDASTIDER, U., MA, Y., FILIPEK, S., WANG, X., DOTSCH, V. & BERNHARD, F. 2014. High-level cell-free production of membrane proteins with nanodiscs. *Methods in molecular biology,* 1118**,** 109-30.
ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology,* 5**,** 172.
ROSENBAUM, D. M., CHEREZOV, V., HANSON, M. A., RASMUSSEN, S. G., THIAN, F. S., KOBILKA, T. S., CHOI, H. J., YAO, X. J., WEIS, W. I., STEVENS, R. C. & KOBILKA, B. K. 2007. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science,* 318**,** 1266-73.
ROTHBLATT, J. A. & MEYER, D. I. 1986. Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system. *Cell,* 44**,** 619-28.
ROTHMAN, J. E. & ORCI, L. 1992. Molecular dissection of the secretory pathway. *Nature,* 355**,** 409-15.
RYU, E. K., KIM, T. G., KWON, T. H., JUNG, I. D., RYU, D., PARK, Y. M., KIM, J., AHN, K. H. & BAN, C. 2007. Crystal structure of recombinant human stromal cell-derived factor-1alpha. *Proteins,* 67**,** 1193-7.

RYU, S. E., KWONG, P. D., TRUNEH, A., PORTER, T. G., ARTHOS, J., ROSENBERG, M., DAI, X. P., XUONG, N. H., AXEL, R., SWEET, R. W. & *ET AL.* 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature,* 348**,** 419-26.

RYU, S. E., TRUNEH, A., SWEET, R. W. & HENDRICKSON, W. A. 1994. Structures of an HIV and MHC binding fragment from human CD4 as refined in two crystal lattices. *Structure,* 2**,** 59-74.

SADIR, R., BALEUX, F., GROSDIDIER, A., IMBERTY, A. & LORTAT-JACOB, H. 2001. Characterization of the stromal cell-derived factor-1alpha-heparin complex. *The Journal of biological chemistry,* 276**,** 8288-96.

SAKAI, K., HAMANAKA, R., YUKI, H. & WATANABE, M. 2009. A novel fractionation method of the rough ER integral membrane proteins; resident proteins versus exported proteins? *Proteomics,* 9**,** 3036-46.

SALEHI, A. S., SMITH, M. T., BENNETT, A. M., WILLIAMS, J. B., PITT, W. G. & BUNDY, B. C. 2016. Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. *Biotechnology journal,* 11**,** 274-81.

SAMMALKORPI, M., KARTTUNEN, M. & HAATAJA, M. 2007. Structural properties of ionic detergent aggregates: a large-scale molecular dynamics study of sodium dodecyl sulfate. *The journal of physical chemistry. B,* 111**,** 11722-33.

SANDERS, R. W., VAN ANKEN, E., NABATOV, A. A., LISCALJET, I. M., BONTJER, I., EGGINK, D., MELCHERS, M., BUSSER, E., DANKERS, M. M., GROOT, F., BRAAKMAN, I., BERKHOUT, B. & PAXTON, W. A. 2008. The carbohydrate at asparagine 386 on HIV-1 gp120 is not essential for protein folding and function but is involved in immune evasion. *Retrovirology,* 5**,** 10.

SANGRA, M., ESTELRICH, J., SABATE, R., ESPARGARO, A. & BUSQUETS, M. A. 2017. Evidence of Protein Adsorption in Pegylated Liposomes: Influence of Liposomal Decoration. *Nanomaterials,* 7.

SAPHIRE, E. O., PARREN, P. W., PANTOPHLET, R., ZWICK, M. B., MORRIS, G. M., RUDD, P. M., DWEK, R. A., STANFIELD, R. L., BURTON, D. R. & WILSON, I. A. 2001. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science,* 293**,** 1155-9.

SARRAMEGNA, V., TALMONT, F., DEMANGE, P. & MILON, A. 2003. Heterologous expression of G-protein-coupled receptors: comparison of expression systems from the standpoint of large-scale production and purification. *Cellular and molecular life sciences : CMLS,* 60**,** 1529-46.

SASAKI, Y., ASAYAMA, W., NIWA, T., SAWADA, S., UEDA, T., TAGUCHI, H. & AKIYOSHI, K. 2011. Amphiphilic polysaccharide nanogels as artificial chaperones in cell-free protein synthesis. *Macromolecular bioscience,* 11**,** 814-20.

SATTENTAU, Q. J. & MOORE, J. P. 1991. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *The Journal of experimental medicine,* 174**,** 407-15.

SAUNDERS, C. J., MCCAFFREY, R. A., ZHARKIKH, I., KRAFT, Z., MALENBAUM, S. E., BURKE, B., CHENG-MAYER, C. & STAMATATOS, L. 2005. The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *Journal of virology,* 79**,** 9069-80.

SAWASAKI, T., OGASAWARA, T., MORISHITA, R. & ENDO, Y. 2002. A cell-free protein synthesis system for high-throughput proteomics. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 14652-7.

SCHARF, L., SCHEID, J. F., LEE, J. H., WEST, A. P., JR., CHEN, C., GAO, H., GNANAPRAGASAM, P. N., MARES, R., SEAMAN, M. S., WARD, A. B., NUSSENZWEIG, M. C. & BJORKMAN, P. J. 2014. Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell reports,* 7**,** 785-95.

