Neutrons reveal the structure and dynamics of osteopontin phosphopeptides and nanocluster formation

Keele University

Samuel Lenton

A thesis submitted for the degree of

Doctor of Philosophy

March 2017

Abstract

In bovine milk the highly phosphorylated cleavage fragment of osteopontin (OPN1-149) is capable of stabilising high concentrations of calcium phosphate ions by formation of calcium phosphate nanoclusters (CPN). Here OPN1-149 isolated from bovine milk is characterised by a range of biophysical techniques as adopting an intrinsically disordered conformation in solution. The transient ensemble of conformations sampled by OPN1-149 is characterised by small angle X-ray scattering (SAXS) in combination with the ensemble optimisation method. Formation of CPN by OPN1-149 is characterised using dynamic light scattering and small angle neutron scattering (SANS). The dynamics of the protein and the effects of sequestration on the molecular fluctuations are quantified on the nanosecond-angstrom resolution by elastic incoherent neutron scattering. The molecular fluctuations of the free phosphopeptide are found to represent a highly flexible protein. Upon sequestration of calcium phosphate, stiffening of OPN1-149 is reflected in molecular fluctuations more closely resembling those characteristic of globular proteins. The quantification of the effects provides a handle for future comparisons and classification of molecular fluctuations. The formed OPN1-149 calcium phosphate nanoclusters are characterised according to a core shell model using SANS and static light scattering. The structure of the core is probed using neutron and X-ray diffraction. The results bring insight into the modulation of the activity of OPN1-149 and phosphopeptides with a role in the control of biomineralisation.

A recombinant human fragment (hOPN1-157) equivalent to OPN1-149 is produced and characterised in order to extend the work further towards the human regulatory system. The effects of phosphorylation, a key requirement for sequestration of calcium phosphate, on the structure and dynamics are determined using SAXS and elastic incoherent neutron scattering. The results indicate that phosphorylation of the protein has limited impact on both the structure and dynamics of hOPN1-157

Contents

List of Figures

Chapter 1

Introduction - Biological background

1.1 Fundamentals of biomineralisation

Mineralised tissue is characterised by the presence of deposited mineral such as hydroxyapatite (HA), which has the chemical formula $Ca_5(PO_4)_3(OH)$. Healthy soft tissue, on the other hand, is characterised by a lack of mineral content and the deposition of minerals in soft tissue is often detrimental to health [1]. In vertebrates the most abundant mineralised tissue is that of bone, from which the skeleton is constructed. The most important property of bone material is its high resilience through which it fulfils its fundamental function of providing form and strength to the mammalian body. The high resilience is due to the nanoscale hierarchical structure, where a mineral, often HA, is deposited within in an organic matrix such as collagen [2]. The hierarchical structure of bone is shown in Figure 1.1. In mammalian bone crystalline HA is deposited within an organic matrix consisting of long strands of collagen protein [3, 4]. The strength afforded to mineralised tissue depends on the nanoscale on not only the amount of HA deposited but also on the interaction between HA and collagen [5].

In vivo soft and hard tissue are often in contact with each-other and certain biological fluids can permeate both tissue types. For example teeth contain large amounts of tissue mineralised by crystalline calcium phosphate. In the mouth teeth are constantly exposed to saliva which due to its acidic pH would be expected to increase the rate

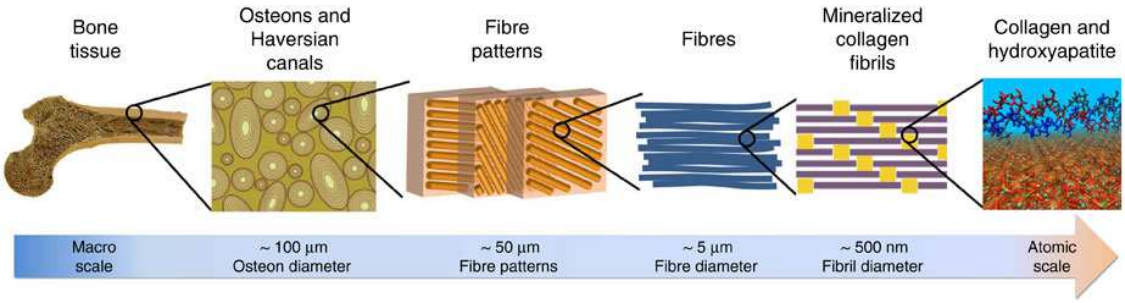

Figure 1.1: Illustration of bone structure at different resolutions, on the atomic scale bone consists of deposits of crystalline HA within an organic matrix of collagen. Image taken from Nair *et al* [3]

of tooth demineralisation [6, 7, 8]. In order to maintain the integrity of the tooth mineralised tissue, mammals produce saliva containing a high contents of calcium and phosphate [9]. Saliva is in almost constant contact with the tooth surface and therefore the mineral contents of tooth is constantly replenished [9]. The same is true for other biological fluids which are in contact with mineral fronts, such as those found in the extracellular matrix of bone. Milk contains the highest concentration of calcium phosphate of all biofluids [10]. The high concentration of calcium phosphate is required in order to provide the neonate with sufficient calcium and phosphate to develop and maintain it's skeletal structure [11]. Milk is transported from the mother to the neonate through the soft tissue present in the mammary gland, the calcification of which is rare [11]. A consequence of this is the almost paradoxical observation that biological fluids are supersaturated with respect to hydroxyapatite yet do not contain precipitate and the different soft tissues which they permeate remain unmineralised [12, 13, 14, 15].

At levels of supersaturation calcium phosphate is expected to precipitate out of solution. At physiological pH the precipitate does not directly form the most thermodynamically stable phase of HA, instead it passes through a number of different calcium phosphate phases of decreasing solubility. The process can be described by Ostwald's rule of stages which states that nucleation will occur with the formation of a critical

nucleus with the lowest possible surface energy and not the lowest energy phase [16]. This means that in general the least stable phase crystallises first and consequently increasingly stable crystalline phases form. During the precipitation of calcium phosphate in a supersaturated solution the first phase encountered is that of a highly unstable and short-lived amorphous calcium phosphate (ACP-1) which is followed by a second amorphous phase (ACP-2) [17]. Subsequently, crystalline octacalcium phosphate (OCP) is formed, which matures into the most stable hydroxyapatite phase, this process is represented in Figure 1.2 [14, 17, 18]. It is the HA mineral phase which is often deposited at sites of ectopic calcification and in sites of required calcification such as the collagen matrix [19, 20, 21, 22].

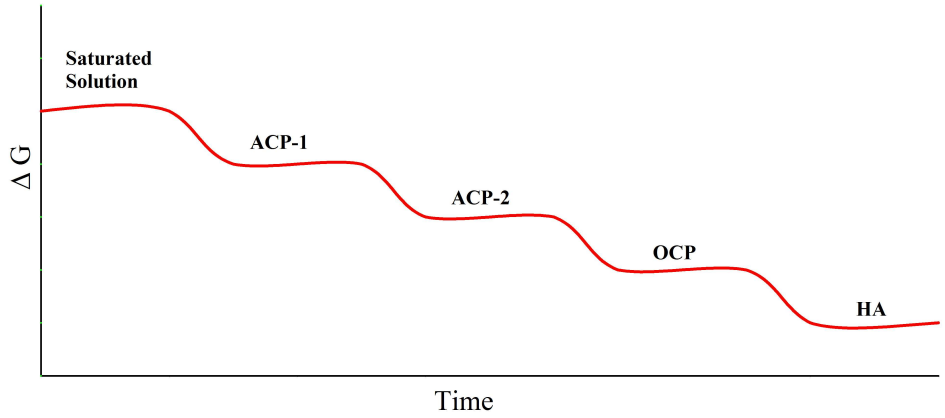

Figure 1.2: Schematic of the calcium phosphate phase separation leading to the formation of HA. The process proceeds via several transient intermediate crystalline and semi crystalline calcium phosphate phases. Based on an image taken from Holt *et al.* [23]

.

Biological fluids are complex aqueous suspensions that contain a range of different types of organic and inorganic constituents. Including biological macromolecules (proteins, lipids and carbohydrates), exogenous substances (e.g. drugs and microorganisms) and ionic species [9, 24]. This is because most biological fluids perform several functions. A good example of this is saliva the functions of which include facilitating the taste and nutrient content of food, maintaining tooth homoeostasis as well as

defending the host from bacteria and other exogenous pathogens [9].

As previously mentioned saliva also contains significant concentrations of different ionic species including calcium and phosphate, which are required to maintain the integrity of tooth enamel [9]. For an ionic species in solution there exists a physical limit of the amount that can be dissolved before the solution is considered saturated. Typically a saturated solution will form either a precipitate of solid or gas.

The saturation level of different biological fluids can be calculated with respect to different solid phases and it can be determined whether a solution is considered stable, metastable or unstable. The stability of a biological fluid can be determined by using thermodynamic principles, by calculation of the thermodynamic equilibrium constant. In solution, ionic compounds disassociate into their respective constituent species. In classical thermodynamics, an equilibrium can be established between a solid phase and the solution, in the case of calcium phosphate ions forming a solid HA phase:

$$\mathrm{Ca_5(PO_4)_3(OH)_{(s)} \rightleftharpoons OH^-_{(aq)} + 3\,{PO_4}^{3-}{}_{(aq)} + 5\,Ca^{2+}_{(aq)}}$$

The solubility of the formed solid phases is given by the solubility constant, K_S which can be obtained as a product of all ion activities in the solution. For HA the ion activity product is expressed as :

$$K_S = \left\{\mathrm{OH^-_{(aq)}}\right\}\left\{{\mathrm{PO_4}}^{3-}{}_{(aq)}\right\}^2\left\{\mathrm{Ca^{2+}_{(aq)}}\right\}^3$$

The brackets around the ions correspond to the activities of the ions, for ideal solutions these are equal to their concentrations, but for concentrated solutions like biofluids the ion activities are significantly smaller then their concentrations. When a solution is in equilibrium with the solid phase the ion activity value is invariant [10]. The ion activity product of a solid phase forming is dependent on the stochiometry, for example for a dicalcium phosphate the ion activity product is expressed as $\{Ca^{2+}\}^1 \cdot \left\{HPO_4^{2-}\right\}^1$, compared to that of a tricalcium phosphate, $\{Ca^{2+}\}^3 \cdot \left\{PO_4^{3-}\right\}^2$. For

these two examples the ion activities take the same form however their magnitudes differ due to the exponent, so that each calcium phosphate phase has a characteristic ion activity product. In a solid phase in equilibrium with a solution the ion activity does not change even when the composition of the solution changes. Therefore, if a invariant ion activity product is found in the solution, then the structure of the solid phase can be determined.

The calculation of ion equilibria in biological fluids is now well established [10], and has been applied to several biological fluids [10]. The model developed by Holt allows for the activities of each ionic species to be determined and from this the activity products of different crystalline phases can be obtained [10]. The state of equilibrium can then be determined by dividing the ion activity product (IAP) by the solubility product of the solid phase (K_S), yielding the saturation index defined as Log(S). If Log(S) > 1 the solution is supersaturated with respect to the solid phase and hence unstable. On the other hand if Log(S) < 1 the solution is undersaturated with respect to the solid phase and hence stable.

Thus by calculating Log(S) from experimental data on the composition and ionic strength of different bodily fluids it is straightforward to determine the level of supersaturation with respect to HA or any other mineral phase. Such calculations have been performed by Holt *et al.* [10] on the supersaturation of different bodily fluids with respect to HA. The results are shown in Table 1.1, they indicate that several biological fluids are supersaturated with respect to HA. In Figure 1.3 the supersaturation of different samples of milk calculated with respect to different solid phases of calcium phosphate are shown. It shows that milk is supersaturated with HA but undersaturated with respect to an amorphous calcium phosphate phase [10].

Unwanted mineralisation of soft tissues, known as ectopic calcification, is often detrimental to mammalian health [25, 26]. Ectopic calcification can occur in several different organs including the lungs, circulatory tissues and kidneys [1, 27, 28, 29, 30,

Table 1.1: Salt composition (mM), ionic strength, free ion concentrations (mM) and state of saturation with respect to calcium phosphate (Log(S)) of human artificial urine and ultrafiltrates of serum, saliva and milk (Table adapted from Holt *et al.* [10]).

	Serum	**Urine**	**Saliva**	**Milk**
Calcium	1.8	4.1	0.8	7.3
Magnesium	0.6	2.9	-	0.9
Sodium	144.9	148.4	31.6	2.2
Potassium	4.4	42.0	31.0	13.3
Phosphate	1	20.5	2.5	2.2
Chloride	151.8	177.4	58.0	24.1
Citrate	-	2.3	-	2.1
Ionic strength	150	187	65.6	35.3
pH	7.4	5.9	7.4	6.8
Ca^{2+}	1.1	1.8	0.5	4.7
Mg^{2+}	0.4	1.4	-	0.6
Log(S)	**7.74**	**3.77**	**7.85**	**10.02**

31, 32]. In humans one of the most prevalent types of ectopic calcification is that of the aortic valve, known as aortic stenosis [33, 34]. Calcification of the aortic valve leads to a reduction of the function [33, 35]. The risk factors and schematic progress of aortic calcification are shown in Figure 1.4 [36].

The likelihood of developing aortic stenosis increases with age [37], consequently the disease is especially prevalent in the elderly population. The known increased risk factor is caused by build up of calcium phosphate deposits within the aortic valve over prolonged time periods [38]. The impact of aortic stenosis on national health in western countries is increasing as the ageing population increases [39]. Aortic stenosis and other diseases of mineralisation are caused by a defect in the system which regulates biomineralisation. All ectopic calcifications are a result of the precipitation

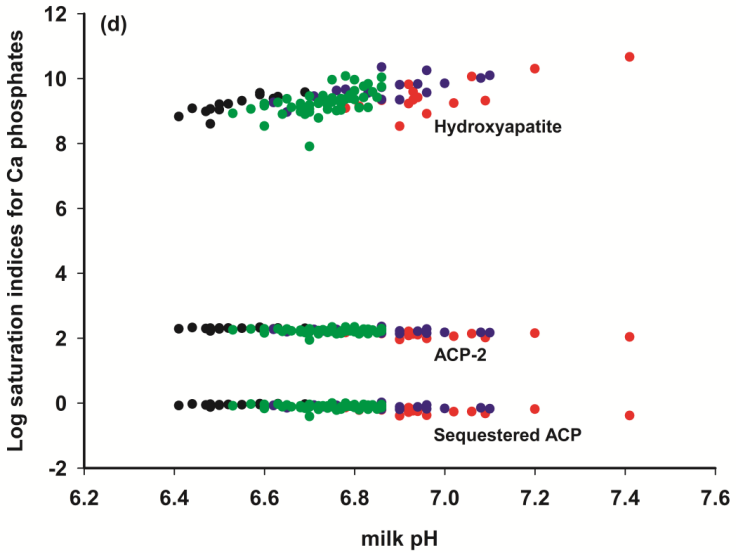

Figure 1.3: Calculated log(S) of various samples of bovine milk with respect to different calcium phosphate phases. Image taken from Holt *et al.* [10]

of calcium phosphate at unwanted locations within soft tissue. Despite this biofluids do not readily precipitate calcium phosphate, it is thought that inhibitors of precipitation such as magnesium ions, citrate and most powerfully secreted phosphoproteins regulate the supersaturation [23, 40]. Examples of inhibitory phosphopeptides include caseins in milk and osteopontin (OPN) which is present in all biofluids [23, 41]. A fundamental understanding of the regulatory system(s) involved is essential to support the design of new materials and therapies, and this a strong motivation for the work described in this thesis.

1.1.1 The role of phosphoproteins in biomineralisation

Depending upon the type and level of post-translational modifications present OPN can either inhibit or increase the rate of hydroxyapatite growth [42, 43, 44]. It is for example known that sequences of 8 or more phosphorylated residues, such as those found in the C-terminal half of OPN, can in fact promote the maturation of amorphous calcium phosphate [45].

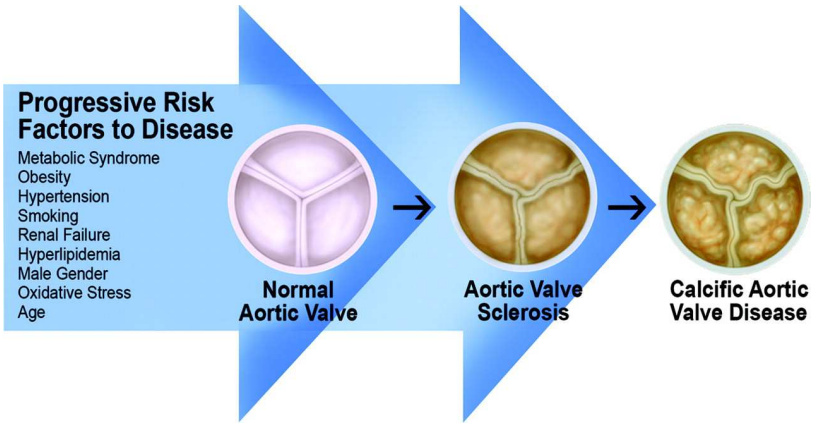

Figure 1.4: Aortic stenosis risk factors, known to increase the chance of developing aortic valve calcification. Image taken from Rajamannan *et al* [36]

Due to their levels of saturation with respect to hydroxyapatite, artificially created mimetic biological fluids with the same ionic strength and pH of those produced *in vivo* are highly unstable and readily precipitate calcium phosphate [46, 47, 48]. In order for soft and mineralised tissue to coexist a regulatory system must exist which inhibits ectopic calcification. One of the major regulatory roles involves of highly phosphorylated proteins and the control they exert on solutions supersaturated with respect to HA [23, 45, 49].

Osteopontin and Fetuin A, for example, can influence the rate of the calcium phosphate phase transformation, from the formation of a fresh precipitate of calcium phosphate all the way through to the formation of HA [50, 51, 52, 53, 54, 55]. Highly phosphorylated proteins including OPN and bone sialoprotein II are also capable of binding to crystalline hydroxyapatite through densely phosphorylated regions [52, 56, 57]. Caseins are another family of well studied phosphoproteins that have been shown to stabilise supersaturated milk by the sequestering high levels of calcium phosphate within a structure known as the casein micelle [15, 58].

Examples demonstrating the importance of phosphorylated proteins in the regulation of mineralisation can also be found in several other biological fluids. For example in urine and blood the widely expressed extracellular matrix proteins fetuin-A and OPN

have been shown to play important roles in many aspects of mineralisation including the inhibition of HA formation, the promotion of HA formation and the controlled mineralisation of collagen [59, 44, 57, 60, 61, 62, 63, 64]. Highly phosphorylated proteins such as Caseins, Fetuin-A and OPN can affect the time period of every step in the conversion of fresh precipitate of calcium phosphate to hydroxyapatite [43, 65, 66]. Highly phosphorylated regions are present in many proteins that regulate biomineralisation. If a protein containing such regions is present in sufficient molar concentration in a supersaturated solution then the calcium phosphate is sequestered within so called calcium phosphate nanoclusters. Small-angle X-ray and neutron scattering studies have shown that the formed CPNs are core-shell particles with a core of calcium phosphate surrounded by the sequestering protein [67]. The act of calcium phosphate sequestration prevents the formation of hydroxyapatite. Evidence is now accumulating which suggests that the sequestration of calcium phosphate is a general regulatory mechanism which stabilises supersaturated biological fluids [10, 68]. The most well studied case of calcium phosphate sequestration is that found in bovine milk. The function of milk is to facilitate the transport of large concentrations of calcium phosphate from the mother to the neonate through soft tissue [11, 69]. Consequently bovine milk is supersaturated with respect to HA, yet calcification of the mammary gland is rare [23]. About two thirds of the calcium and half the phosphate is associated with the casein family of proteins, which together form the larger casein micelle structure [70, 71]. The casein family of proteins can be separated into the calcium sensitive casein α, β and the non calcium sensitive κ-casein [71]. In milk calcium phosphate is sequestered by the calcium sensitive β-casein resulting in the formation of calcium phosphate nanoclusters (CPNs) [72, 67, 73]. The act of sequestration leads to the formation of CPN and the stabilisation of the supersaturated biofluid [67, 74]. β casein is a highly phosphorylated rheomorphoic or intrinsically disordered protein [75, 76], in bovine milk, β casein contains a sequence of 4 consecutive serines which

are phosphorylated [75, 76, 77, 78]. Holt *et al.* [45] described a number of examples of these highly phosphorylated regions and named them as phosphate centers (PCs). A PC is a short acidic sub-sequence in which three or more residues are phosphorylated. A typical example from caseins is the sequence of the single PC of bovine β casein: -E-pS-L-pS-pS-pS-E-E-, where pS is a phosphoseryl residue [45]. Other examples include sequences in OPN and other phosphoproteins from mineralised tissues including the protein orchestin [45]. The non-phosphorylated acidic residues in PCs and their flanking sequences may also influence the strength of binding. The effect of adding Asp residues to a PC was explored by recombinant methods applied to casein-like sequences [79]. Non-PC less highly phosphorylated sequences and unphosphorylated acidic sequences can, for example, bind to crystalline calcium phosphates or inhibit the maturation of ACP [52] but they have not, so far, been demonstrated to form the equilibrium nanocluster complexes. Amino acid sequences can only be considered phosphate centres if three or more of the potential sites of phosphorylation are phosphorylated. It is through these highly phosphorylated regions that β-caseins interact with calcium phosphate and facilitates the formation of calcium phosphate nanoclusters [67, 74, 23]. Other phosphate centre containing proteins including OPN have also been shown to be able to sequester calcium phosphate in a similar manner [79, 45]. In order to understand the act of calcium phosphate sequestration a thermodynamic description has been developed [72, 74].

1.1.2 A thermodynamic model of calcium phosphate sequestration

The formation of calcium phosphate nanoclusters can be understood in terms of a thermodynamic model of calcium phosphate sequestration. Formation of stable equilibrium core-shell nanoparticles can be divided into two parts, the energy of formation of the core and the sequestration of ACP by the protein shell. The chemical potential of a system is defined by classical thermodynamics as a form of potential energy that

is released or absorbed during a phase transition. In solution the variation of chemical potential is expressed by relating the quantities with the concentration. Classical thermodynamics allows the calculation of the chemical potential of a species (μ_i) from the activity of the species in solution (a_i) so that:

$$\mu_i = \mu_i^{std} + RT \ln a_i \tag{1.1}$$

where μ_i^{std} is the potential in a standard state, R is the Boltzmann constant and T the temperature. In the case of calcium phosphate nanoclusters there exists a chemical potential for both the protein monomer (μ_m) and the nanocluster form (μ_{CPN}).

$$\mu_m = \mu_m^{std} + RT \ln a_m \tag{1.2}$$

At equilibrium $\mu_m = \mu_i$ then the Gibbs free energy is $\Delta G_i^{\cdot} = \mu_i - \mu_m$ then if K_i is the equilibrium constant for the nanocluster formation then:

$$\Delta G_i^{\cdot} = -RT \ln \left(\frac{a_i}{a_1} \right) = -RT \ln K_i \tag{1.3}$$

The chemical potential of the calcium phosphate bulk phase is equal to the chemical potential of the monomer in a saturated solution:

$$\Delta G_{core} = RT \ln a_s \tag{1.4}$$

The Gibbs free energy of formation of an equilibrium complex between a core of calcium phosphate and a sequestering phosphopeptide (ΔG_{cpn}), can be divided into the free energy of formation of the calcium phospahte core (ΔG_{core}) and the free energy of sequestration of the shell of phosphopeptide (ΔG_{seq}), to form the resulting core-shell particle. For a complex containing i moles of calcium phosphate sequestered by a shell containing j moles of phosphopeptide:

$$\Delta G_{cpn} = i\Delta G_{core} + j\Delta G_{seq} \tag{1.5}$$

The CPN is then assumed to have a spherical structure, with the number of phosphopeptides proportional to the surface of a sphere:

$$\Delta G_{cpn} = i\Delta G_{core} + ki^{2/3}\Delta G_{seq}/A \tag{1.6}$$

where $k = (36\pi V_{acp}^2)^{1/3}$, A is the core surface area occupied by the sequestering phosphopeptide, V_{acp} is the molar volume of ACP. The average modal core radius (r_{core}) is then given by:

$$r_{core} \approx \left(\frac{2k\gamma}{3RTln\left(\frac{a_1}{a_s}\right)}\right)\left(\frac{3V_{core}}{4\pi}\right)^{\frac{1}{3}} \tag{1.7}$$

where γ is the surface energy of the protein shell. Replacing the Gibbs free energy of the core formation with the activity of ACP in solution (a_1) containing the complex (a_s) saturated with respect to the bulk ACP phase, is then:

$$r_{core} \approx \left(\frac{2k\Delta G_{seq}}{-3A_{core}\Delta G_{core}}\right)\left(\frac{3V_{core}}{4\pi}\right)^{\frac{1}{3}} \tag{1.8}$$

If there is a reaction between the phospopeptide and the core surface then ΔG_{seq} will be negative and hence $a_1 < a_s$ in which case the surface energy is negative and the solution is undersaturated and hence stable. The above equation shows that the radius of a calcium phosphate core varies inversely with the Gibbs free energy of the calcium phosphate core, (ΔG_{core}). A crystalline core would have a much higher lattice energy compared to an amorphous core, therefore decreasing the likelihood that the core is crystalline.

Another condition for the formation of a stable calcium phosphate nanocluster is an stoichiometric excess of phosphate centres, the fraction of reacted phosphate centres

(α) required for a stable solution is then [72]:

$$\alpha = \frac{[p]}{f[pp]R_p} \tag{1.9}$$

Where $[p]$ is the molar concentration of phosphate, $[pp]$ is the molar concentration of phosphorylated peptides containing f phosphate centres and R_p is the stoichiometric ratio from the empirical formula of a calcium phosphate nanocluster which is given by:

$$Ca_{Rca}H_{Rh}(Pi)_{Rp}(H_2O)_{Rw}R{-}PC$$

For a solution to be stable α must be equal to or less then 1. Above an α value of 1 the solution can remain metastable but over time a precipitate will form.
Based on the model, calculations can be performed on biological fluids to determine the region of stability and the molar concentration of phosphate centers required to stabilise the solution. The calculation is explained in detail by Holt [72], and requires certain parameters shown in Table 1.2

Table 1.2: Parameters required for the calculation of stability of different biofluids, table adapted from Lenton *et al.* [68].

Parameter	**Description**
C_1, C_2, C_n	Concentration of the n principal salts
R_{C1}, R_{C2}, R_{Cn}	Molar ratio of each salt constituent to PC in the complex
$[PC]$	Concentration of phosphate centres
y	acidity of the calcium phosphate
K_s	Solubility constant
v(C_1, C_2, C_n)	Isotherm describing the affinity of the salt to the PC

The theory of nanocluster formation and their ability to stabilise biological fluids has been tested experimentally [10]. Figure 1.5 shows that there is a close agreement

between the predicted stability of a synthetic biofluids and the observed stability. Showing that sufficient concentration of phosphate centre containing proteins can stabilise synthetic biofluids, phosphate centre concentration is in stoichiometric excess of the calcium phosphate ions.

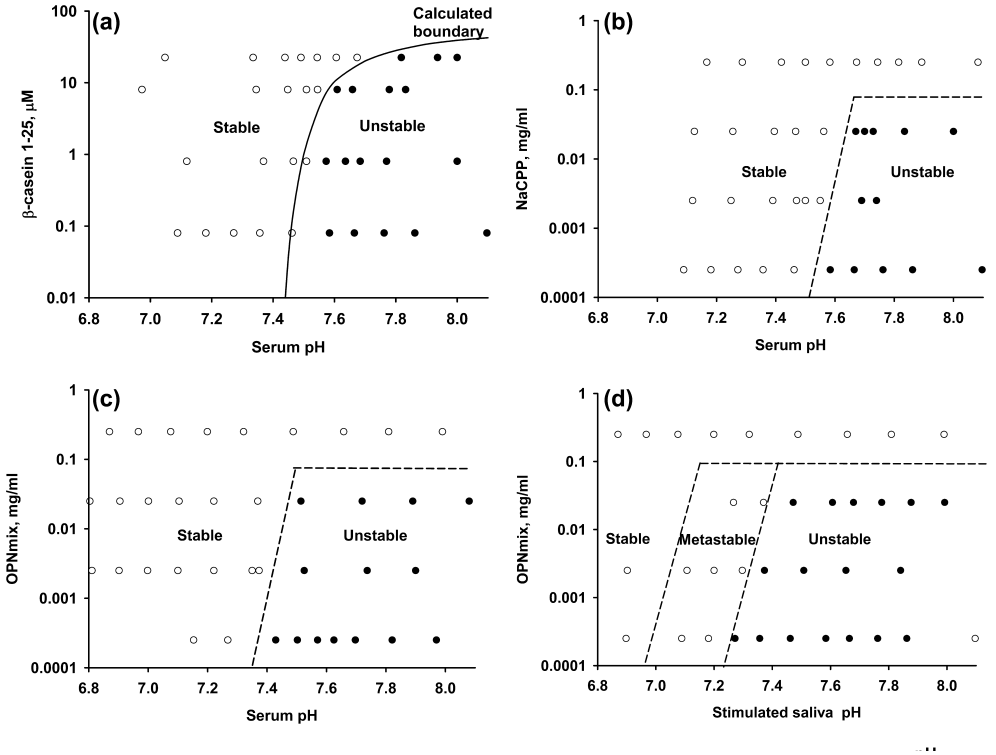

Figure 1.5: Stability diagrams of artificial biofluids, using different phosphate centre containing proteins. The calculated border of stability is indicated by a dashed line, the circles indicate synthetic mimetic biofluids. Filled circles are samples which formed precipitate. **(A)** Stabilisation of synthetic serum by β-casein isolated from milk. **(B)** Stabilisation of serum by NaCPP another phosphopeptide. **(C)** Stabilisation of serum by OPNmix. **(D)** Stabilisation of saliva by OPNmix. Image taken from Holt *et al.* [10].

1.2 The structure and function of osteopontin

Osteopontin (OPN) is a highly phosphorylated sialoprotein expressed in numerous different bodily tissues including brain, kidney and mineralised tissue [80]. In contrast to OPN function in biofluids, the main function of OPN in the brain and kidney is that of a signalling molecule [80]. Significant concentrations of OPN have been

found in all bodily fluids where it has been sought including in milk and the extracellular matrix of bone and teeth [81, 82, 83]. OPN is a member of the small integrin binding ligand, N-linked glycoproteins (SIBLING) family of proteins [84]. This group of proteins share several common features including a large degree of post-translational modification and a high level of flexibility in the protein structure [84]. Characteristically of SIBLING proteins OPN is extensively modified by post-translational modifications including phosphorylation and glycosylation, resulting in a distribution of different OPN peptides with varying functional groups and molecular masses [84, 85, 59, 86, 87]. Post-translational modifications of OPN are often tissue dependent and allow OPN to mediate many different cellular and biological processes *in vivo* [88]. Procceses regulated by post-translational modification of OPN include bone remodelling [89], cell signalling [90] and regulation of the inflammation response [91]. There are abundant examples of the functional relevance of phosphorylation in OPN [88]. The process of bone remodelling and growth are greatly affected by the levels of phosphorylation of OPN. For example OPN promotes the attachment of osteoclasts provided that the protein is phosphorylated [92], a condition that is also essential to the ability in inhibiting hydroxyapatite formation [59, 44]. Likewise in urine and soft tissues phosphorylation of OPN is required to inhibit crystallisation of calcium phosphate [93, 94, 95]. The kinases responsible for phosphorylating secreted osteopontin have been identified as Casein kinase IIα and FAM20C [96, 97, 98].

In humans OPN is expressed as a 300 amino acid sequence but matures into a 264 residue chain with a molecular mass of 42.9 kDa [86]. The mature human sequence contains 36 sites of phosphorylation and 5 sites of glycosylation, compared with 28 and 3 respectively for bovine OPN (Figure 1.6) [99, 86]. OPN isolated from bovine milk is known to bind calcium in several different locations in the sequence in a non phosphorylation dependent manner, with a slight increase in the association constant observed upon phosphorylation of OPN [100]. This is thought to be due to the

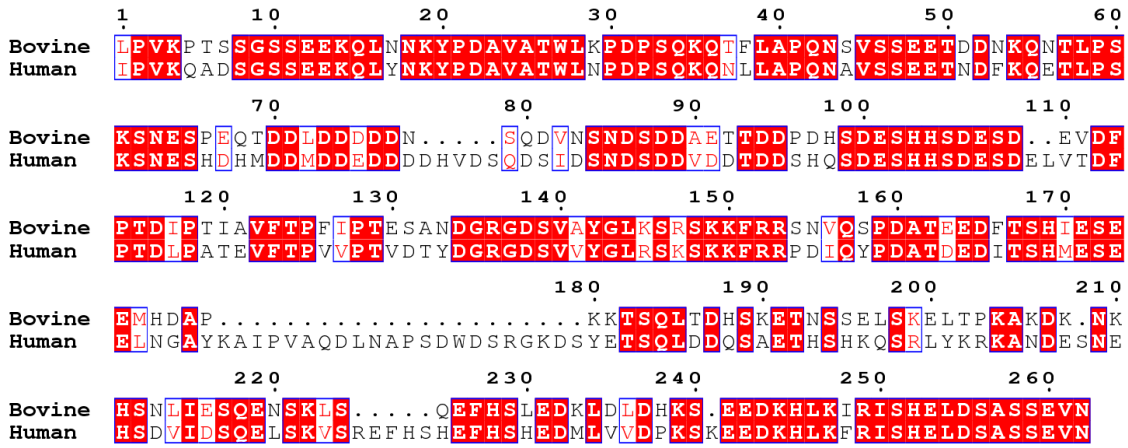

Figure 1.6: Sequence alignment of human and bovine OPN. Amino acid residues highlighted in red are conserved accross the species.

higher charge charge interaction introduced by the phosphorylated residues [100]. Upon going down the Hofmeister series the affinity of OPN to different ions decreases significantly proving that the association is a charge charge interaction and not due to conformational space requirements [100]. The calcium binding locations in bovine OPN have been identified, and are shown in Figure 1.7.

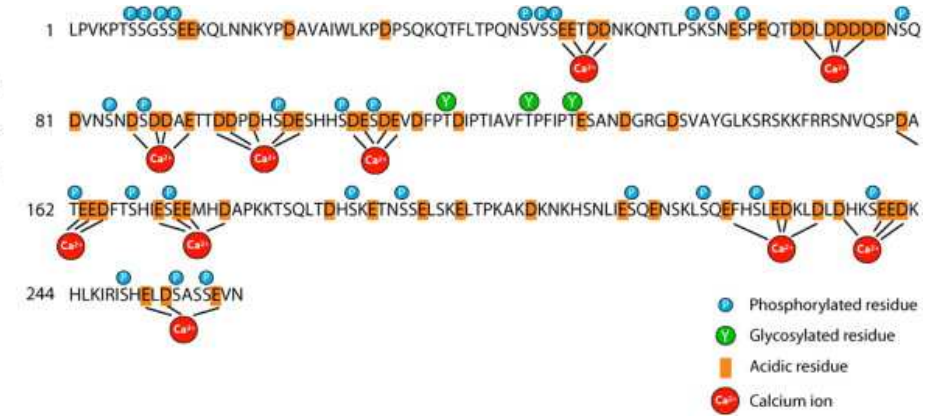

Figure 1.7: Sequence of full length bovine OPN with the 28 phosphorylation sites indicated (blue circles) and acidic residues (yellow). Calcium binding sites proposed based on experimental data are indicated by red circles representing Ca^{2+} ions. Image taken from Kläning *et al.* [100].

The sequence of OPN is highly conserved across mammalian species (for example see Figure 1.6) in particular regions containing motifs of serines and threonines [99]. The

structures of OPN from different species have been characterised using several different techniques. NMR studies of quail OPN determined a highly flexible structure of the protein and addition of calcium ions to the buffer did not induce structural changes [101]. Other studies completed by one dimensional NMR on human OPN from recombinant sources showed a similar high flexibility and lack of secondary structure elements [84]. It is now generally accepted that the lack of a folded or globular structure, along with the many post-translational modifications it undergoes, are central to the functional diversity and specificity of OPN [50]. Many proteins involved in the regulation and formation of biominerals in mammals are either known to have or are predicted to have a large degree of disorder [102, 103]. These include the casein family of proteins and more recently OPN. These proteins are capable of sequestering calcium phosphate thereby forming stable calcium phosphate nanoclusters which prevent supersaturated biofluids from precipitating. It is believed that intrinsic protein disorder is vital for the formation of these nanoclusters and hence the stabilisation of many different biological fluids.

1.3 Intrinsically disordered proteins

It was previously thought that proteins require a well-defined three dimensional structure to carry out their biological function, the so called structure function paradigm [104]. In recent years, however, it has become increasingly obvious that there exists a large group of proteins that are either completely disordered or contain large domains of disorder, and that therefore well defined three-dimensional structures are not necessarily required for protein function [104]. Figure 1.8 schematically highlights this dichotomy.

Disordered proteins or disordered protein regions are defined as dynamic ensembles in which atomic positions and Ramachandran angles vary significantly over time without

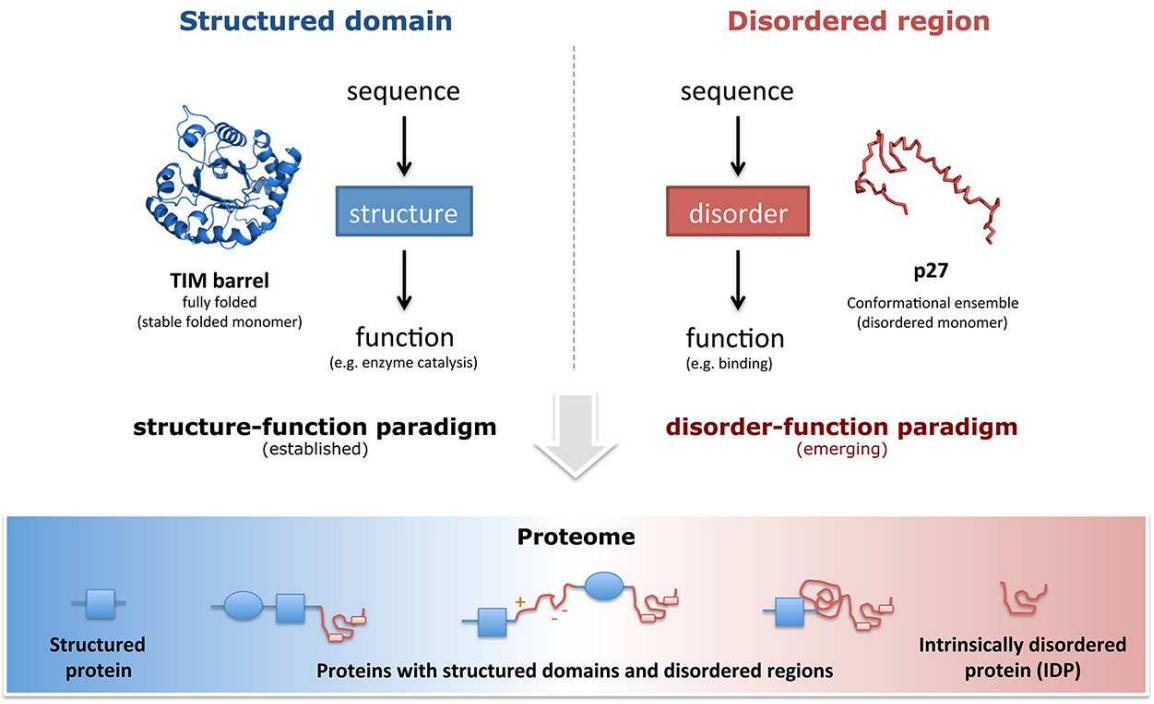

Figure 1.8: Structured regions and disordered regions are fundamental protein building blocks. In recent years the idea of a disorder-function paradigm has become accepted. Many proteins consist of a combination of both structured and unstructured regions and can be classified by their amount of disorder. Image taken from Van der Lee *et al.* [105].

a well defined equilibrium value, in contrast to folded proteins which only have small amplitude thermally driven motions about an equilibrium position [105]. In recent years it has become clear that intrinsically disordered proteins (IDPs) are involved in many key biological regulatory processes including regulation of the cell cycle and signal transduction [106]. It is now known that around 40% of the protein coding genes in the human genome contain disordered regions of at least 30 amino acid residues [107]. The traditional structure function paradigm states that proteins fold into a well defined three-dimensional structure due to the preferred thermodynamic stability. The collapsed globular state of a folded protein is stabilised by several different forces, including the hydrophobicity of the amino acid residues in the core of the globular protein and the formation of disulfide bridges. It is now clear that proteins can contain structured or disordered regions and that both of these fundamentally different types of protein regions must be considered individually to study function.