SCHOBORG, J. A., HODGMAN, C. E., ANDERSON, M. J. & JEWETT, M. C. 2014. Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis. *Biotechnology journal,* 9**,** 630-40.

SCHREIBER, G. & URBAN, J. 1978. The synthesis and secretion of albumin. *Reviews of physiology, biochemistry and pharmacology,* 82**,** 27-95.

SCHWARZ, D., DOTSCH, V. & BERNHARD, F. 2008. Production of membrane proteins using cell-free expression systems. *Proteomics,* 8**,** 3933-46.

SCHWEET, R., LAMFROM, H. & ALLEN, E. 1958. The Synthesis of Hemoglobin in a Cell-Free System. *Proceedings of the National Academy of Sciences of the United States of America,* 44**,** 1029-35.

SHILOACH, J., KAUFMAN, J., GUILLARD, A. S. & FASS, R. 1996. Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by Escherichia coli BL21 (lambdaDE3) and Escherichia coli JM109. *Biotechnology and bioengineering,* 49**,** 421-8.

SHIMIZU, Y., INOUE, A., TOMARI, Y., SUZUKI, T., YOKOGAWA, T., NISHIKAWA, K. & UEDA, T. 2001. Cell-free translation reconstituted with purified components. *Nature biotechnology,* 19**,** 751-5.

SHIN, J. & NOIREAUX, V. 2010. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of biological engineering,* 4**,** 8.

SHORE, G. C. 2009. Apoptosis: it's BAK to VDAC. *EMBO reports,* 10**,** 1311-3.

SICILIANO, S. J., ROLLINS, T. E., DEMARTINO, J., KONTEATIS, Z., MALKOWITZ, L., VAN RIPER, G., BONDY, S., ROSEN, H. & SPRINGER, M. S. 1994. Two-site binding of C5a by its receptor: an alternative binding paradigm for G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 1214-8.

SIEKEVITZ, P. 1952. Uptake of radioactive alanine in vitro into the proteins of rat liver fractions. *The Journal of biological chemistry,* 195**,** 549-65.

SINIBALDI, R., ORTORE, M. G., SPINOZZI, F., DE SOUZA FUNARI, S., TEIXEIRA, J. & MARIANI, P. 2008. SANS/SAXS study of the BSA solvation properties in aqueous urea solutions via a global fit approach. *European biophysics journal : EBJ,* 37**,** 673-81.

SISSONS, C. H. 1974. Yeast protein synthesis. Preparation and analysis of a highly active cell-free system. *The Biochemical journal,* 144**,** 131-40.

SMITH, E. W., LIU, Y., GETSCHMAN, A. E., PETERSON, F. C., ZIAREK, J. J., LI, R., VOLKMAN, B. F. & CHEN, Y. 2014. Structural analysis of a novel small molecule ligand bound to the CXCL12 chemokine. *Journal of medicinal chemistry,* 57**,** 9693-9.

SONG, R., FRANCO, D., KAO, C. Y., YU, F., HUANG, Y. & HO, D. D. 2010. Epitope mapping of ibalizumab, a humanized anti-CD4 monoclonal antibody with anti-HIV-1 activity in infected patients. *Journal of virology,* 84**,** 6935-42.

SORANZO, T., CORTES, S., GILDE, F., KREIR, M., PICART, C. & LENORMAND, J. L. 2016. Functional characterization of p7 viroporin from hepatitis C virus produced in a cell-free expression system. *Protein expression and purification,* 118**,** 83-91.

SORANZO, T., MARTIN, D. K., LENORMAND, J. L. & WATKINS, E. B. 2017. Coupling neutron reflectivity with cell-free protein synthesis to probe membrane protein structure in supported bilayers. *Scientific reports,* 7**,** 3399.

SPIRIN, A. S., BARANOV, V. I., RYABOVA, L. A., OVODOV, S. Y. & ALAKHOV, Y. B. 1988. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science,* 242**,** 1162-4.

SPURLINO, J. C., LU, G. Y. & QUIOCHO, F. A. 1991. The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *The Journal of biological chemistry,* 266**,** 5202-19.

STAUNTON, D., SCHLINKERT, R., ZANETTI, G., COLEBROOK, S. A. & CAMPBELL, I. D. 2006. Cell-free expression and selective isotope labelling in protein NMR. *Magnetic resonance in chemistry : MRC,* 44 Spec No**,** S2-9.

STECH, M., HUST, M., SCHULZE, C., DUBEL, S. & KUBICK, S. 2014. Cell-free eukaryotic systems for the production, engineering, and modification of scFv antibody fragments. *Engineering in life sciences,* 14**,** 387-398.

STECH, M., MERK, H., SCHENK, J. A., STOCKLEIN, W. F., WUSTENHAGEN, D. A., MICHEEL, B., DUSCHL, C., BIER, F. F. & KUBICK, S. 2012. Production of functional antibody fragments in a vesicle-based eukaryotic cell-free translation system. *Journal of biotechnology,* 164**,** 220-31.

STEVENSON, K., MCVEY, A. F., CLARK, I. B., SWAIN, P. S. & PILIZOTA, T. 2016. General calibration of microbial growth in microplate readers. *Scientific reports,* 6**,** 38828.