IDPs contain a different distribution of amino acid composition compared to globular proteins, often including large stretches of low complexity amino acid residues [108]. Low complexity regions are defined as those containing little diversity in their amino acid composition and often contain repeats of the same amino acid residues. The evolution of largely disordered proteins differs from that of their ordered counterparts mainly due to the lack of structural constraints within the amino acid sequence [109] resulting in a more rapid evolution of IDPs compared to globular proteins [110]. An example of this is the linker region in a unstructured linker domain in RAP70 [111] where the amino acid sequence conservation is low yet the disordered behaviour is conserved across species. Due to their low sequence complexity, lack of hydrophobic amino acids and tendency to be enriched in charged and polar amino acids it is possible to analyse sequences and predict the tendency for protein disorder [112, 113]. A number of different disorder predictors now exist based on several different techniques. *Ab initio* methods are one of the more popular methods used as they rely only on the primary amino acid sequence of the protein [114]. For example IUpred [115] predicts the disorder tendency of an amino acid sequence based on the capacity of the sequence to form inter-residue interactions to stabilise the globular state [115].

1.3.1 The energy landscape of IDPs

A protein in solution samples a variety of conformations based on thermal fluctuations which can be described by the dependence of free energy on the amino acid configuration in three dimensional space, see Figure 1.9. Folded proteins are characterised by a well defined energy minimum within which the secondary and tertiary structures of the globular protein are located. IDPs on the other hand, do not adopt a well defined tertiary structure, instead the structure of the protein fluctuates within a conformational ensemble [116]. The free energy landscape of IDPs is therefore best described by a more featureless energy landscape with several local energy minima

separated by low energy barriers (Figure 1.9) [117, 118]. Changes to the environmental conditions may therefore have a great effect on the conformational ensemble of an IDP present in solution. For example Figure 1.9 C shows the effects of increasing amounts of folding upon binding of an IDP to the sampled energy landscape [119].

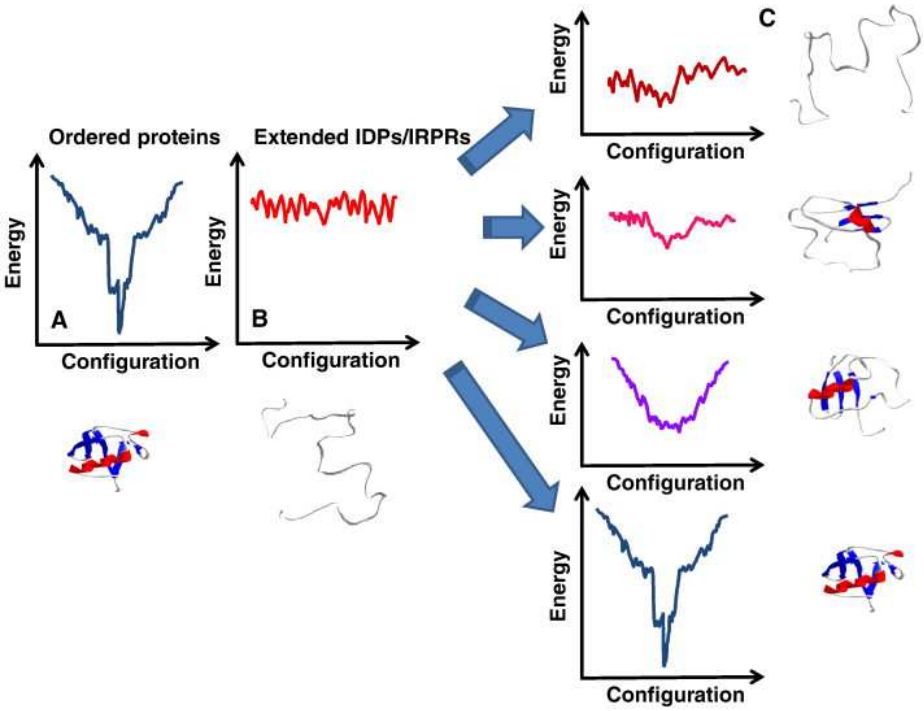

Figure 1.9: Energy landscapes depicted for **(A)** A globular protein. **(B)** A IDP and **(C)** Different degrees of folding upon binding of an IDP. Image taken from [119].

1.3.2 Post-translational modifications of IDPs

A major characteristic of IDPs is their increased tendency to undergo post-translational modifications (PTMs) compared to globular proteins [120, 121]. Computational predictions have shown that there is a strong link between the amount of predicted protein disorder and the extent of PTMs [122, 123]. PTMs are classified as chemical changes inflicted upon a peptide chain after translation. Currently around 300 different types of post-translational modifications are known to occur under physiological conditions [124]. The different types of PTMs range from the addition of chemical

groups, such as phosphates and carbohydrates, to the digestion of the amino acid chain by proteolytic enzymes [124, 125]. The extent and amount of post-translational modifications greatly affect the structural and functional properties of proteins, including the stability and localisation in the cell as well as the affinity of proteins to their ligands [126]. The nature of and amount of PTMs effecting different IDPs are often highly cell specific which allows for tuning of the protein function for specific cellular processes [88]. The reversible nature of certain PTMs (e.g. phosphorylation and acteylation) provides a regulatory switch turning on and off specific biological processes, this type of regulation is often found in signalling pathways [127]. For example the circadian clock is highly regulated by different types of kinase and phosphatase enzymes [128]. The higher rate of PTMs found in IDPs or IDP regions is thought to be in part due to their increased flexibility which allow for easy recognition and consequent modification by PTM enzymes, as more of the amino acid chain is accessible to both solvent and the enzymes that undertake PTMs [122]. The rapid turning on and of off IDP function by PTM enzymes makes them ideal for regulating cell signalling resulting in a high degree of IDP participation in these processes [129, 130].

1.3.3 Mechanisms of IDP complex formation

One of the key features of IDPs is their ability to bind many different types of proteins or substrates, an ability which seems to be unique to IDPs and as mentioned previously influenced by the amount and type of PTM [131]. IDPs that are capable of interacting with many binding partners are known as hub proteins [132, 131]. The conformational freedom of IDPs allows them to adopt different conformational structures when bound to their partners which are often highly diverse [133]. The result is a range of different types of complexes with a wide range of structural and dynamical features. The following is a non-exhaustive list of some of the key mechanisms that

underpin IDP interactions with their partners. They can be broadly separated into static disorder based complexes and dynamic disorder based complexes [134].

1.3.4 Static-disorder based complexes

This type of complex is based on the recognition and binding of a specific three dimensional conformation of an IDP by an often globular partner. The formed complexes are referred to as static because they can be described by a well defined three dimensional shape. The two main types of static-disorder based complexes are those formed by conformational selection and those formed by binding induced folding.

Conformational selection states that the intrinsically disordered substrate protein is fluctuating through its heterogeneous ensemble in solution. The binding partner selects one or several of the compatable heterogeneous conformations thereby shifting the heterogeneous population towards a selected conformational state [135, 136]. The model therefore suggests that the heterogeneous ensemble of structural conformations contains structures that resemble the bound state [137]. In the case of IDPs this allows one disordered protein to be a binding partner for multiple proteins, each selecting for a different structural conformation. Distributions of the heterogeneous ensemble of conformations not only depends upon the amino acid sequence but also external effects such as post-translational modifications which are capable of shifting the structural conformational ensemble towards one which is more favourable to binding [138, 139].

Classified as a disorder to order mechanism binding-induced folding states that folding of certain IDPs takes place upon partner binding [140]. The first step in binding-induced folding is the recognition of a partner through a certain motif which is followed by folding of the IDP partner [141, 142]. It is believed that by adopting an unfolded conformation the protein has a larger capture radius allowing it to more rapidly interact with the partner, but its biological function is completed once bound in the folded state.

1.3.5 Dynamic-disorder based complexes

Several IDPs contain regions within which their structural heterogeneity is conserved upon complex formation. Increasing the rigidity or removing these regions affects the binding affinity and specificity [143]. The maintenance of structural disorder upon complex formation is termed fuzziness and the complexes formed are known as fuzzy complexes. The maintenance of conformational heterogeneity in fuzzy complexes is of critical importance and often the disordered regions provide sites for other proteins to bind or post-translational modifications to occur [143].

1.3.6 Intrinsic disorder, osteopontin and biomineralisation

Intrinsically disordered proteins are proteins that do not feature any significant secondary or tertiary structural elements. Therefore their crystal structure can not be determined. However, the function of intrinsically disordered proteins may be related to their structure in a similar manner to globular proteins. For example, all proteins shown to sequester calcium phosphate thereby forming calcium phosphate nanoclusters are shown to be intrinsically disordered. The ability to sample a large area of conformational space and rapidly recognise binding partners is a major benefit of adopting an intrinsically disordered structure. The efficient sampling of a large area of conformational space also allows intrinsically disordered proteins to recognise and bind a large subset of binding partners.

1.4 Scope of this thesis

In this thesis some of the questions of the role of intrinsic disorder in the sequestration of calcium phosphate are answered using a range of different techniques. The heterogeneous conformations of OPN1-149 in solution are determined by SAXS com-

bined with the EOM method. The rigidity of OPN1-149 is then determined by elastic incoherent neutron scattering and related to the heterogeneous structural ensemble of OPN1-149. The effects of sequestering calcium phosphate on the rigidity is also determined by elastic incoherent neutron scattering, providing evidence that IDPs can maintain a disordered structure but acquire different dynamical properties. The core-shell nature of calcium phosphate nanoclusters stabilised by OPN1-149 is confirmed by small-angle neutron scattering in combination with contrast variation. The core of the calcium phosphate nanocluster is determined to have amorphous properties by neutron and X-ray diffraction. That the core is amorphous is important for the biological function of the formed calcium phosphate nanoclusters, allowing them to act as reservoirs of calcium phosphate for hard tissue mineralisation. Furthermore, human osteopontin is prepared using recombinant methods in order to determine the effects of phosphorylation on the protein structure and dynamics. The results provide further insight into the function of this regulatory protein *in vivo*.

Chapter 2

Methods - theoretical background

The work within this thesis utilises a range of biophysical techniques to characterise the structure and dynamics of different osteopontin peptides. This chapter aims to provide a general, non-exhaustive, introduction to several of the key techniques employed. More detail about the experiments performed are given in the corresponding results chapters.

2.1 Properties of the neutron

The neutron is a nuclear particle that decays into a proton, an electron and an electron antineutrino. A neutron has no net negative charge and a slightly higher mass then a proton, some of the fundamental properties of the neutron are shown in Table 2.1 and compared against those of the proton.

Neutrons can be described as waves with the wavenumber, K, defined as

$$K = \frac{2\pi}{\lambda} \tag{2.1}$$

Where λ is the neutron wavelength. The momentum of a neutron can be calculated based on it's mass m_n and it's velocity v by:

Table 2.1: Properties of the neutron and the proton.

	Neutron	**Proton**
Mass	$1.675 \cdot 10^{-27}$ kg	$1.672 \cdot 10^{-27}$ kg
Charge	0	1
Spin	1/2	1/2
Magnetic moment	$0.966 \cdot 10^{26} J \cdot T^1$	$1.410 \cdot 10^{26} J \cdot T^1$

$$p = m_n v = \hbar K \tag{2.2}$$

Where $\hbar$ is Planck's constant and v the neutron velocity. The kinetic energy (E) of the neutron can then be calculated by:

$$E = (\hbar^2 / 2m_n) K^2 \tag{2.3}$$

The wavelength of a neutron is thus dependent on it's energy, which in turn depends upon temperature:

$$\lambda = \frac{\hbar}{m_n v} \tag{2.4}$$

$$v = \frac{E}{\hbar} \tag{2.5}$$

$$E = K_B T \tag{2.6}$$

where K_B is the Boltzmann constant and T the temperature. In order to study matter neutrons with similar wavelengths to interatomic distances are required, cor-

responding to energies on the order of meV. Neutrons emerging from either fission or spallation sources have energies of an order of MeV. In order to produce neutrons of lower energy the produced neutrons are slowed down by a moderator. A moderator lowers the neutron energy by thermal equilibration typically by inelastic collisions with light atoms. The resulting neutrons are characterised by their temperature and corresponding wavelengths shown in Table 2.2. At the Institut Laue-Langevin neutrons are moderated to thermal energies using liquid D_2O to cold temperatures.

Table 2.2: Neutron energies and corresponding properties.

	Ultra cold	**Cold**	**Thermal**	**Epi-thermal**
Energy (E)	0.25 μeV	1 meV	25 meV	1000 meV
Temperature (K)	3 mK	12 K	290 K	1000 K
Wavelength(λ)	57 nm	0.9 nm	0.18 nm	0.029 nm
Velocity (v)	6.9 ms^{-1}	440 ms^{-1}	2200 ms^{-1}	1400 ms^{-1}

2.2 The neutron scattering process

Neutrons interact with the atomic nuclei of many atoms in the sample. The scattering per atom is given by the double differential cross section:

$$\frac{d^2\sigma}{d\Omega dE_{out}} = \frac{\sigma_s}{4\pi\hbar} \cdot \frac{K_{in}}{K_{out}} \cdot S(q,\omega) \tag{2.7}$$

where K_{in} and K_{out} are the incident beam and scattered beam, respectively, $S(q,\omega)$ is the scattering cross section and $\hbar$ is Planck's constant. The difference, between the incident and scattered beam is then defined as the scattering vector $\vec{q}$, given by:

$$\vec{q} = K_{out} - K_{in} \tag{2.8}$$

This process is represented schematically in Figure 2.1 .

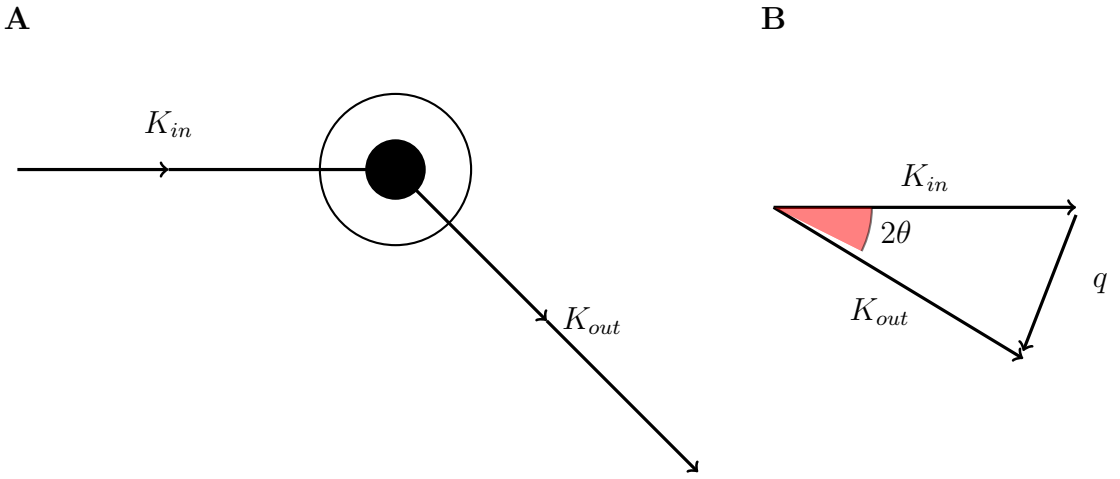

Figure 2.1: (A) The scattering process is described by two vectors the incident beam, K_{in} and the scattered beam K_{out} **(B)** The momentum transfer (q) is defined $K_{in} - K_{out}$

The total scattering cross section, $\sigma_{tot}S(q,\omega)$, contains coherent and incoherent contributions:

$$\sigma_{tot}S(q,\omega) = [\sigma_{coh}S_{coh}(q,\omega) + \sigma_{inc}S_{inc}(q,\omega)] \tag{2.9}$$

where $S_{coh}(q,\omega)$ and $S_{inc}(q,\omega)$ are the coherent and incoherent scattering functions and σ_{coh} and σ_{inc} are the coherent and incoherent scattering cross sections. The interaction of a neutron with the atomic nuclei of a sample can result in several different processes including diffraction, inelastic scattering and coherent inelastic scattering.

During a diffraction type experiment the angular dependence of the scattered intensity, defined as $S(q)$, is measured:

$$S(q) = \int S(q, \omega) d\omega \tag{2.10}$$

A diffraction measurement reflects the spatial arrangement of atoms in the sample and the scattering profile depends upon this arrangement of atoms in 3-dimensional space. In a crystalline material diffraction gives rise to Bragg peaks, in a semi crystalline material the scattering is more diffuse and if the sample is in solution as is the case for small-angle scattering experiments the resulting pattern is further diffused, examples are shown in Figure 2.2.

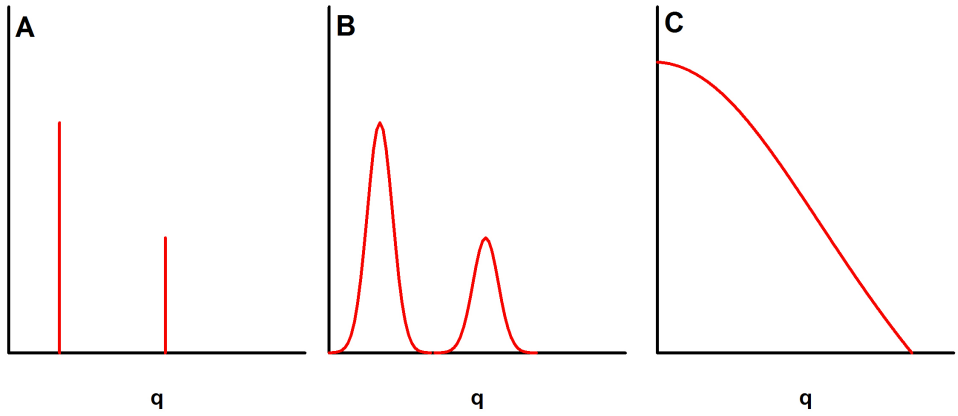

Figure 2.2: Scattering function shown for three different types of sample. **(A)** Bragg diffraction observed for crystalline material. **(B)** Diffuse scattering observed for semicrystalline material. **(C)** Small-angle scattering observed for a heterogeneous non crystalline solution.

During a small-angle scattering experiment the angular dependency of the scattered intensity ($S(q)$) is measured, which reflects the spatial rearrangement, ($p(r)$), of the atoms. To remain independent of the wavelength (λ) of the incident radiation the scattering patterns are usually represented as a function of q, defined as:

$$q = \frac{4\pi}{\lambda} \cdot sin(\theta) \tag{2.11}$$

where the scattering angle between the incident beam and the recorded intensity is 2θ.

2.3 Small-angle scattering

Measurement of the scattered intensity as a function of scattered angle at small angles can provide information about the structure of a particle in solution at low resolution. This technique is known as small-angle scattering. The information provided by small-angle scattering depends on the sample studied and can range on structural information of individual particles to interactions between them. The exact length scale studied depends upon the wavelength of the incident beam and the angle measured. Both neutrons and X-rays can be produced at the appropriate range of wavelengths, and the underlying principles for the analysis of the scattering data are similar. A schematic diagram of a small-angle scattering experiment is shown in Figure 2.3. During a basic SAS experiment a solution containing the biological sample is inserted into a beam of monochromatic radiation (neutrons or X-rays) and the intensity of the scattered bean is recorded as a function of the scattered angle. An identical measurement is then performed on the buffer in which the sample was measured and is subtracted from sample scattering, the difference in signal arising from scattering of the biological macro-molecule this process is represented schematically in Figure 2.3. In order to determine the structural properties of biological macromolecules dilute solutions have to be measured where interparticle interactions are absent. This concentration normally lies around a few mg/ml. Typical protein samples measured by SAS are two phase systems, with scattering contributions from the macromolecule and from the solvent. The difference between the scattering of the sample and it's solvent is known as contrast. One of the major differences between SAXS and SANS

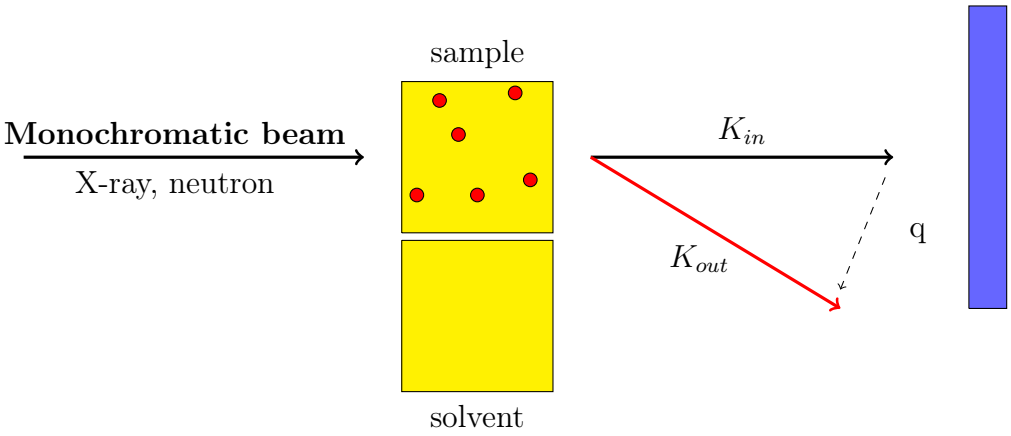

Figure 2.3: Schematic layout of a basic SAS experiment, the momentum transfer q is defined as the incident beam (K_{in}) minus the scattered beam (K_{out}). The detector is represented by the blue bar.

is the cause of contrast. For small-angle X-ray scattering the contrast arises due to the difference in electron density between the sample and the solvent. If the sample to be measured ($M1$) is present in a solute ($M2$) then the contrast (Dr) is defined as:

$$Dr = M1 - M2 \tag{2.12}$$

Because biological macromolecules are mainly constructed of lighter atoms this difference can be quite small, therefore experiments are often performed at synchrotron radiation sources which provide a X-ray beam with high intensity and brightness.
For a SANS experiment depends upon the difference between the scattering length of the sample ($p_{(s)}$) and solvent (p^o). So that the contrast ($S(q)$) between the sample and solvent is defined as:

$$S(q) = p_{(s)} - p^o \tag{2.13}$$

During a small-angle neutron scattering experiment the scattered intensity of the incident beam depends upon the excess neutron scattering length of the atomic constituents. Unlike electron density which increases with increasing molecular mass the neutron scattering length is dependent on nuclear properties (spin, energy levels) and there is no linear dependency of molecular mass, even isotopes of the same atom can have different neutron scattering lengths (Figure 2.4).

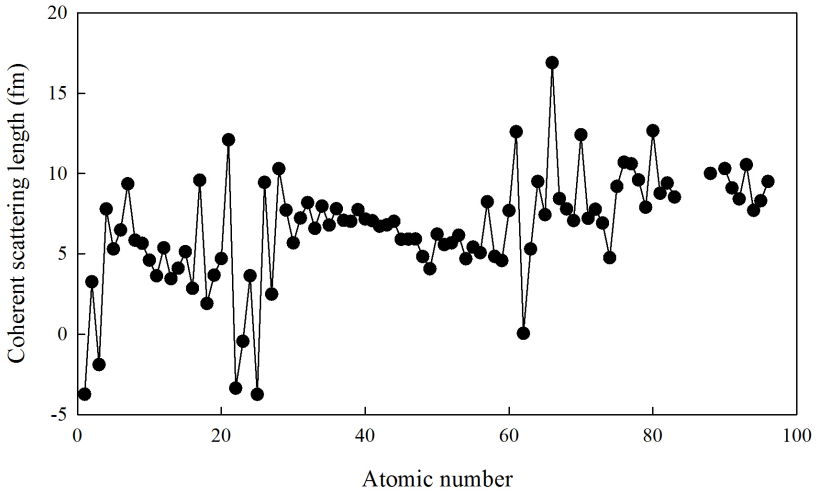

Figure 2.4: The variation of the coherent scattering length does not increase linearly with atomic mass.

A useful feature of neutron scattering is the ability to alter the contrast of the solvent using isotope replacement. For example deuterium has a neutron scattering length of $6.671fm$ compared to $-3.742fm$ of hydrogen. Therefore it is possible to change the scattering length of the solvent and match out scattering of different constituents of the sample. This phenomenon of varying the percentage of deuterium in the solvent is known as contrast variation and is shown schematically in Figure 2.5.

2.3.1 Model independent analysis of small-angle scattering data

As previously mentioned during a SAS experiment the intensity of the scattered beam is measured as a function of the angle ($I(q)$). The recorded spectra contains

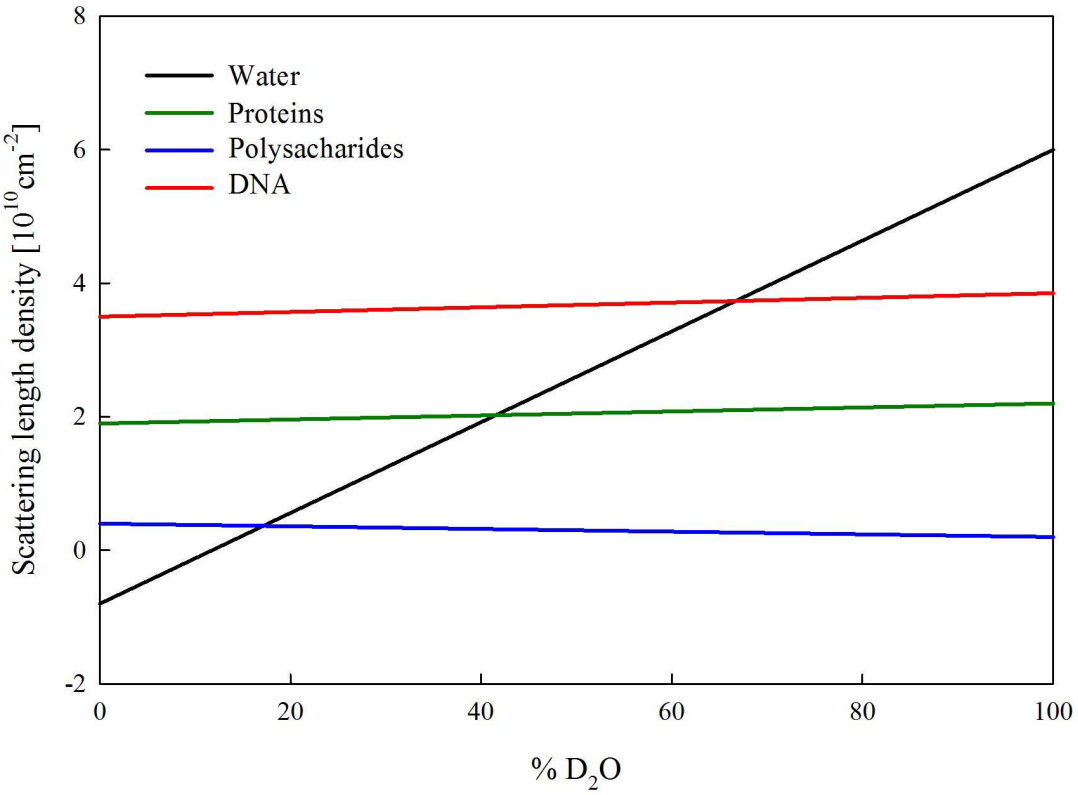

Figure 2.5: Average neutron scattering length densities for molecules commonly present in biological molecules as a function of the percentage of D_2O in solution. By altering the concentration of deuterium in the solvent it is possible to match-out contributions from different samples in the solution. For example at 40% D_2O contributions from protein scattering are matched out.

information about the size and shape of the particle on the size scale determined by the measured q-range. Several model independent characteristics of the scattering sample can be determined from the scattering spectra including the molecular mass, the radius of gyration (R_G) and the maximum particle diameter (D_{max}). The most straightforward model independent data analysis is the Guinier analysis developed by Guinier in the 1930s [144]. The Guinier analysis allows for the extraction of the (R_G) and the forward scattering intensity (I(0)), by the Guinier equation:

$$I(q) = I(0)exp\left(-\frac{1}{3}R_G^2 q^2\right) \tag{2.14}$$

A linear fit to the plot of $lnI(q)$ against q^2 therefore allows for the extraction of the radius of gyration from the slope and $I(0)$ from the intercept of the slope. The Guinier approximation is only valid for low q regions for globular proteins this region corresponds to $q*R_G \leq 1.3$. A deviation from a linear behaviour in the Guinier region defined by $q * R_G \leq 1.3$ is indicative of poor sample caused by either interparticle interactions which can be caused by radiation damage induced by the incident beam. Once it has been determined from the Guinier analysis that there is no radiation damage or interparticle effects the value of the forward scattering intensity ($I(0)$) can be used to determine the molecular mass if a standard of known mass has also been measured. According to the relationship [145]:

$$M_P = I(0)_p/C_p \cdot \frac{M_{st}}{I(0)_{st}/C_{st}} \tag{2.15}$$

Where M_P and M_{st} are the molecular masses of the protein sample and the standard respectively. $I(0)_p$ and $I(0)_{st}$ are the forward scattering intensities of the sample and standard and C_p and C_{st} are the concentrations of the sample and standard, in grams per litre. The accuracy of this analysis is limited (10 - 20 %) but is suficient to determine the oligomeric state of the protein [145].

Information about the compactness of the scattering particle can be obtained by the Kratky plot ($q^2 * I(q)$ against q). Where the Guinnier plot looks at structural features at low q the Kratky plot emphasises structural features at higher values of q. Examples of different types of biological macromolecules and the corresponding shape of the Kratky plot are shown in figure Figure 2.6. By the Kratky plot analysis it is therefore relatively straightforward to determine whether the sample studied adopts a globular or unfolded conformation in solution.

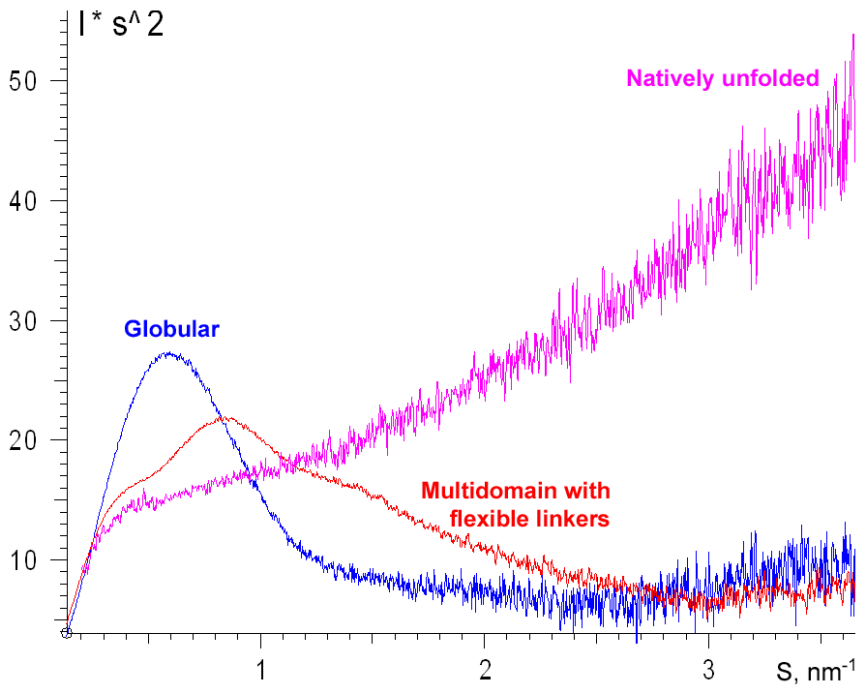

Figure 2.6: Kratky plot of different biological macromolecules. Due to their well defined globular shape folded proteins show a defined maxima, whereas natively unfolded proteins increase at higher values of q. Image taken from Bernado *et al.* [146].

It is also possible to extract information on the size and shape of the particle through indirect Fourier transform methods. A transform can be applied to the scattering data to obtain the spherically averaged auto-correlation function of the excess scattering density yielding the pair distribution function, $p(r)$. The $p(r)$ function essentially provides a histogram of all distances between pairs of points in the scattering particle. The $p(r)$ function is defined as:

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty srI(s)sin(sr)ds \tag{2.16}$$

The $p(r)$ is a function in real space, which is related to the measured scattering intensity $I(s)$ by an inverse transform:

$$I(s) = 4\pi \int_0^\infty p(r)\frac{sin(sr)}{sr}dr \tag{2.17}$$

The value of p(r) reaches zero when the radius exceeds the maximum pair of distance in the sample, a point known as D_{max}. Therefore the upper limit on the inverse transform can be replaced by D_{max} (instead of infinity). The $p(r)$ function can then be obtained indirectly from this inverse transform. This method, known as the indirect Fourier transform method [147], assumes that $p(r)$ can be represented by a linear combination (K) of orthogonal functions (Wk) in the range 0 to D_{max}, where D_{max} must be obtained.

$$p(r) = \sum_{K=1}^{K} c_k \phi_k(s_k) \tag{2.18}$$

The optimal value of c_k is obtained by minimisation of

$$\phi = \chi^2 + \alpha P(p) \tag{2.19}$$

where χ is a measure of the goodness of fit between experimental and calculated data. The calculated data is obtained by the following direct transform of the p(r) function. The variable α provides a balance between the fit of the curve to the data and the smoothness of the curve, and P(p) smooths out the $p(r)$ function (equation ahead):

$$\chi^2 = \frac{1}{N-1} \cdot \left[\frac{I_{exp}(S_J) - cI_{calc}(S_J)}{\sigma(S_j)} \right] \tag{2.20}$$

where N is the number of data points, σ the standard deviation and c is the scaling factor.

$$P(p) = \int_0^{D_{max}} \left[P'\right]^2 dp \tag{2.21}$$

The $p(r)$ distribution not only provides information about the maximum distance but also on the shape of the particle, see Figure 2.7.

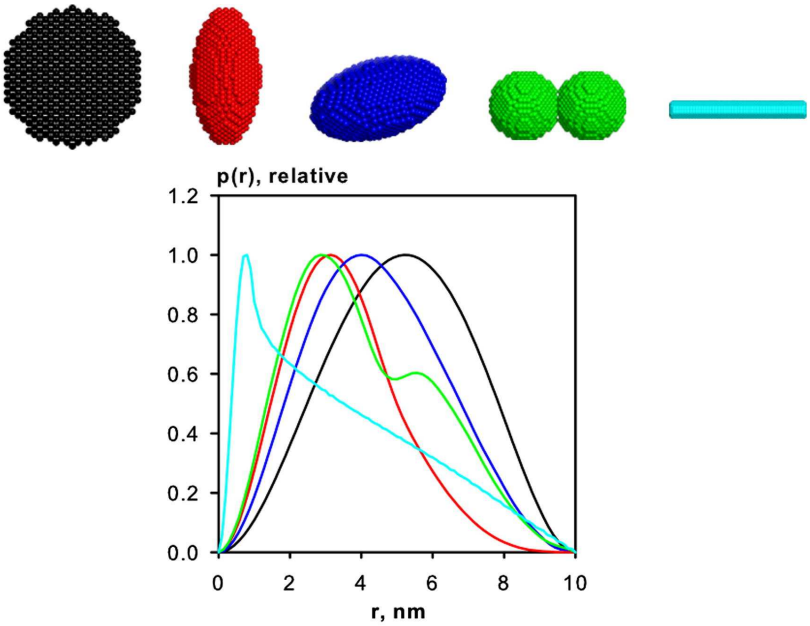

Figure 2.7: Distance distribution functions calculated for different geometrical shapes. Image taken from Mertens [148].

The indirect Fourier transform program, GNOM, automatically performs the calculation and provides as output the $p(r)$ function for a given solution [147].

2.3.2 *Ab initio* modelling

The $p(r)$ function obtained by indirect Fourier-transform methods can be used to produce scattering envelopes of the biological sample. The most common used software for *ab initio* modelling is the DAMMIN software provided by the ATSAS package [149]. DAMMIN stands for dummy atom model minimisation, the software builds model by assuming that each scattering particle is built up of a collection of beads

within a spherical search volume which has a maximum size corresponding to the D_{max}. Each bead is randomly assigned to either solvent (black), or structure (yellow), in random positions (Figure 2.8). From the random starting position, the beads are re-arranged, one at a time, and the fit of the new model to the scattering curve is determined and evaluated by χ^2 in equation 2.20. Limits are applied to the model including that all beads must be connected and that the resulting model must be compact. The algorithm of DAMMIN is essentially as follows:

1. Start from a random position (X_0)
2. Select a random atom, change it from solvent to sample (X_2) and calculate the change ($\delta = X_2 - X_0$)
3. If $\delta =< 0$ then X_2 is the new position and move to next step, else repeat step 2
4. After T steps, if no change is observed for the movement of any atom then finish modelling. Where T is given by the user.

There are several drawbacks to using *ab initio* methods to recover a low resolution structure. Most important to remember is that using these methods a three dimensional shape is being reconstructed from essentially a one dimensional data set.

2.3.3 Flexibility analysis of SAS data by ensemble optimisation methods

Highly flexible systems such as intrinsically disordered proteins (IDPs) are not represented by a single molecule in solution. Instead these structures exist as a heterogeneous ensemble of different conformational structures. Using SAS it is possible to gain several different types of data from IDPs some of which have already been discussed. One of the most useful applications of small-angle scattering in terms of structure determination is the possibility to asses the flexibility of IDPs. Classical analysis of small-angle scattering data results in a single scattering envelope which represents the studied biological macromolecule and fits to the experimentally determined scattering

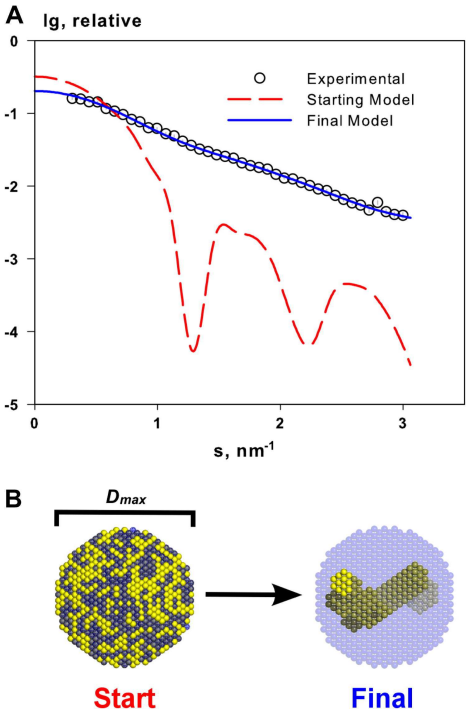

Figure 2.8: The *ab initio* modelling process using DAMMIN. **(A)** From the p(r) distribution of the experimental curve a maximum spherical search volume is found. The fitting procedure is then started until the calculated scattering curve matches the experimentally determined curve. **(B)** Graphical depiction of the spherical search volume determined by D_{max}, with beads assigned going from the starting point to the final bead model. Image taken from Mertens *et al.* [148]

curve. In reality flexible systems exist as a large distribution of different conformational structures. A method to analyse data based on this has been developed and is known as the ensemble optimisation method (EOM). This method represents these heterogeneous structures present in solution by an ensemble of structures containing N different conformations. The scattering from this ensemble can be calculated from the average scattering of N structures present in the ensemble.