STUEBER, D., IBRAHIMI, I., CUTLER, D., DOBBERSTEIN, B. & BUJARD, H. 1984. A novel in vitro transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. *The EMBO journal,* 3**,** 3143-8.

SULLIVAN, N., SUN, Y., SATTENTAU, Q., THALI, M., WU, D., DENISOVA, G., GERSHONI, J., ROBINSON, J., MOORE, J. & SODROSKI, J. 1998. CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *Journal of virology,* 72**,** 4694-703.

SUN, Z. Z., HAYES, C. A., SHIN, J., CASCHERA, F., MURRAY, R. M. & NOIREAUX, V. 2013. Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of visualized experiments : JoVE***,** e50762.

SVENSSON, O., GILSKI, M., NURIZZO, D. & BOWLER, M. W. 2018. Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta crystallographica. Section D, Structural biology,* 74**,** 433-440.

SVENSSON, O., MALBET-MONACO, S., POPOV, A., NURIZZO, D. & BOWLER, M. W. 2015. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta crystallographica. Section D, Biological crystallography,* 71**,** 1757-67.

SVERGUN 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallography,* 25**,** 495-503.

SVERGUN, D. I. & KOCH, M. H. 2003. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics,* 66**,** 1735 - 1782.

SVERGUN, D. I., RICHARD, S., KOCH, M. H., SAYERS, Z., KUPRIN, S. & ZACCAI, G. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 2267-72.

SWEET, R. W., TRUNEH, A. & HENDRICKSON, W. A. 1991. CD4: its structure, role in immune function and AIDS pathogenesis, and potential as a pharmacological target. *Current opinion in biotechnology,* 2**,** 622-33.

TAKEDA, H., OGASAWARA, T., OZAWA, T., MURAGUCHI, A., JIH, P. J., MORISHITA, R., UCHIGASHIMA, M., WATANABE, M., FUJIMOTO, T., IWASAKI, T., ENDO, Y. & SAWASAKI, T. 2015. Production of monoclonal antibodies against GPCR using cell-free synthesized GPCR antigen and biotinylated liposome-based interaction assay. *Scientific reports,* 5**,** 11333.

TAKEDA, M., IKEYA, T., GUNTERT, P. & KAINOSHO, M. 2007. Automated structure determination of proteins with the SAIL-FLYA NMR method. *Nature protocols,* 2**,** 2896-902.

TAMAMIS, P. & FLOUDAS, C. A. 2014. Molecular recognition of CCR5 by an HIV-1 gp120 V3 loop. *PloS one,* 9**,** e95767.

TARUI, H., MURATA, M., TANI, I., IMANISHI, S., NISHIKAWA, S. & HARA, T. 2001. Establishment and characterization of cell-free translation/glycosylation in insect cell (Spodoptera frugiperda 21) extract prepared with high pressure treatment. *Applied microbiology and biotechnology,* 55**,** 446-53.

THIEBE, R. 1975. Aminoacylation of tRNA. Magnesium requirement and spermidine effect. *FEBS letters,* 51**,** 259-61.

TONELLI, M., SINGARAPU, K. K., MAKINO, S., SAHU, S. C., MATSUBARA, Y., ENDO, Y., KAINOSHO, M. & MARKLEY, J. L. 2011. Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J Biomol NMR,* 51**,** 467-76.

TREWHELLA, J., DUFF, A. P., DURAND, D., GABEL, F., GUSS, J. M., HENDRICKSON, W. A., HURA, G. L., JACQUES, D. A., KIRBY, N. M., KWAN, A. H., PEREZ, J., POLLACK, L., RYAN, T. M., SALI, A., SCHNEIDMAN-DUHOVNY, D., SCHWEDE, T., SVERGUN, D. I., SUGIYAMA, M., TAINER, J. A., VACHETTE, P., WESTBROOK, J. & WHITTEN, A. E. 2017. 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta crystallographica. Section D, Structural biology,* 73**,** 710-728.

TSE, T. P. & TAYLOR, J. M. 1977. Translation of albumin messenger RNA in a cell-free protein-synthesizing system derived from wheat germ. *The Journal of biological chemistry,* 252**,** 1272-8.

TSUBOI, T., TAKEO, S., IRIKO, H., JIN, L., TSUCHIMOCHI, M., MATSUDA, S., HAN, E. T., OTSUKI, H., KANEKO, O., SATTABONGKOT, J., UDOMSANGPETCH, R., SAWASAKI, T., TORII, M. & ENDO, Y. 2008. Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infection and immunity,* 76**,** 1702-8.

UEDA, H., SIANI, M. A., GONG, W., THOMPSON, D. A., BROWN, G. G. & WANG, J. M. 1997. Chemically synthesized SDF-1alpha analogue, N33A, is a potent chemotactic agent for CXCR4/Fusin/LESTR-expressing human leukocytes. *The Journal of biological chemistry,* 272**,** 24966-70.

VAN DALEN, A., HEGGER, S., KILLIAN, J. A. & DE KRUIJFF, B. 2002. Influence of lipids on membrane assembly and stability of the potassium channel KcsA. *FEBS letters,* 525**,** 33-8.

VAN OERS, M. M., PIJLMAN, G. P. & VLAK, J. M. 2015. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *The Journal of general virology,* 96**,** 6-23.