In order to pick an ensemble of structures first a large pool (M) of structures is generated at random based on the amino acid sequence of the protein. The amount of structures in this pool is enough so that all areas of conformational space are covered. From this large pool of structures a subset of conformations is picked that best represents the experimental scattering curve. This is achieved by a Monte Carlo

based search. If it is assumed that each of the different conformational models are uniformly populated, so that the intensity of a subset $I(q)$ containing N conformers is

$$I(q) = \frac{1}{N} \sum_{n=1}^{N} I_n(q) \tag{2.22}$$

Where $I_n(q)$ is scattering from the n-th conformation and q is the momentum transfer (as previously defined). The scattering curve is calculated for each model according to

$$\chi^2 = \frac{1}{K-1} \sum_{j=1}^{K} \left[\frac{\mu I(q_j) - I_{exp}(q_j)}{\sigma(q_j)} \right]^2 \tag{2.23}$$

Where K is the number of experimental points on the scattering curve, $I_{exp}(q)$ is the experimentally determined scattering data, $\sigma(q)$ is the error and μ is a scaling factor. The conformations with the lowest χ^2, and hence with the best fit to the data, are selected for further analysis. The final conformation typically results in around 50 structures, a far smaller amount then required to fully explain the flexible behaviour of a protein in solution that samples a large area of conformational space. However these ensembles are sufficient to provide information on the overall properties of flexible proteins in terms of their size and shape. The structures produced by the EOM method are not truly present in solution, instead the structures are said to represent the population of conformational ensembles present in the solution. The distributions of size in terms of D_{max} and R_G are representative. The EOM method is able to find minima in terms of distribution but not in terms of truly representative conformations. A comparison can be drawn by comparing the truly random pool

generated at the start of the algorithm with the selected ensemble. An example of this type of analysis is shown in Figure 2.9.

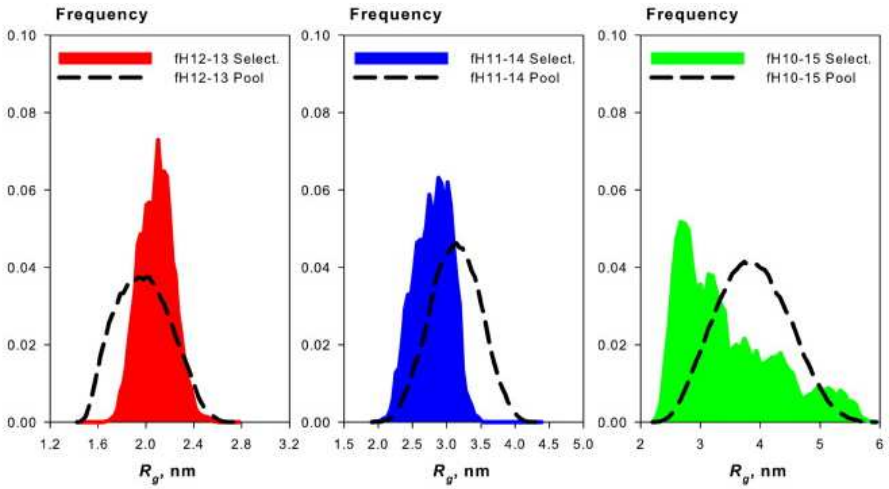

Figure 2.9: The radius of gyration distribution determined by the EOM method applied to three different proteins. The random pools are shown as black dashed lines and the selected ensembles in colour. The distributions are skewed towards either extended structures (fH12-13) or compact structures (fH10-14 and fH10-15). Image taken from Mertens *et al.* [148]

2.3.4 Instruments used for small-angle scattering experiments

Small-angle X-ray diffractometer BM29

BM29 is a pinhole geometry SAXS beamline located on the synchrotron light source at the ESRF in Grenoble [150]. The beamline has been designed specifically for high throughput protein solution scattering experiments. It contains a 2.867 m evacuated tube with a Pilatus 1m detector. Unlike the SANS beamlines the detector is kept at a fixed position, the q-range is selected by altering the energy of the incident beam between 7-15keV. Due to the much higher flux available at synchrotron beamlines the measurement time is considerably shorter for the same sample conditions. In order to achieve a higher throughput of samples the BM29 beamline is fitted with an biological SAXS sample changer, developed by the EMBL [151]. This system automates the measurement of samples and allows for a high trough put of biological samples.

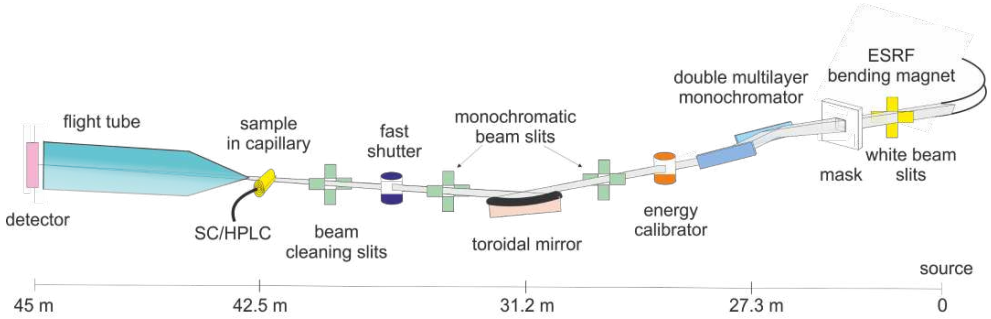

Figure 2.10: Schematic of BM29 the BIO-SAXS beamline at the ESRF. Image taken from the BM29 website (www.esrf.eu/home/UsersAndScience/Experiments/MX/bm29.html)

2.3.5 Small-angle neutron diffractometer D33

D33 is a small-angle neutron scattering instrument at the Institut Laue-Langevin in Grenoble [152]. The instrument can be run in either a time-of-flight (TOF) mode or in a monochromatic mode. In this thesis only the monochromatic mode was used. In monochromatic mode neutrons of a specified energy are selected by a velocity selector to provide the chosen q-range.The extended evacuated tube within which the detectors are located mean that it can collect a large dynamic q-range even in monochromatic mode, in TOF mode a much larger q-range can be collected in a single measurement (Figure 2.11). The downside of measuring a large dynamic q-range in TOF mode is the decrease of signal to noise at low q due to decreased delivery of neutrons from the reactor at the corresponding wavelength. However this option is considered useful for kinetic measurements of certain systems. In monochromatic mode it is however possible to cover a large q-range by measurement of multiple sample to detector distances, this is the typical set up for most biological measurements. The detector is a large area helium 3 multi detector situated in a 20 meter evacuated tube, movement of the detector to different distances is controlled by computer.

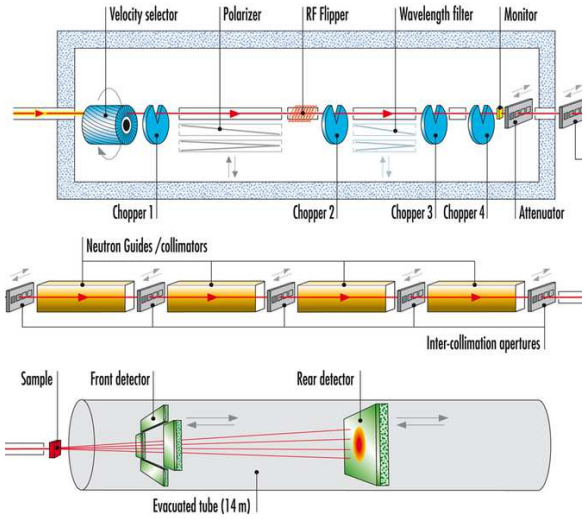

Figure 2.11: Schematic drawing of the D33 instrument at the Institut Laue-Langevin, taken from the ILL website

2.4 Elastic incoherent neutron scattering

Previously the information available from the coherent scattering length was explained. However, according to equation 2.9, the scattered neutrons also contain incoherent contributions. The incoherent scattering length B_{inc} is defined as:

$$B_{inc}^2 = \left\langle B^2 \right\rangle - \left\langle B \right\rangle^2 \tag{2.24}$$

Whereas the coherent scattering length is the average scattering length of the element, the incoherent scattering length is the deviation from the average position ($\langle B^2 \rangle - \langle B \rangle^2$). Incoherent scattering therefore contains information about the dynamics of the atoms present in the studied sample. Measurement of energy resolved neutron scattering spectra provides access to the study of protein motions on the pico to nanosecond time-scale and on the angstrom length-scale. Where the exact time scale studied depends upon the instrumental energy resolution, and the length-scale dependent upon the accessible q-range. An example of a spectra obtained by a typical energy resolved incoherent neutron scattering experiment is shown in Figure 2.12.

During a real experiment such spectra are collected as a function of the scattering vector q. The spectra shows examples of elastic, quasi-elastic and incoherent scattering. At zero energy transfer there is a large elastic peak, these scattered neutrons do not exchange with the sample. The intensities of the elastic peaks as a function of the scattering vector q contain information about the motions of the sample studied. When a neutron interacts with a sample the theoretical scattering function (S_{th}) can be written as [153]:

$$S_{th}(q,\omega) = e^{-<x^2>q^2} \cdot \left[A_0(q) \cdot \delta(\omega) + \sum_n A_n(q) \cdot L_n(\omega) \right] \tag{2.25}$$

The equation contains several different components, $A_0(q)$ is referred to as the elastic incoherent structure factor, which contains information about the motion of the sample studied. The delta function $\delta(\omega)$ needs to be considered when there are motions in the sample slower than the resolution of the instrument. The second part of the equation contains a sum of quasi elastic components, described by a sum of Lorentzians ($L_n(\omega)$) with the quasi elastic incoherent structure factor ($A_n(q)$). The above equation is a simplification of the theoretical scattering function which ignores the presence of long range translational diffusion. Fast vibrational motions in the sample are contained within the Debye-Waller factor ($e^{-<x^2>q^2}$).

2.4.1 Mean squared displacements determined by elastic incoherent neutron scattering

During a fixed window elastic temperature scan the neutron intensity is measured at zero energy transfer, from this data a model independent mean square displacement can be obtained. The intermediate scattering function ($I(q,t)_{inc}$) is defined as the Fourier transform of the Van Hove function, which contains a self and a distinct part. The intermediate scattering function of a component containing N atoms is

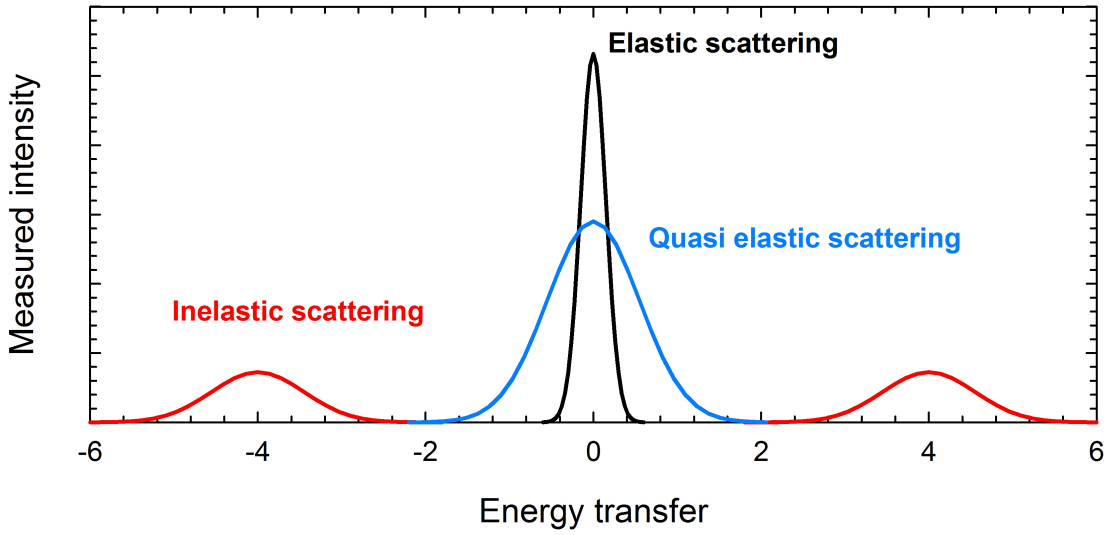

Figure 2.12: Example of a spectra obtained for a typical time-resolved incoherent scattering experiment.

$$I(q,t)_{inc} \approx \frac{1}{n}\sum_{i} exp(iq\left[r_{[i]}(0) - r_{[i]}(t)\right] \tag{2.26}$$

where $r_{[i]}(0)$ and $-r_{[i]}(t)$ are position vectors at time = 0 and time = t. If it is assumed that the distribution of atoms around their average position can be described by a Gaussian approximation the expression can be simplified in a Gaussian approximation. As q approaches zero, the approximation is also valid for motions following a non Gaussian distribution so long as the motions are localised in the length and time window of the spectrometer used. Under this assumption ($I(q,t)_{inc}$ = can be simplified [154]:

$$I(q,t)_{inc} \approx \frac{1}{n}\sum_{i}^{e} xp\left(-\frac{1}{6}\cdot q^2\left\langle\left[r_{[i]}(0) - r_{[i]}(t)\right]^2\right\rangle\right) \tag{2.27}$$

The mean square displacement of each atom i, $\langle U_i^2 \rangle$, corresponding to the amplitude

of motions present in the sample can then be introduced [155] :

$$\left\langle U_i^2 \right\rangle = \left\langle \left[r_{[i]}(0) - r_{[i]}(t) \right]^2 \right\rangle \tag{2.28}$$

The incoherent scattering function $S(q,0)_{inc}$ is then

$$S(q,0)_{inc} = \frac{1}{N} \sum_i eexp\left(-\frac{1}{6} \cdot q^2 \left\langle u_i^2 \right\rangle \right) \tag{2.29}$$

The average mean squared displacements of all incoherent scatterers in the sample can then be obtained from the slope of the plot of log scattered intensities against q^2.

$$\left\langle u^2 \right\rangle = \frac{-3 \cdot \Delta lnS(q,0)}{\Delta q^2} \tag{2.30}$$

This approximation is similar to the Guinier approximation used in small-angle scattering experiments [156]. The approximation is valid up until $\langle u^2 \rangle \cdot q^2 \approx 2$ [156].
By far the largest contributor to incoherent scattering in biological systems are hydrogen atoms. Hydrogens are spread quite evenly on the sidechains of amino acids on the protein backbone. It can therefore by implied that the average movement of hydrogens represents the average movement of the whole protein.
Often mean squared displacements are measured as a function of temperature in a so called fixed window elastic temperature scan. The intensities measured at zero energy transfer are large and therefore this method allows for good statistics to be collected in a relatively short period of time. An example of a fixed window temperature scan performed on the backscattering instrument IN16 at the ILL is shown in Figure 2.13.

The Gaussian approximation linear fit and the deviation thereof at higher q can be seen in Figure 2.13 A.

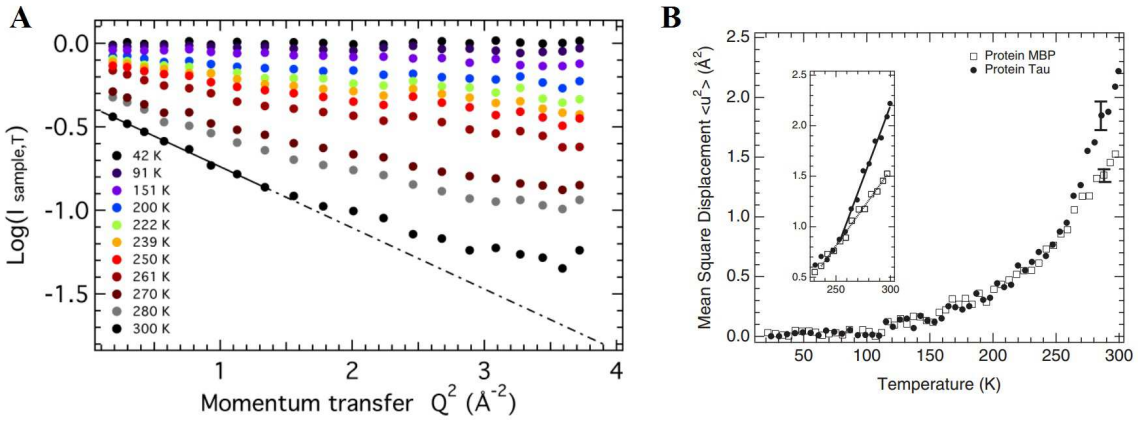

Figure 2.13: Fixed window temperature scan performed on a hydrated protein powder on IN16.**(A)** The Gaussian approximation applied with the linear fit applied to low q. **(B)** The obtained mean squared displacements plotted as a function of temperature. Image taken from Gallat *et al.* [157]

A mean force constant $< k' >$ can obtained from the dependence of the mean squared displacement against temperature, shown in the inset in Figure 2.13 B [158]. According to Zaccai [158] the force constant can be described by:

$$< k' >= \frac{0.00276}{\frac{d<u^2>}{dT}} \tag{2.31}$$

Where the top half of the fraction results in $< k' >$ in N/m. The mean force constant obtained provides information about the rigidity of the protein sample and can be compared across different samples so long as the instrumental space and time window remains constant [158].

2.4.2 The relationship between instrumental resolution and motion

Different neutron scattering instruments have different energy resolutions (ΔE) with a corresponding time resolution (ΔT):

$$\Delta T = \frac{\hbar}{\Delta E} \tag{2.32}$$

where $\hbar$ is Planck's constant. The length window of motions studied depends upon the accessible q range. Only motions that are accessible on the corresponding time and length scale can be observed. Therefore the obtained mean square displacements are instrument dependent. Motions studied can range from the pico second to the nanosecond and often different instruments are used to measure motions over different timescales (see Table 2.3).

2.4.3 Instruments used for incoherent neutron spectroscopy

The basic principle of neutron spectroscopy instruments is the measurement of scattered neutron intensity as a function of both the energy transfer ΔE and the scattering vector q. In this thesis two different types of instruments are used to achieve this goal, time-of-flight instruments and backscattering instruments. The basic principles are similar,at reactor based neutron sources a monochromatic beam needs to be obtained from the polychromatic stream of neutrons produced by the reactor. This is achieved by the use of a monochromator, backscattering instruments often use a single crystal that obtains a monochromatic beam using Bragg reflections. Time-of-flight instruments can also use single crystals but often multiple choppers are used to extract a monochromatic beam. Time of flight spectrometers chop the beam so that a pulsed beam of neutrons with the same energy arrive simultaneously at the sample. It is

known at what time the neutrons arrive at the sample, the scattered intensity is then measured as a function of time. If a neutron gains energy from the sample it's speed will increase and arrive first at the detector, followed by neutrons that did not exchange energy with the sample and finally by neutrons that lose energy to the sample. The energy resolution reached by a time of flight spectrometer is based on the energy bandwidth of the incident beam, and ranges from around 5 eV to 200 eV. The principle underlying this type of instrument is explained in detail elsewhere [159].

Backscattering on the other hand uses the principle of Bragg reflection to obtain a higher energy resolution. The underlying principle of backscattering is described in detail elsewhere [160]. Briefly, if a neutron beam is reflected by a crystal close to a 90 degree angle the resulting reflected beam will have a very small energy dispersion, ΔE. The neutron beam that has undergone Bragg reflection is then scattered by the sample towards analyser plates which reflect the beam under close to backscattering conditions towards neutron detectors. A typical layout of a backscattering instrument is shown in Figure 2.14. This allows a backscattering instrument to reach much lower energy resolutions from around 1 eV upto 8 eV. Due to the small energy bandwidth backscattering spectroscopy typically has lower counting statistics than similar time of flight instruments, however lower energy resolutions can be reached. By using a combination of different instruments with different energy resolutions it is therefore possible to cover several orders of magnitude in energy and time resolution (see Table 2.3). The decreased counting statistics obtained by backscattering instruments mean that large amounts of sample are required (typically around 100 mg), this combined with the low sample quantity typically available for protein samples (micro to milligrams) means that often only the elastic incoherent measurement is performed.

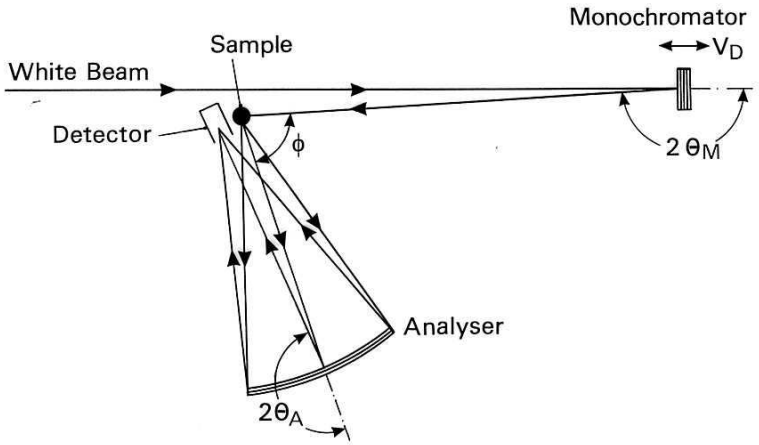

Figure 2.14: Schematic drawing of a typical neutron backscattering instrument. Image taken from www.ill.eu/othersites/BS-review/

2.5 Dynamic light scattering theory

Dynamic light scattering refers to the measurement of the fluctuation in intensity of scattered light over time. The change in intensity is interpreted and related to the Brownian motions of the scattered particles.This interpretation analyses the scattering intensity fluctuations by using correlation functions. The results of such an

Table 2.3: Resolution and timescale reached by different neutron spectrometers at the Institut Laue Langevin. BS = backscattering and TOF = time of flight instruments.

Instrument	Type	Resolution	Timescale
IN16	BS	$0.9\mu eV$	$\approx$ 1 Nanosecond
IN13	BS	$8\mu eV$	$\approx$ 100 Picoseconds
IN6	TOF	$50\mu eV$	$\approx$ 20 Picoseconds

experiment provides information about the hydrodynamic radius of the scattered particle in solution. Where the hydrodynamic radius is the radius of a hard sphere that diffuses at the same rate as the sample.

All information about the diffusion of a particle through a solvent can be described by the dynamic structure factor $(F_s, (q\tau))$. The dynamic structure factor is the Fourier Transform of the van Hove self-correlation function $(G_s(r,t))$.

$$F_s(q,\tau) = \int G_s(r,\tau) exp(iqr) dr \tag{2.33}$$

And the Van Hove correlation function is:

$$G_s(r,\tau) = \langle n(0,t) n(r,t+\tau) \rangle_{V,T} \tag{2.34}$$

Where $n(r,t)$ is the amount of scattering particles diffusion with Brownian motion, within a volume centred on r at time t. For dilute solutions $G_s(r,\tau)$ is the probability of finding a particle at time $t+\tau$ at position r if the particle was previously at position O. For dynamic light scattering measurements only difference vector, $r, -0$, and time difference τ are used for interpreting results. Therefore the measurement is relative diffusion of the particle over time.

The Brownian motion of a protein in solution can be described by a random walk, where the van Het Hove self correlation function depends only upon the distance. to give:

$$G_s(r,\tau) = \left[\frac{2\pi}{3} \delta R(\tau)^2 \right]^{\left[\frac{3}{2}\right]} exp\left(-2 \cdot \frac{3r(\tau)^2}{2 < \delta R(\tau)^2 >} \right) \tag{2.35}$$

Where $\delta R(\tau)^2$ is the square of the Brownian motion that the particle undergoes in time τ. The random movement of the particle through the solvent can be described by

$$\delta R(\tau)^2 = 6D_s\tau \tag{2.36}$$

Where D_s is known as the self diffusion coefficient.

During a dynamic light scattering measurement the Fourier transforms of the van Hove correlation function ($G_s(r,\tau)$ is measured. The Fourier transform gives the dynamic structure factor:

$$F_s(q,\tau) = exp\left(-D_s q^2 \tau\right) \tag{2.37}$$

Once the self diffusion coefficient (D_2) has been measured it is possible to determine the hydrodynamic radius of the scatterer by using the Stokes-Einstein relationship:

$$D_s = \frac{kT}{f} = \frac{kT}{6\pi\eta R_H} \tag{2.38}$$

Where R_H is the hydrodynamic radius, or the radius of a sphere equal to the size of the scattering particle, which undergoes the same Brownian motion.

Chapter 3

The dynamic footprint of sequestration in the molecular fluctuations of osteopontin

Experimental work presented in the following chapter has been submitted for peer review and publication:

- **Lenton, S.,** Nylander, T., Holt, C., Seydel, T., Haertlein, M. Teixeira, S, and Zaccai, G. The dynamic footprint of sequestration in the molecular fluctuations of osteopontin

Abstract

Different length fragments of osteopontin are known to adopt unfolded conformations in solution. In milk the highly phosphorylated OPN1-149 phosphopeptide represents the largest species. Previous studies by light scattering and SAXS have shown that the OPN1-149 fragment can sequester amorphous calcium phosphate and form a type of calcium phosphate nanocluster. The sequestration of calcium phosphate by unfolded proteins is fundamental to the stabilisation of biofluids supersaturated with respect to hydroxyapatite. The unfolded state of osteopontin is thought to be a prerequisite for this activity. Here the effects of sequestration on the molecular fluctuations of OPN1-149 are quantified on the nanosecond-angstrom resolution by elastic incoherent neutron scattering. The molecular fluctuations of the free phosphopeptide are found to represent a highly flexible protein. Upon sequestration of calcium phosphate, stiffening of osteopontin is reflected in molecular fluctuations more closely resembling those characteristic of globular proteins. The results bring insight into the modulation of the activity of osteopontin and phosphopeptides with a role in the control of biomineralisation. The quantification of the effects provides a handle for future comparisons and classification of molecular fluctuations.

3.1 Introduction

Osteopontin (OPN) has been called "a molecule for all seasons" because it participates in numerous cellular functions including cell adhesion, signalling, migration, and survival, in addition to roles in immunomodulation and biomineralisation [161]. The ability of a single protein to participate in multiple cellular functions is in part because *in vivo* OPN is heavily modified by post translational enzymes, resulting in a range of OPN peptides with different amino acid lengths and different functional groups attached [86, 99]. OPN belongs to the family of Small Integrin-Binding Ligand, N-linked Glycoproteins (SIBLINGs) [162]. The SIBLINGs are a sub-group of a larger, paralogous group of secreted proteins to which the caseins also belong. A key characteristic of the SIBLINGs is their high degree of phosphorylation and the effect this has on the function of individual members. For example OPN is highly phosphorylated in the Golgi apparatus by the FAM20C golgi kinase, and then secreted [96]. Mutations that inhibit the golgi kinase from phosphorylating OPN leads to a bone disease known as Raine syndrome [85, 96, 163]. The secreted phosphorylated OPN protein is present in all biofluids where it has been sought. The role of OPN in biofluids is to prevent the precipitation of high levels of calcium phosphate [10]. This is achieved by sequestration of calcium phosphate by OPN leading to the formation of so called calcium phosphate nanoclusters (CPNs). Sequestration of calcium phosphate by OPN phosphopeptides is phosphorylation dependent, and the formed CPN are core shell structures containing a core of calcium phosphate surrounded by an OPN shell. In bovine milk the highly phosphorylated OPN is processed proteolytically between 157-158 by the plasmin protease. This fragment is referred to in the literature as OPN1-149 for historical purposes. The OPN1-149 phosphopeptide contains regions which are highly phosphorylated through which it sequesters calcium phosphate.

The recombinant expression of OPN for experimental study is complicated due to the presence of significant and functionally important post-translational modifications in the native protein. The length of the full OPN chain makes NMR assignments in the native protein somewhat difficult and has required work with labelled but (to-date) unphosphorylated and unglycosylated recombinant material. Due to these limitations only limited characterisations on several different, usually short, OPN fragments has been completed. Fisher *et al.* determined, using 1-D 1H NMR and transverse relaxation times, that the solution structure of native OPN was consistent with that of an intrinsically disordered protein (IDP) [84]. The performed NMR experiments were limited and only obtained information which showed that the studied fragment was non globular in solution. More recently Kurzbach *et al.* used more advanced NMR techniques to show that OPN displays compact and extended conformations in solution and interconverts between these conformations through cooperative phase transitions [101]. Further structural insights were obtained by the experiments of Kurzbach *et al.*, however these experiments were based on a limited OPN fragment due to problems with protein expression [164]. Yamaguchi *et al.* have shown that a recombinantly produced fragment of mouse OPN exhibits transient long-range intramolecular interaction between its N- and C-terminal regions that can shield the central part of the chain from integrin interactions [165]. These experiments show that different OPN peptides behave differently but all share a common non globular structure. However all of the listed experiments performed on OPN lack any kind of post-translational modifications found *in vivo*, which have been shown to be fundamental for OPN function. A structural study of the previously mentioned OPN1-149 isolated from a bovine source has been completed using SAXS. The results from this study indicated that bovine OPN1-149 adopts an unfolded conformation in solution, however analysis of the data reported in this paper was limited to only model independent methods [45]. From the results it was hypothesised that the adaptation of

an unfolded conformation was vital for OPN1-149 to sequester calcium phosphate in a stable CPN [45].

The literature therefore provides broad evidence that several OPN phosphopeptides are proteins which lack secondary structure elements, known as intrinsically disordered proteins (IDPs). IDPs are a unique family of macromolecules that are capable of fulfilling many biological functions despite entirely or partially lacking a unique 3D structure (for recent reviews see, for example, Habchi *et al.* and Oldfield Dunker [106, 166, 167]). A disordered conformation is thought to convey several functional benefits [106, 166, 167]. For example, IDPs are able to adapt to different environments and interact rapidly with one or more partners. A possible combination of intrinsic disorder with a high degree of controllable post-translational modification is why a single protein such as OPN can regulate so many cellular functions [102]. Several proteins have been identified as able to sequester calcium phosphate in a similar manner to OPN1-149. Apart from a high degree of phosphorylation all recognised proteins adopt a largely or completely unfolded structure [23]. A key advantage of adopting an unfolded conformation is the ability to achieve a high density of packing of sequestering phosphopeptide chains on the core surface and thereby maximising the free energy of sequestration by the shell [72]. It has been suggested that another advantage of adopting an IDP conformation is the higher flexibility afforded to an IDP [23]. This would allow IDPs such as OPN1-149 to sample a large area of conformational space and rapidly sequester calcium phosphate prior to a nuclei of calcium phosphate forming [67, 41]. Despite significant advances in NMR and other techniques, relatively little experimental evidence is available regarding this increased flexibility of IDPs. Therefore a full characterisation of OPN1-149 requires not only information about the structure but also it's flexibility in solution. The flexible behaviour of proteins covers several orders of magnitude in timescale ranging from electronic re-arrangements on the femto second scale to protein folding on the order

of seconds or minutes (Figure 3.1) [168, 169]. The atomic thermal fluctuations of a protein can be related to the shape of the potential energy well within which the protein moves [158]. Elastic incoherent neutron scattering has been extensively used to study the molecular fluctuations of globular proteins [170, 171].

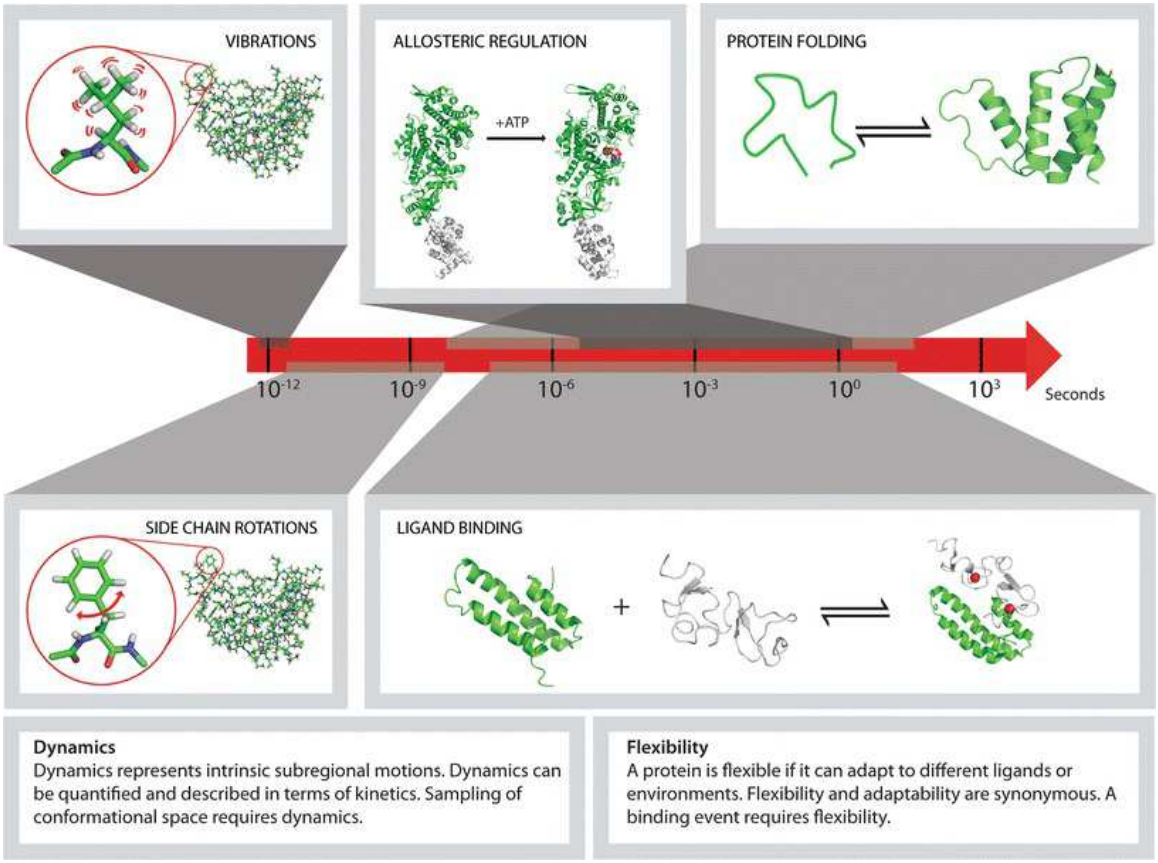

Figure 3.1: Protein motions over periods corresponding to several orders of magnitude in time, ranging from the vibration of hydrogen atoms to protein folding events. image taken from [172].

In more recent years this technique has been used to study intrinsically disordered proteins, including casein and Tau protein [157, 173]. The disordered Tau protein was studied by elastic incoherent neutron scattering and was shown to be more flexible compared to a folded protein of similar molecular mass [157]. Similar results were obtained for a β casein phosphopeptide by inelastic neutron scattering which also showed behaviour of a highly flexible protein [173]. The results indicate that in general disordered proteins are more flexible compared to globular proteins of similar

molecular weight. A high degree of flexibility is thought to be vital for the multi-functional ability that many IDPs show. However so far no dynamical studies have been completed on IDP in different bound states and the effects of being bound to a partner on protein dynamics is not well explored.

In this study elastic incoherent neutron scattering is used to investigate the temperature dependent protein dynamics of bovine OPN1-149. The effects of sequestration on the molecular dynamics are also investigated. Due to the distribution of hydrogen in protein samples elastic incoherent neutron scattering provides information on the displacement of hydrogen atoms and the side-chains to which they are attached, which is used as a measure of protein flexibility. The length scale studied is on the order of Ångstroms and the timescales range from nano to pico second depending on instrumental resolution [156]. The results obtained indicate a high flexibility of OPN1-149 which is reduced upon sequestration of calcium phosphate. The structural counterpart of the dynamics study is presented in the form of small-angle X-ray scattering (SAXS), which shows that the unfolded protein adopts a heterogeneous ensemble of structures in solution. The structural and dynamics data is then combined to show that OPN1-149 can rapidly sample large areas of conformational space and this is related to the protein function *in vivo*.

3.2 Materials and methods

3.2.1 Bioinformatic structure predictions

Structural disorder tendency of OPN1-149 was predicted by IUPred [174]. The accuracy of this predictor is high at around 80-90 % [174]. The predictor uses only the amino acid sequence of OPN1-149 and returns a score that ascribes the possibility of structural disorder on a per residue level [174]. A score given above 0.5 means the region is likely to adopt a disordered conformation. Further analysis was completed

by plotting the mean net charge of the sequence against the hydrophobicity determined by the Kyte-Doolittle hydrophobicity scale, using the online PONDR server [175]. The online PONDR server also adds the position of a reference set of globular and intrinsically disordered proteins for comparison, with a boundary determined by Uversky *et al*. [175].

3.2.2 Preparation of native bovine OPN1-149

Native bovine osteopontin (OPNmix) was kindly provided as a fraction of osteopontin purified from bovine milk by the method of Sørensen *et al* by Dr C Holt (Glasgow University) [99]. From the OPNmix the OPN1-149 peptide was purified to homogeneity by size exclusion chromatography (SEC) as described by Holt *et al* [45]. In brief, osteopontin fragments were separated by size on a 120 ml volume column containing a Superdex 75 (G.E. Healthcare) stationary phase, connected to a Biologic dual flow purification system (Bio-Rad). Size exclusion was performed at room temperature, the column was equilibrated with 2 column volumes of elution buffer (50 mM phosphate, pH 7.0, 300 mM NaCl, 0.02% NaN_3) at a flow rate of 0.3 ml/min. 100 mg of OPNmix was then dissolved in the same elution buffer and loaded onto the column at a rate of 0.3 ml/min. Elution from the column was constantly monitored by U.V. absorption at $\lambda = 280$ nm, fractions were collected on volumes of 1 ml, elution was stopped after 2 column volumes of elution buffer. Two main elution peaks were observed, aliquots were taken from the eluted fractions and analysed by SDS-PAGE . N-terminal sequencing was used to confirm the identify of OPN1-149 in the indicated peak. Fractions deemed to be homogeneous by SDS-PAGE analysis were pooled, extensively dialyzed against water, flash frozen and lyopholized until dry as determined by weighing. The resulting dry OPN1-149 protein was stored in a desiccator over phosphorus pentoxide at room temperature and used as required.

3.2.3 Formation of calcium phosphate nanoclusters stabilised by OPN1-149

Calcium phosphate nanoclusters were formed using the mixing method described previously by Holt [45]. In the presence of an excess of osteopontin the pH of a solution was raised slowly by addition of calcium phosphates salts of different basicity (mono, di and trisodium phosphate), whilst continuously mixing, until a pH of 7.4 was reached.