VANATALU, K., PAALME, T., VILU, R., BURKHARDT, N., JUNEMANN, R., MAY, R., RUHL, M., WADZACK, J. & NIERHAUS, K. H. 1993. Large-scale preparation of fully deuterated cell components. Ribosomes from Escherichia coli with high biological activity. *European journal of biochemistry,* 216**,** 315-21.

VELDKAMP, C. T., PETERSON, F. C., PELZEK, A. J. & VOLKMAN, B. F. 2005. The monomer-dimer equilibrium of stromal cell-derived factor-1 (CXCL 12) is altered by pH, phosphate, sulfate, and heparin. *Protein science : a publication of the Protein Society,* 14**,** 1071-81.

VELDKAMP, C. T., SEIBERT, C., PETERSON, F. C., DE LA CRUZ, N. B., HAUGNER, J. C., 3RD, BASNET, H., SAKMAR, T. P. & VOLKMAN, B. F. 2008a. Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Science signaling,* 1**,** ra4.

VELDKAMP, C. T., SEIBERT, C., PETERSON, F. C., DE LA CRUZ, N. B., HAUGNER, J. C., 3RD, BASNET, H., SAKMAR, T. P. & VOLKMAN, B. F. 2008b. Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci Signal,* 1**,** ra4.

VERA, A., GONZALEZ-MONTALBAN, N., ARIS, A. & VILLAVERDE, A. 2007. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and bioengineering,* 96**,** 1101-6.

VERHULST, P.-F. 1845. Mathematical Researches into the Law of Population Growth Increase. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles,* 18**,** 1–42.

WAGNER, S., BADER, M. L., DREW, D. & DE GIER, J. W. 2006. Rationalizing membrane protein overexpression. *Trends in biotechnology,* 24**,** 364-71.

WALDIE, S., MOULIN, M., PORCAR, L., PICHLER, H., STROHMEIER, G. A., SKODA, M., FORSYTH, V. T., HAERTLEIN, M., MARIC, S. & CARDENAS, M. 2019. The Production of Matchout-Deuterated Cholesterol and the Study of Bilayer-Cholesterol Interactions. *Scientific reports,* 9**,** 5118.

WALDMANN, T. A. 2003. Immunotherapy: past, present and future. *Nature medicine,* 9**,** 269-77.

WALKER, L. M., HUBER, M., DOORES, K. J., FALKOWSKA, E., PEJCHAL, R., JULIEN, J. P., WANG, S. K., RAMOS, A., CHAN-HUI, P. Y., MOYLE, M., MITCHAM, J. L., HAMMOND, P. W., OLSEN, O. A., PHUNG, P., FLING, S., WONG, C. H., PHOGAT, S., WRIN, T., SIMEK, M. D., KOFF, W. C., WILSON, I. A., BURTON, D. R. & POIGNARD, P. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature,* 477**,** 466-70.

WALKER, L. M., PHOGAT, S. K., CHAN-HUI, P. Y., WAGNER, D., PHUNG, P., GOSS, J. L., WRIN, T., SIMEK, M. D., FLING, S., MITCHAM, J. L., LEHRMAN, J. K., PRIDDY, F. H., OLSEN, O. A., FREY, S. M., HAMMOND, P. W., KAMINSKY, S., ZAMB, T., MOYLE, M., KOFF, W. C., POIGNARD, P. & BURTON, D. R. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science,* 326**,** 285-9.

WALLIN, E. & VON HEIJNE, G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein science : a publication of the Protein Society,* 7**,** 1029-38.

WATANABE, M., MIYAZONO, K., TANOKURA, M., SAWASAKI, T., ENDO, Y. & KOBAYASHI, I. 2010. Cell-free protein synthesis for structure determination by X-ray crystallography. *Methods in molecular biology,* 607**,** 149-60.

WEBB, N. R., MADOULET, C., TOSI, P. F., BROUSSARD, D. R., SNEED, L., NICOLAU, C. & SUMMERS, M. D. 1989. Cell-surface expression and purification of human CD4 produced in baculovirus-infected insect cells. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 7731-5.

WELSH, J. P., BONOMO, J. & SWARTZ, J. R. 2011. Localization of BiP to translating ribosomes increases soluble accumulation of secreted eukaryotic proteins in an Escherichia coli cell-free system. *Biotechnology and bioengineering,* 108**,** 1739-48.

WHITAKER, W. R., LEE, H., ARKIN, A. P. & DUEBER, J. E. 2015. Avoidance of truncated proteins from unintended ribosome binding sites within heterologous protein coding sequences. *ACS synthetic biology,* 4**,** 249-57.

WHITE, G. F., RACHER, K. I., LIPSKI, A., HALLETT, F. R. & WOOD, J. M. 2000. Physical properties of liposomes and proteoliposomes prepared from Escherichia coli polar lipids. *Biochimica et biophysica acta,* 1468**,** 175-86.

WHITE, S. 1998 - 2018. *Membrane proteins of known 3D structures* [Online]. Stephen White laboratory at UC Irvine. Available: http://blanco.biomol.uci.edu/mpstruc/ [Accessed 23.05.2018 2018].

WILHELM, D., BEHNKEN, H. N. & MEYER, B. 2012. Glycosylation assists binding of HIV protein gp120 to human CD4 receptor. *Chembiochem : a European journal of chemical biology,* 13**,** 524-7.