3.2.4 Circular dichroism experiments

Circular dichroism (CD) spectra were obtained on a J-720 spectrometer (JASCO) at 0.5 mg/mL (50 mM phosphate pH 7.4 , 150 mM NaCl). The sample environment was kept at 20 °C; the spectrum was recorded between 190 nm and 250 nm accumulating data over 3 runs with a band width of 1 nm. The raw data was normalised for protein concentration and the number of amino acid residues. Secondary structure content was analysed using the Dichroweb analysis server [176], using a reference data set which contains the spectra of denatured proteins [177].

3.2.5 Small-angle X-ray scattering experiments

Small angle X-ray scattering experiments were performed at the I1911-SAXS beamline at the MAX II ring of the MAX IV laboratory (Lund, Sweden) [178]. Prior to the measurements, the samples were exhaustively dialyzed against buffer (50 mM phosphate pH 7.4 , 150 mM NaCl), centrifuged and filtered (0.22 μm Sartorius syringe filter), after which the concentration was determined by U.V. absorption at 280 nm. Three OPN1-149 concentrations (1.5, 3 and 5 mg/ml) were measured and compared to rule out concentration dependent effects. Temperature controlled measurements (20 °C) were performed in a 1 mm capillary. Scattering intensities were recorded using a 2D sensitive detector. The sample to detector distance was 1.45m with an incident

radiation wavelength (λ) of 0.91 Å, providing an accessible q-range of 0.01-0.4 $Å^{-1}$, where $q = 4\pi * sin\theta/\lambda$ and 2θ is the scattering angle. 6 frames were recorded for 10 seconds per frame; each frame was inspected for signs of radiation damage before averaging. Corresponding buffers from dialysis were measured before and after each sample, using the same measurement settings and the resulting curve subtracted from sample scattering.

3.2.6 Small-angle X-ray scattering data analysis

Scattering envelopes of OPN1-149 were obtained using two different *ab intio* approaches: dummy atom and dummy residue modelling from the DAMMIF [149] and GASBOR [179] software, respectively. The DAMMIF assembly of closely packed dummy atoms are packed within a sphere which has a diameter equal to the D_{max}, determined by the indirect Fourier transform with the GNOM software [147]. Using simulated annealing the software searches for a conformation within the defined sphere that matches the experimental data. This approach will find different solutions that fit the data [180], therefore 20 independent DAMMIF computations were performed. The models were then averaged using DAMAVER [149]. A filtered final model was produced by removing low occupancy space. The flexibility of OPN 1-149 was further analysed using the ensemble optimisation method EOM from the ATSAS package [146]. EOM exploits a genetic algorithm to create a large heterogeneous ensemble of models based on the amino acid sequence. A pool of 10 000 conformers was created, from which an ensemble of 20 structures were picked by minimisation of the discrepancy between the calculated scattering profile and the experimental scattering curve. Analysis of the resulting data provided information on the size distribution of the protein and the 20 protein conformers that most accurately represent the ensemble as a whole.

3.2.7 Bioinformatic flexibility predictions

Predictions of OPN 1-149, tau and maltose binding protein (MBP) backbone flexibility were also made using the Dynamine predictor with the corresponding amino acid sequences [181]. Tau was chosen for comparison to an IDP, while the MBP provides a comparison to a globular protein of a similar size (both taken from Gallat *et al.*) [157]. The predictor functions by comparing sequence information with a database of recorded NMR spectra. It predicts an average S2 value for every residue in a given sequence, where S2 describes how restricted the movement of the atomic bond N-H protein backbone vector is with respect to the molecular reference frame. A value of 1 reflects complete order in a stable conformation, while a value of 0 means random bond vector movement (highly dynamic). Values of S2 predicted using DynaMine are on an absolute scale, allowing for comparison between predictions for different proteins. Values of S2 less than 0.8 are considered to indicate a flexible region [181].

3.2.8 Elastic incoherent neutron scattering experiments

Lyophilized powders (150 mg) of OPN 1-149 and OPN CPN were re-hydrated in a desiccator over pure D_2O until a level of 0.44 grams of D_2O per gram of protein was reached, as determined by weighing. Samples were then sealed with indium wire within a flat aluminium sample holder. Powder samples were chosen so that global motions (diffusion and rotation of the entire molecule which are significant in the experimental resolution windows chosen) are suppressed on the chosen time scale, while the humidity ensures a functional sample environment.

EINS experiments were carried out on the backscattering IN16 spectrometer at the Institut Laue-Langevin (ILL, France). An instrumental resolution of 0.9 μeV (full width half maximum) was used corresponding to motions occurring on the 1 ns timescale. The wavelength of the incident neutron beam was 6.27 Å, corresponding to a mea-

sured q-range of 0.02-1.9 $Å^{-1}$. Each sample was inserted into an orange ILL cryostat at room temperature before decreasing to 11K. The sample was held at a 135 °angle with respect to the incident neutron beam. Elastically scattered neutrons were continuously counted whilst raising the temperature from 11K to 300K at a rate of 0.2K/min. The samples were probed for crystalline ice formation using the diffraction setting on IN16. No Bragg peaks were observed. The elastically scattered neutrons of the empty cell $I_{(empty.T)}(q)$ were also recorded in a similar manner for subtraction of sample holder effects. The elastic contributions from the sample scattering were then obtained according to the formula

$$I_{sample.T}(q) = \frac{I_{total.T}(q) - t * I_{empty.T}(q)}{I_{total.T}(q) - t * I_{empty.T}(q)} \tag{3.1}$$

3.2.9 EINS data analysis

Taking only incoherent scattering into account, the inelastic scattering law can be described as a function of q and the frequency Ω ($\hbar\Omega$ is the energy exchange experienced by the sample):

$$S(q,\Omega) = \frac{1}{2\pi}\int I(q,t)exp(-i\Omega t)dt \tag{3.2}$$

The intermediate scattering function is

$$I(q,t) = \int_{-\infty}^{\infty} g(r,t)exp(iq\cdot r)d^3r \tag{3.3}$$

For N hydrogen atoms in position R, the Van Hove auto correlation function $g(r,t)$ can be described as:

$$g(r,t) = \frac{1}{N}\left\langle \sum_{l=1}^{N} \delta\left\{r + R(0) - R(t)\right\}\right\rangle \tag{3.4}$$

Where δ is a Dirac δ function, $R(0)$ and $R(t)$ are the position vectors for atomic motions with no preferential orientation and no time dependence on the instrumental resolution complying with:

$$q^2 \left\langle u^2 \right\rangle \leq 2 \tag{3.5}$$

Where u^2 is the mean squared displacement (MSD). The elastic intensity can then be approximated by applying the Gaussian approximation, analogous to the Guinier approximation used for SAS, where the shape described by the motion of the particle is:

$$\left\langle u^2 \right\rangle = 2R_g^2 \tag{3.6}$$

The elastic scattering intensity is approximated by:

$$I_{sample.T}(q) = exp\left(\frac{-q^2 < u^2 >}{6}\right) \tag{3.7}$$

The atomic MSDs were extracted from the normalised data using the program LAMP provided by the ILL, from a linear fit of the q dependency on the elastic intensity described by the Gaussian approximation shown in equation 3.7.

3.3 Results

3.3.1 Purification of OPN1-149

Native bovine osteopontin is secreted as a 276 amino acid protein at relatively high concentrations in most bodily fluids [182]. In bovine milk osteopontin is cleaved by plasmin and other milk proteases between the residues: *Lys145-Ser146, Arg147-Ser148, Lys149-Lys150, Phe151-Arg152, Arg152-Arg153, and Arg153-Ser154* [183]. The resulting osteopontin peptides contain an average of 150 amino acids, initially this

group of osteopontin peptide fragments was thought to contain only a single fragment cleaved between residue 149 and 150, termed OPN1-149 [45]. Although incorrect, for readability the historical name OPN1-149 is used in this thesis to describe this group of similarly sized peptides. The desired OPN1-149 peptide was purified from OPNmix by single step size exclusion chromatography. The elution trace was monitored by U.V. absorption as a function of retention time and is shown in Figure 3.2. Two major peaks were obtained as previously observed during purification of OPN1-149 from OPNmix [45]. The second peak has previously been identified as the fraction containing OPN1-149 [45], which was confirmed by mass spectrometry analysis.

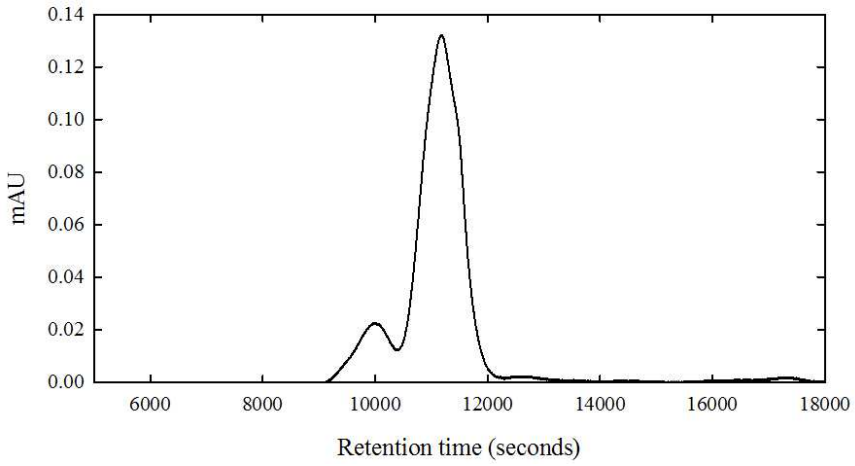

Figure 3.2: U.V. at trace 280 nm of the elution from the S75 column after loading with OPNmix. The second larger peak was consequently identified as OPN1-149 .

3.3.2 Primary sequence analysis

Many computational tools have been developed to predict the propensity of a certain amino acid sequence to adopt a disordered structure. One of the more simple techniques relies on the fact that, in general, IDPs tend to contain a high proportion of amino acids with a high negative net charge and low hydropathy [175]. A high net charge inhibits the hydrophobic collapse by charge-charge repulsion and a low hydrophobicity prevents the hydrophobic collapse of the amino acid chain [175, 184].

A plot of the normalised net charge against the normalised hydropathy, both of which are calculated from the amino acid sequence, yields the charge-hydropathy plot. This charge-hydropathy plot contains a well defined boundary which separates proteins with a disordered tendency from globular proteins [175, 184]. The amino acid sequence of OPN1-149 was analysed by this technique and compared with a reference set of IDPs and globular proteins. The resulting plot is shown in Figure 3.3 where the boundary between globular and disordered proteins is indicated. The amino acid properties of OPN1-149 show a low mean scaled hydropathy combined with a high net charge. These properties place OPN1-149 well within the charge-hydropathy region expected for intrinsically disordered proteins.

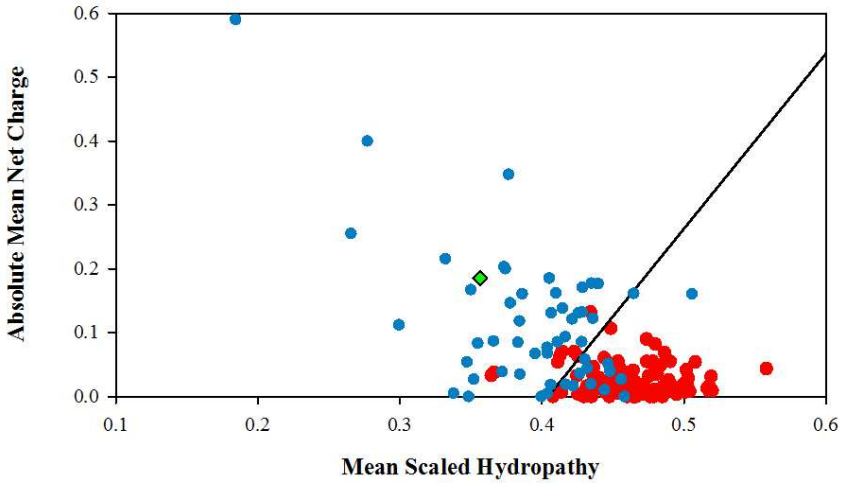

Figure 3.3: A comparison of the mean net charge and mean hydropathy for a set of 275 globular (red circles) and disordered (Blue circles) proteins taken from Uversky *et al.* [175]. The boundary between order and disorder is indicated by a black line. The calculated properties of OPN1-149 is indicated by a green triangle, and lies well within the disordered region.

Predictions of the amount of secondary structure elements within the OPN1-149 chain were undertaken using the residue based disorder predictor IUPred [174]. The resulting profile produced by IUPred is shown in Figure 3.4, the disorder boundary is indicated by a dashed line, anything above this value is considered disordered. The analysis by IUPred predicts that OPN1-149 is almost completely disordered and

according to the prediction there is only one small region which has the possibility of forming globular like regions such as alpha helices and beta sheets. These results emphasise the charge-hydropathy plot analysis that OPN1-149 has an amino acid contents characteristic of those found in intrinsically disordered proteins. Considering the predicted IDP structure of OPN1-149, and what is known about other OPN fragments, several biophysical techniques where chosen to further characterise the protein structure.

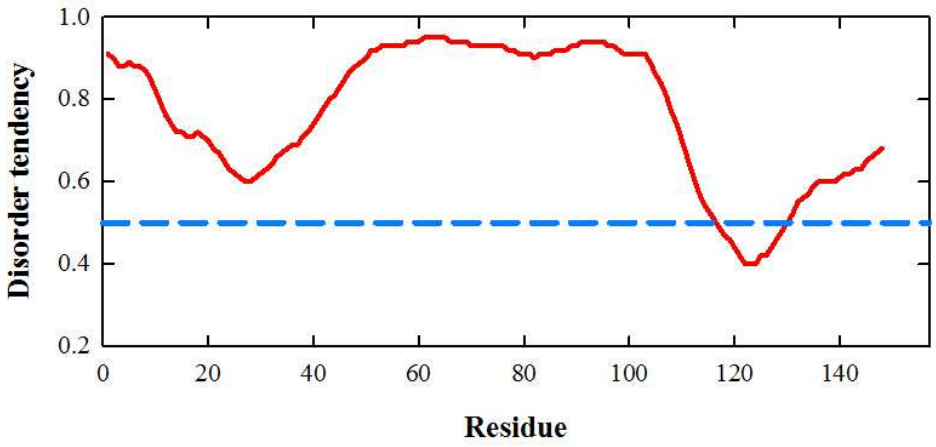

Figure 3.4: Disorder profiles predicted of OPN1-149 by the IUPred method , a dashed blue line shows the value above which the region is considered disordered [174].

3.3.3 OPN1-149 displays biophysical properties of a non globular protein

Circular dichroism spectroscopy was used to investigate the presence of local secondary structure elements in OPN1-149. The CD spectrum of OPN1-149 is shown in Figure 3.5 A. The observed spectrum is typical of a protein lacking in secondary structure elements, such as alpha-helix and beta-sheet, and has the appearance of a spectrum caused mostly by disordered protein structure. Further deconvolution analysis of the CD spectra with a comparison against a reference database indicates that the protein has a secondary structure content consisting of 8% alpha helices,

20% beta-sheet and 72% unordered or random coil structure. The analysis of secondary structure elements within OPN1-149 by circular dichroism indicates that the protein is lacking in well defined secondary structure elements, and agrees with the predictions made by the amino acid analysis.

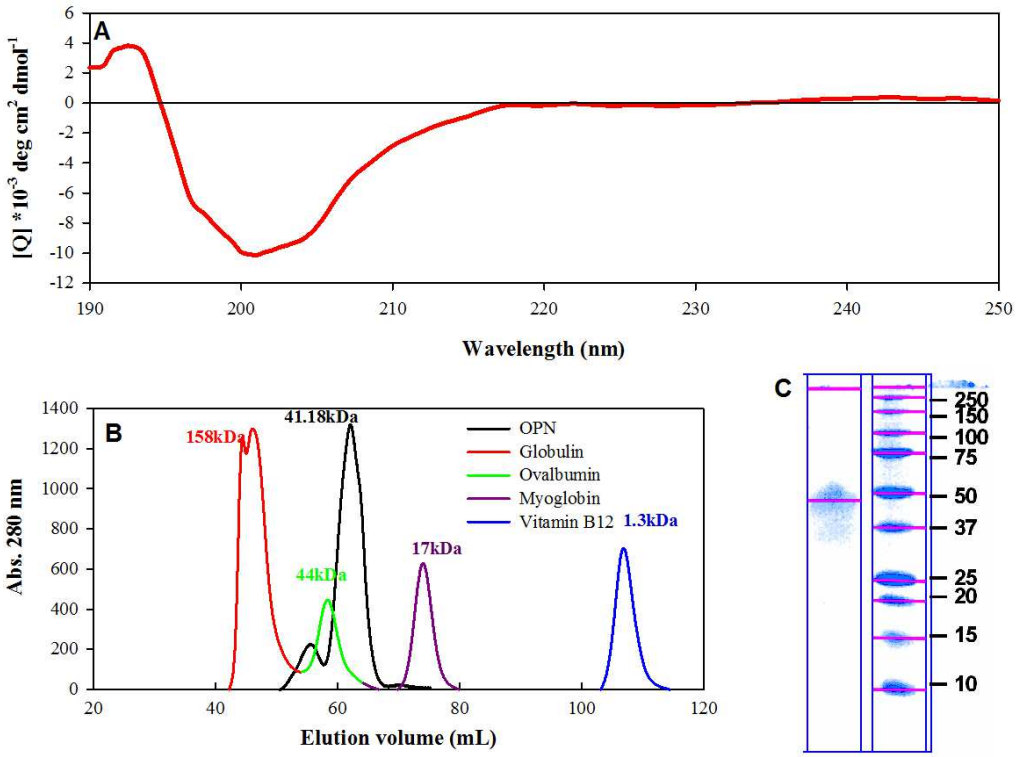

Figure 3.5: **B**iophyiscal characterisation of OPN1-149 **(A)**: Circular dichroism spectra collected for OPN1-149 at 1mg/ml shows a lack of secondary structure elements. **(B)** Analytical size exclusion chromatography indicates that OPN1-149 migrates down a SEC column at a much higher molecular weight compared to the calculated MW. **(C)** OPN1-149 also exhibits unusual migration on a SDS PAGE gel, indicative of an intrinsically disordered protein.

The molecular mass at which OPN 1-149 migrates down the SEC column was obtained by calibrating a column with a set of globular proteins and measuring their retention time (Figure 3.5 B). This analysis yields an apparent molecular mass of $\approx$ 44 kDa for OPN1-149, which is considerably larger then the theoretical molecular mass of $\approx$ 18 kDa based on the amino acid sequence (Figure 3.5 B). A ratio of 2.2 to 2.8

times the apparent molecular mass compared to the theoretical molecular mass is characteristic of disordered proteins [185]. The experimentally determined ratio for OPN1-149 is 2.4, as expected for an IDP. The difference in expected and observed molecular mass is caused by the inability of extended conformations to enter the pores within the stationary phase of the column. This apparent disagreement in observed and expected molecular weight is consistent with SDS-PAGE results, as typically is observed for IDPs [119] . Aberrant migration of IDPs on SDS-PAGE gels is caused by the effects of SDS binding to the characteristically negatively charged proteins. The combined biophysical evidence strongly supports a disordered protein structure for OPN1-149 in solution.

3.3.4 Structural characterisation by SAXS

Experiments were performed at three different concentrations of OPN1-149 (1.5, 3 and 5 mg/ml) to rule out concentration dependent effects. The data were plotted as Guinier plots from which the radius of gyration (R_G) was determined from a linear fit, shown in Figure 3.6. Within error there is no observed effect of concentration on the R_G, indicating no inter or intra particle effects. Therefore concentration dependent effects which are detrimental to the resulting data can be ruled out. The highest concentration measured (5 mg/ml) had the highest signal to noise ratio and was used for all subsequent data analysis.