WORST, E. G., EXNER, M. P., DE SIMONE, A., SCHENKELBERGER, M., NOIREAUX, V., BUDISA, N. & OTT, A. 2015. Cell-free expression with the toxic amino acid canavanine. *Bioorganic & medicinal chemistry letters,* 25**,** 3658-60.

WU, B., CHIEN, E. Y., MOL, C. D., FENALTI, G., LIU, W., KATRITCH, V., ABAGYAN, R., BROOUN, A., WELLS, P., BI, F. C., HAMEL, D. J., KUHN, P., HANDEL, T. M., CHEREZOV, V. & STEVENS, R. C. 2010. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science,* 330**,** 1066-71.

WU, H., KWONG, P. D. & HENDRICKSON, W. A. 1997. Dimeric association and segmental variability in the structure of human CD4. *Nature,* 387**,** 527-30.

WYATT, R., MOORE, J., ACCOLA, M., DESJARDIN, E., ROBINSON, J. & SODROSKI, J. 1995. Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *Journal of virology,* 69**,** 5723-33.

YACHOU, A. & SEKALY, R. P. 1999. Binding of soluble recombinant HIV envelope glycoprotein, rgp120, induces conformational changes in the cellular membrane-anchored CD4 molecule. *Biochemical and biophysical research communications,* 265**,** 428-33.

YAMAGUCHI, H. & MIYAZAKI, M. 2014. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules,* 4**,** 235-51.

YANG, J., KANTER, G., VOLOSHIN, A., LEVY, R. & SWARTZ, J. R. 2004. Expression of active murine granulocyte-macrophage colony-stimulating factor in an Escherichia coli cell-free system. *Biotechnology progress,* 20**,** 1689-96.

YIN, G. & SWARTZ, J. R. 2004. Enhancing multiple disulfide bonded protein folding in a cell-free system. *Biotechnology and bioengineering,* 86**,** 188-95.

YIN, Y., WANG, X. X. & MARIUZZA, R. A. 2012. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 5405-10.

ZAMECNIK, P. C., FRANTZ, I. D., JR. & *ET AL.* 1948. Incorporation in vitro of radioactive carbon from carboxyl-labeled dl-alanine and glycine into proteins of normal and malignant rat livers. *The Journal of biological chemistry,* 175**,** 299-314.

ZAWADA, J. & SWARTZ, J. 2005. Maintaining rapid growth in moderate-density Escherichia coli fermentations. *Biotechnology and bioengineering,* 89**,** 407-15.

ZAWADA, J. & SWARTZ, J. 2006. Effects of growth rate on cell extract performance in cell-free protein synthesis. *Biotechnology and bioengineering,* 94**,** 618-24.

ZAWADA, J. F., YIN, G., STEINER, A. R., YANG, J., NARESH, A., ROY, S. M., GOLD, D. S., HEINSOHN, H. G. & MURRAY, C. J. 2011a. Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnology and bioengineering,* 108**,** 1570-8.

ZAWADA, J. F., YIN, G., STEINER, A. R., YANG, J., NARESH, A., ROY, S. M., GOLD, D. S., HEINSOHN, H. G. & MURRAY, C. J. 2011b. Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnol Bioeng,* 108**,** 1570-8.

ZEIRA, M., TOSI, P. F., MOUNEIMNE, Y., LAZARTE, J., SNEED, L., VOLSKY, D. J. & NICOLAU, C. 1991. Full-length CD4 electroinserted in the erythrocyte membrane as a long-lived inhibitor of infection by human immunodeficiency virus. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 4409-13.

ZEMELLA, A., THORING, L., HOFFMEISTER, C. & KUBICK, S. 2015. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem : a European journal of chemical biology,* 16**,** 2420-31.

ZHANG, F., SKODA, M. W., JACOBS, R. M., MARTIN, R. A., MARTIN, C. M. & SCHREIBER, F. 2007. Protein interactions studied by SAXS: effect of ionic strength and protein concentration for BSA in aqueous solutions. *The journal of physical chemistry. B,* 111**,** 251-9.

ZHANG, S., KRIVOSHEYEVA, A. & NOCHUMSON, S. 2003. Large-scale capture and partial purification of plasmid DNA using anion-exchange membrane capsules. *Biotechnology and applied biochemistry,* 37**,** 245-9.

ZHANG, W., CANZIANI, G., PLUGARIU, C., WYATT, R., SODROSKI, J., SWEET, R., KWONG, P., HENDRICKSON, W. & CHAIKEN, I. 1999. Conformational changes of gp120 in epitopes near the CCR5 binding site are induced by CD4 and a CD4 miniprotein mimetic. *Biochemistry,* 38**,** 9405-16.

ZHANG, W., GODILLOT, A. P., WYATT, R., SODROSKI, J. & CHAIKEN, I. 2001. Antibody 17b binding at the coreceptor site weakens the kinetics of the interaction of envelope glycoprotein gp120 with CD4. *Biochemistry,* 40**,** 1662-70.

ZHDANOV, V. P. & CHO, N. J. 2016. Kinetics of the formation of a protein corona around nanoparticles. *Mathematical biosciences,* 282**,** 82-90.

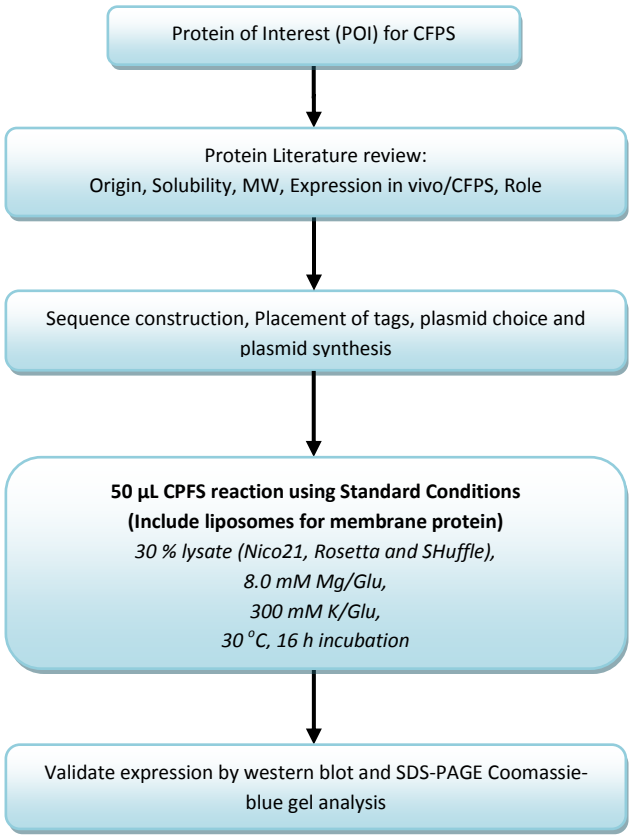
ZOU, Y. R., KOTTMANN, A. H., KURODA, M., TANIUCHI, I. & LITTMAN, D. R. 1998. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature,* 393**,** 595-9.

ZWICK, M. B., PARREN, P. W., SAPHIRE, E. O., CHURCH, S., WANG, M., SCOTT, J. K., DAWSON, P. E., WILSON, I. A. & BURTON, D. R. 2003. Molecular features of the broadly neutralizing immunoglobulin G1 b12 required for recognition of human immunodeficiency virus type 1 gp120. *Journal of virology,* 77**,** 5863-76.

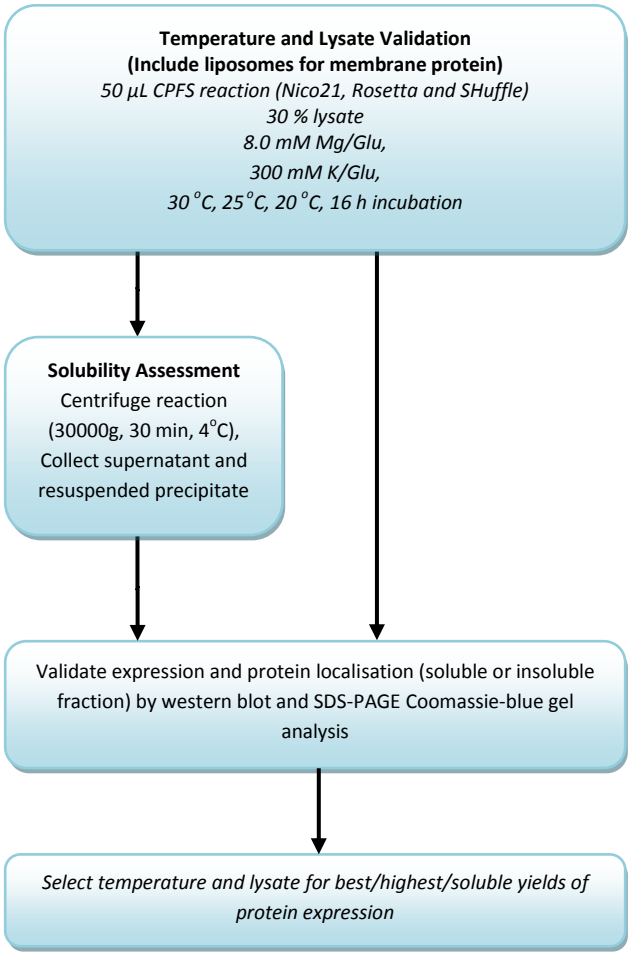
ZWIETERING, M. H., JONGENBURGER, I., ROMBOUTS, F. M. & VAN 'T RIET, K. 1990. Modeling of the bacterial growth curve. *Applied and environmental microbiology,* 56**,** 1875-81.

## APPENDICES: Standard Operating Protocols

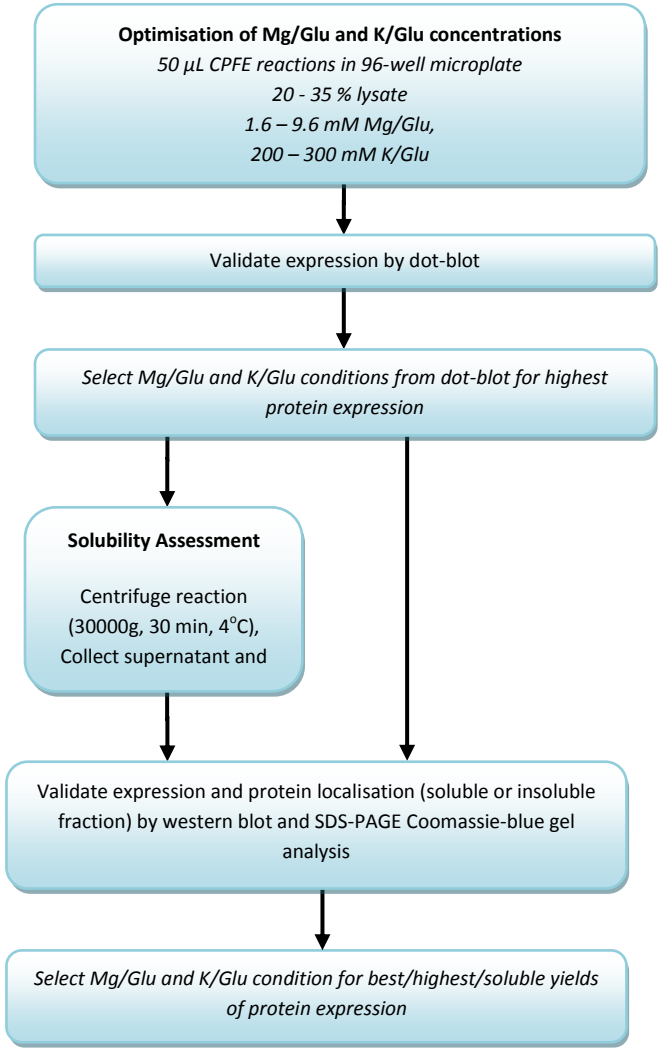
### 1. Preliminary CFPS of new soluble/membrane protein



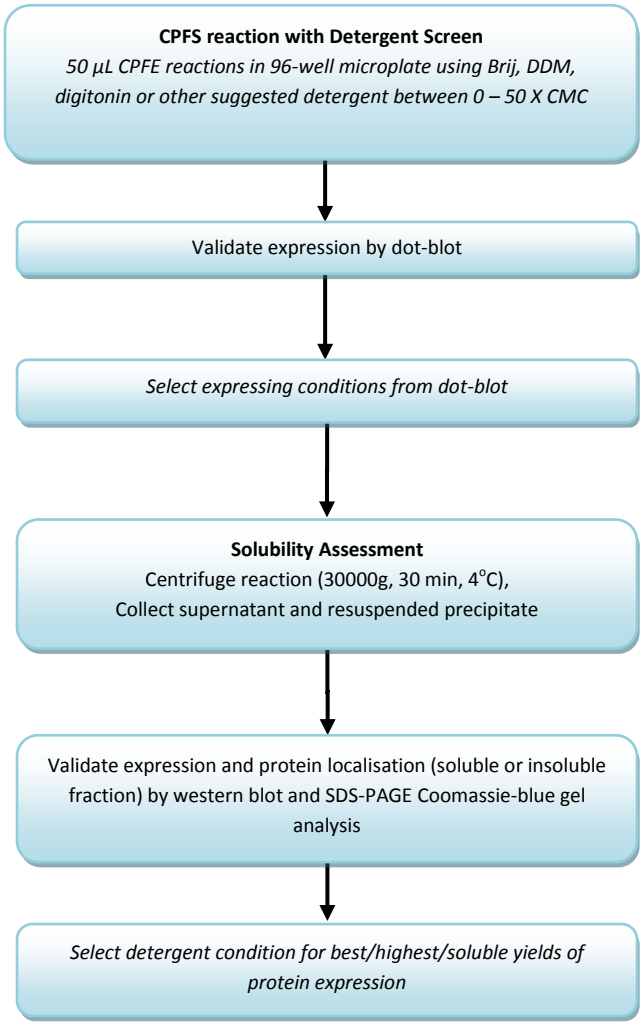

### 2. Optimising soluble/membrane protein CFPS: *temperature and lysate selection*



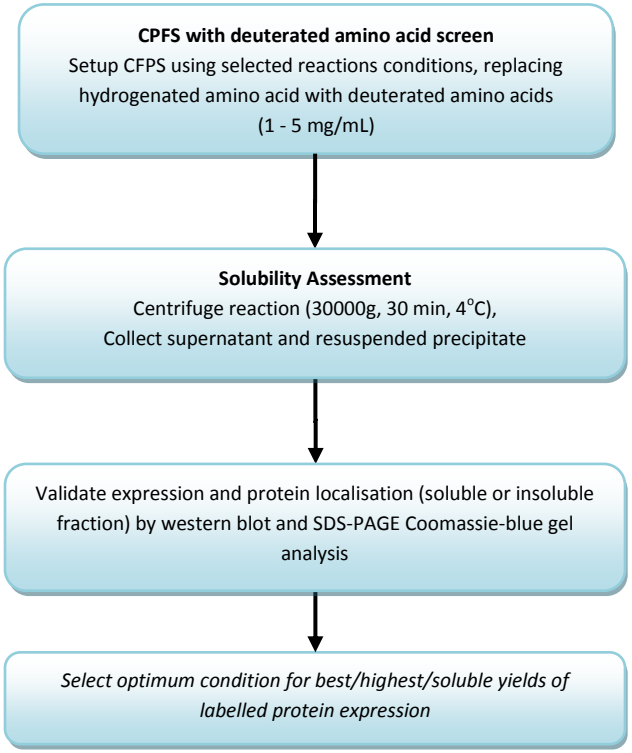
### 3. Optimising soluble/membrane protein CFPS: *magnesium glutamate and potassium glutamate screen*



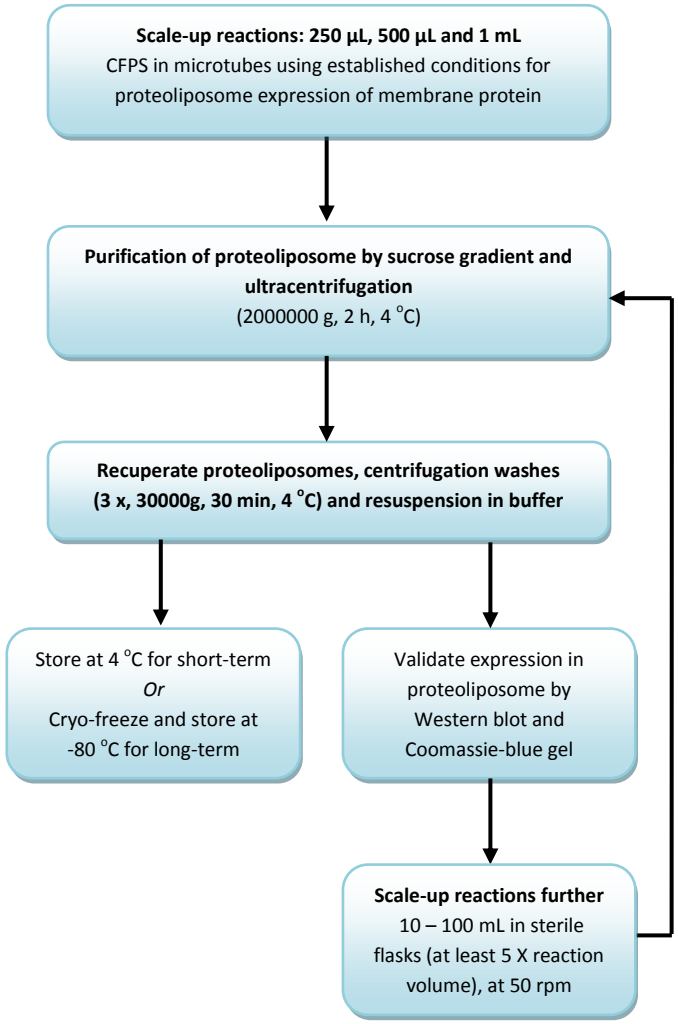

### 4. Optimising CFPS: *solubilised membrane protein in detergent*



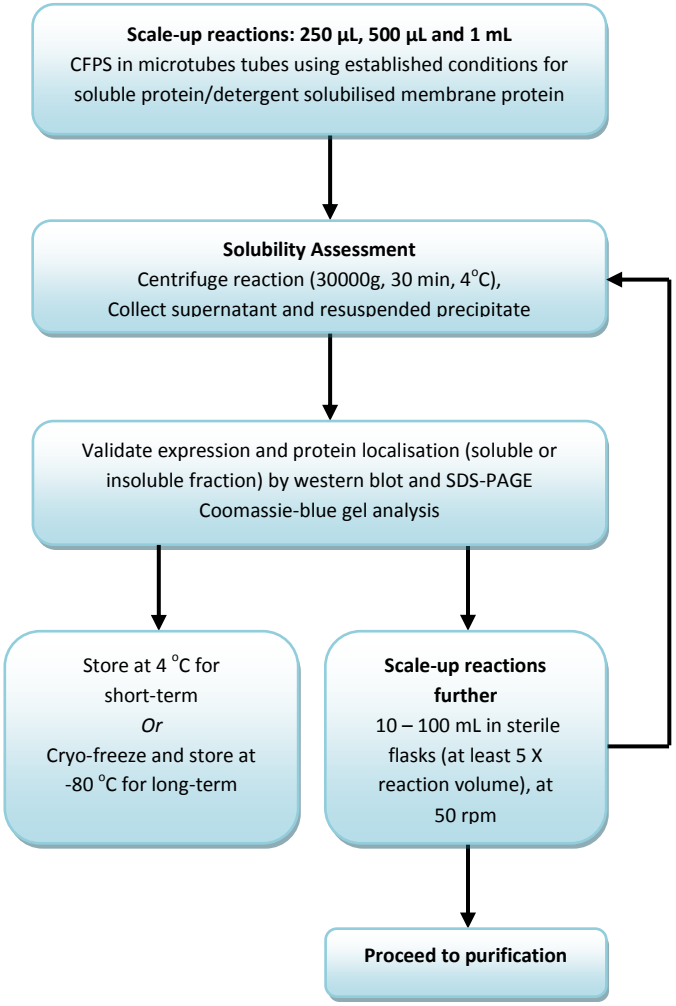
## 5. Optimising CFPS: deuterated protein expression




## 6. Scaling up membrane protein CFPS and proteoliposome purification



## 7. Scaling up soluble protein/detergent solubilised membrane protein CFPS




## 8. Purification of soluble protein/detergent solubilised membrane protein