The R_G of OPN1-149 derived from the Guinier plot was 3.88 $\pm$ 0.26 nm. Predictions on the R_G of disordered proteins based on the amino acid chain length can be made by plotting the Flory equation (Figure 3.7B) [186]. The observed R_G's are consistent with an unfolded protein in solution and scale well with values obtained for other disordered proteins. [157, 187, 188]. Further model independent evidence of the disordered nature of OPN1-149 can be observed in the Kratky plot (Figure 3.7 C). For a folded protein a bell shaped curve with a defined maximum would be expected

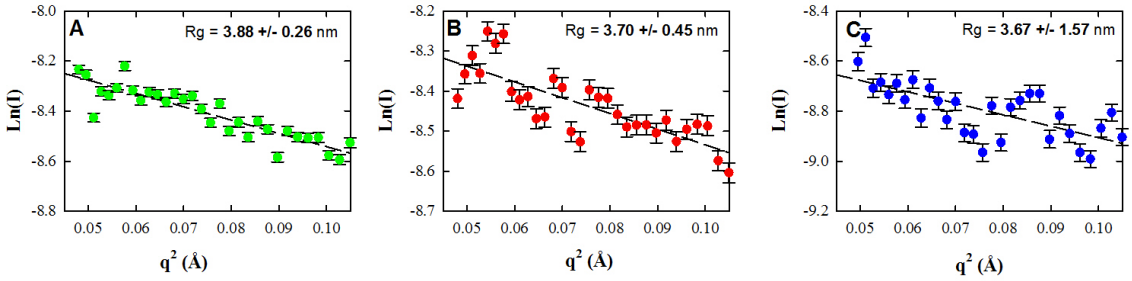

Figure 3.6: Guinier plot of three different concentrations of OPN1-149 collected by SAXS **A)** 5 mg/ml **B)** 3 mg/ml and **C)** 1.5 mg/ml. The radius of gyration was recovered from the slope of the linear fit for each concentration

instead the Kratky plot plateaus at high q. The disordered state of OPN 1-149 is also corroborated by the asymmetric shape of the P(r) distribution, calculated by the indirect Fourier transform method Figure 3.8 [189]. The P(r) distribution function yields a R_G of 3.92 $\pm$ 0.06 nm this value is calculated in real space and agrees well with the value obtained from the Guinier method. The P(r) distribution also provides information on the maximum inter particle distance D_{max} determined to be 12 nm. Overall the model independent parameters obtained by small-angle X-ray scattering suggest a disordered and elongated shape of OPN 1-149 in solution.

3.3.5 SAXS data modelling by *ab initio* methods

Overall molecular envelopes of OPN1-149 in solution were obtained by two *ab initio* approaches, namely DAMMIN and GASBOR. Both modelling approaches use the P(r) calculated from the SAXS data using the indirect Fourier-transform method implemented in GNOM (Figure 3.8). DAMMIN was used to reconstitute scattering envelopes based on the P(r) function. Because the method does not yield unique solution [149, 180], the method was repeated 20 times. A selection of five of the models are shown in green in Figure 3.9, although they each showed slightly different conformations they all displayed a characteristic elongated shape. The normalised

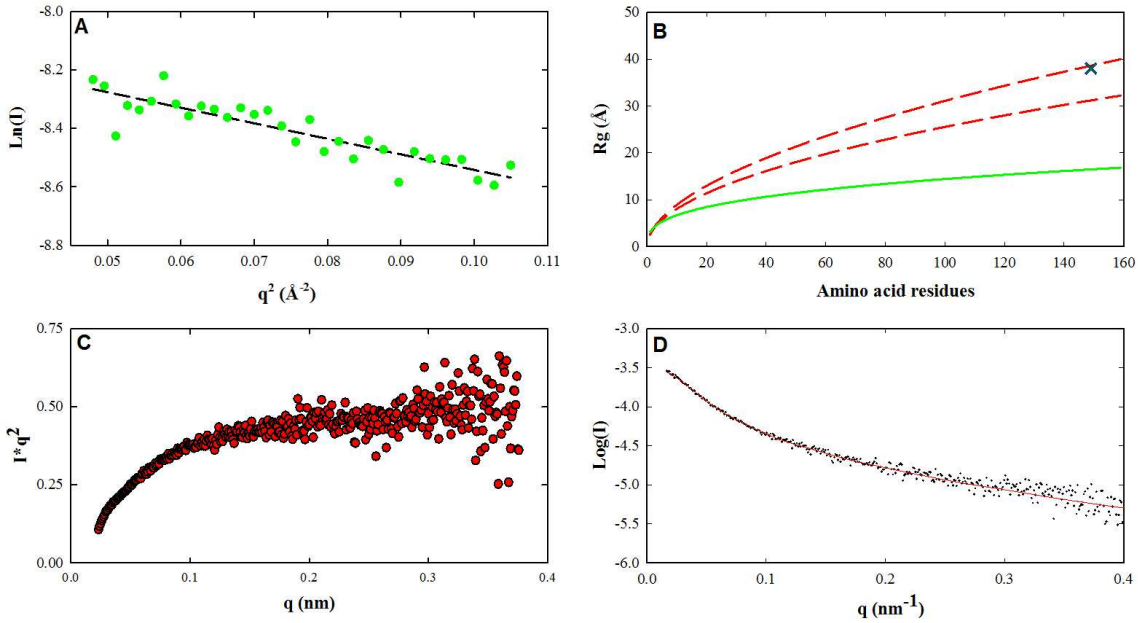

Figure 3.7: A) Guinier plot from which the radius of gyration and I(0) was determined to be 3.882 ± 0.26 nm . **B)** R_G expected for folded proteins and disordered proteins based on protein chain length. Profile of IDP was predicted by the Flory equation. The R_G of OPN1-149 is indicated by the cross, and lies within the region expected for disordered proteins. **C)** Kratky plot of OPN1-149 SAXS data show characteristics of disordered proteins. **D)** Fit of the model produced by DAMMIF (red line) to the experimental data points (black dots) (χ = 1.241.

spatial discrepancies (NSDs), a measure of the similarity between different DAMMIN models was between 0.94 and 1.05 which is considered acceptable for this specific modelling approach [190]. The selected collection of 20 models were averaged to produce the model shown in purple in Figure 3.9.

Using the software DAMFILT, low occupancy areas were removed yielding the final DAMMIF model shown in Figure 3.10. The resulting structure shows an elongated particle, and is similar to structures obtained for other disordered proteins. The low resolution molecular reconstruction of this representation only shows that the conformation of this protein is extended in solution. The resolution obtained by such modelling techniques does not yield information on any secondary structure elements present, however it does provide information on the average structure of the protein

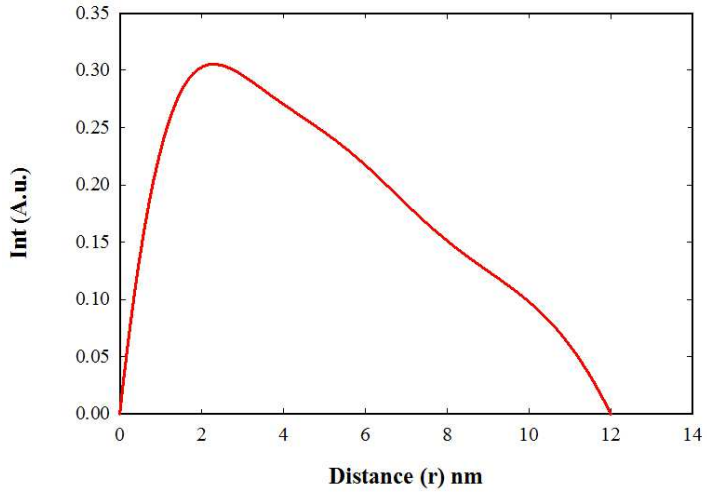

Figure 3.8: The distance distribution function (P(r)) calculated for OPN1-149. The p(r) function shows a large D_{max} at 13 nm and an asymmetric shape characteristic of IDPs. Calculation of p(r) also yields a radius of gyration of 3.92 ± 0.06 nm

adopted in solution. In the case of OPN1-149 the models support the previously obtained evidence that the protein is non globular in solution. Furthermore from the models it can be seen that the protein samples a large area of conformational space. These results follow the lines of the general features obtained for other intrinsically unstructured proteins including *ab initio* SAXS models available in the literature [187].

Ab initio modelling using the dummy residue approach incorporated within the GASBOR software provides qualitatively similar results (Figure 3.11). The computed models provide a good fit to the data, with GASBOR models having slightly higher χ values compared to the DAMMIF models (Figure 3.12). This is expected as the dummy residue approach represents a more realistic description of the different conformers present in a solution of a disordered protein. Further, the scattering envelopes obtained by *ab initio* techniques agree with obtained model independent data from the Kratky plot and the radii of gyration obtained from the p(r) distribution and the Guinier plot that OPN1-149 behaves as an IDP in solution.

Figure 3.9: The average model produced by averaging 20 individual DAMMIN runs (purple) and 5 examples of individual models produced by the DAMMIF software (Green)

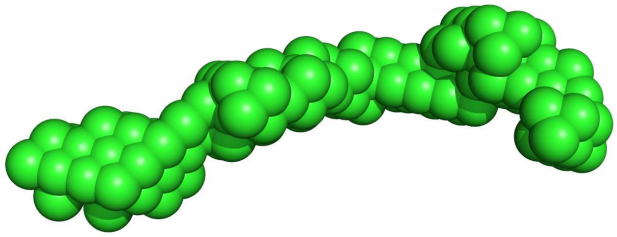

Figure 3.10: Final OPN1-149 model produced by DAMMIF after averaging and filtering excess density

3.3.6 Flexibility analysis by EOM

In solution IDPs consist of a heterogeneous distribution of different conformers, ranging from Gaussian chains to more compact folded conformations. SAXS measures the form factor of a sample, but for IDPs the fluctuations between form factors are much higher than the SAXS time resolution. To extend the information attainable for OPN 1-149, the SAXS data were analysed using the ensemble optimisation method (EOM). The representative ensembles selected by the method to fit the data give information about the preferable conformations of OPN1-149 in solution. The selected EOM ensemble of 20 structures generated for OPN 1-149 produced a good fit to the experimental data with χ of 1.31 (Figure 3.13 A).

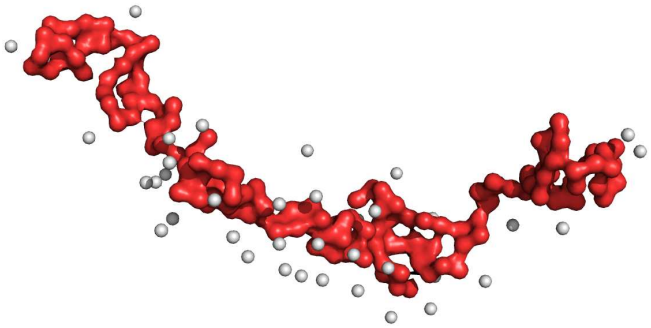

Figure 3.11: Dummy residue model of OPN1-149 produced by GASBOR

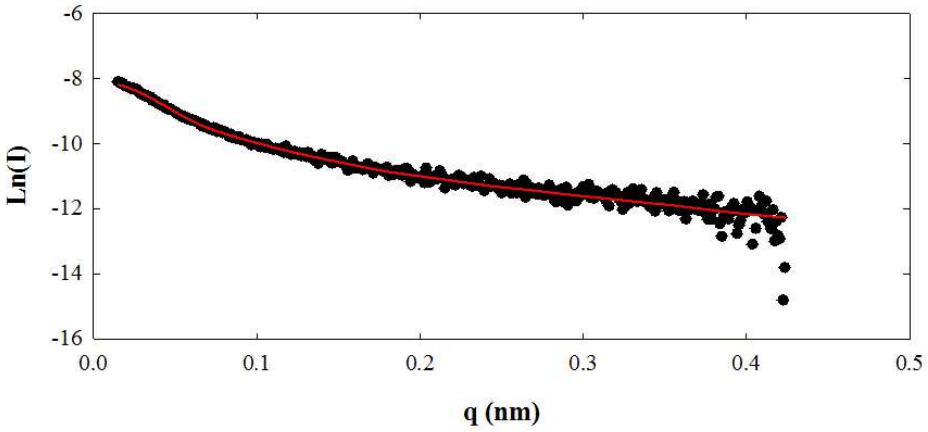

Figure 3.12: Fit of the calculated GASBOR model to the scattering curve, $\chi = 1.84$

Several runs were repeated and yielded a reproducible ensembles shown in Figure 3.13 B + C. The best fit ensemble shows a somewhat bimodal although wide ranging distribution of the R_G Figure 3.13 B. The distribution of the radii peak at higher values than the random pool, indicative of a preference to a somewhat open or extended conformation. The peak for lower R_G however also indicated that the protein exists in a more compact state. This representation of compact and extended structures in solution is also observed in the D_{max} distribution Figure 3.13 C, which shows that the maximum interparticle distance of OPN1-149 ranging from 75 Å up to almost 200 Å. The blue line in Figure 3.13 is the random pool generated assuming a random coil structure. In both the D_{max} and R_G distribution of the generated pool there is a

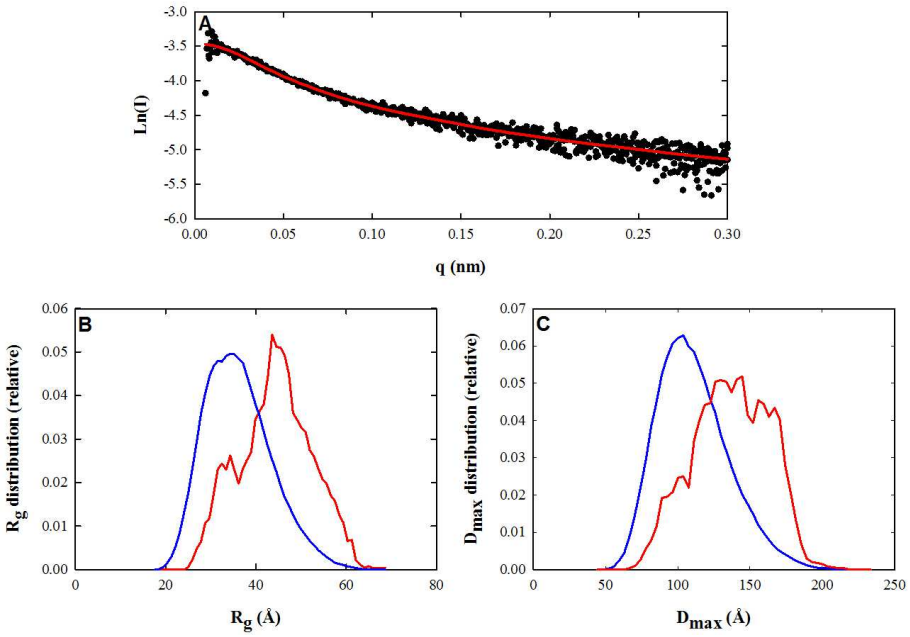

Figure 3.13: EOM analysis of OPN1-149 **(A)** Fit of the selected ensemble (Red line) against experimental data (Black dots) ($\chi = 1.133$). **B** R_G distribution from the selected ensemble (red) and that corresponding to the pool (blue). **C** D_{max} distribution from the selected ensemble (red) and that corresponding to the pool (blue).

smaller peak in these regions. OPN1-149 can therefore sample not only the random coil areas but also more extended structures. The models generated by the EOM method can also be investigated visually. Figure 3.14 shows an example of a compact and extended model from the produced pool. The extended model is visually quite similar to the models produced by GASBOR and DAMMIN. The analysis of SAXS data by EOM provides information about the range of conformations adopted including the more compact conformations that OPN1-149 samples in solution. The results from EOM therefore confirm that OPN1-149 is a disordered protein in solution, capable of sampling a large area of conformational space in by adopting compact and extended conformations. The compact model shown in green corresponds to one of the models present in the peak of the random coil distribution shown in blue in Figure 3.13.

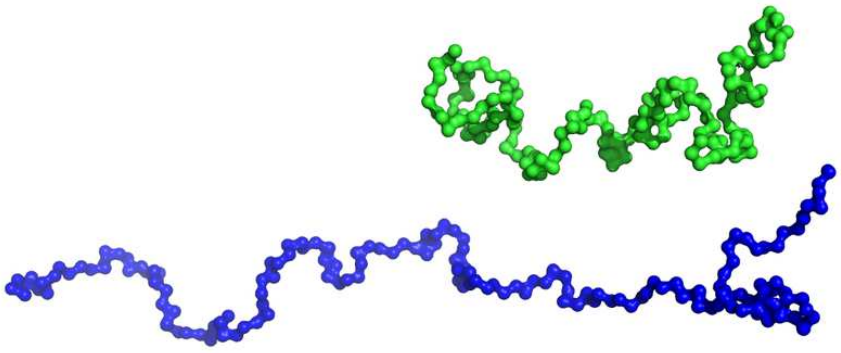

Figure 3.14: Experimentally determined elongated and compact structures of OPN1-149, determined by the EOM method

3.3.7 Characterisation of formed CPN

Dynamic light scattering and small-angle neutron scattering measurements were performed to assess the extent of CPN formation and the solution monodispersity prior to EINS experiments. The DLS results Figure 3.15 A show two main distributions of particle sizes. The hydrodynamic radii of both peaks were determined by the Stokes-Einstein equation yielding hydrodynamic radii of 2.1 nm and 19 nm respectively. These observed hydrodynamic radii correspond with the values expected for both the excess of OPN1-149 peptide and the formed calcium phosphate nanoclusters, as previously measured.

The samples were then further analysed by SANS in order to obtain information about the radius of gyration. The radius of gyration was extracted from the Guinier plot obtained from SANS in H_2O for points that satisfied $qR_G<1.3$ (Figure 3.15). The obtained R_G of 20.5 $\pm$ 0.5 nm, in close agreement with previously obtained results by SAXS which gave an R_G of 21.9 nm, and in agreement with the hydrodynamic radius obtained by DLS measurements. [45, 10]. Some differences in the value of the radius of gyration are to be expected between small-angle neutron scattering and small-angle X-ray scattering because scattering depends on different physical properties (neutron

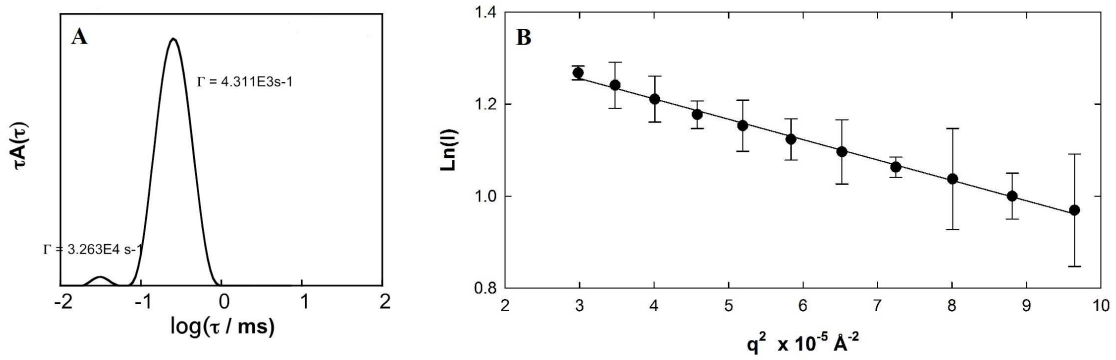

Figure 3.15: A) DLS data of OPN1-149 CPN shows 2 major species in the solution. The larger peak corresponds to the formed CPN with a Rh of 19 nm and the smaller peak corresponds to excess protein in solution with a Rh of 2.1 nm. Guinier approximation of OPN1-149 CPN measured by SANS, yields an R_G of 20.5 ±0.5 nm

scattering length density or electron density, respectively). The results confirm the ability of OPN1-149 to form novel calcium phosphate nanoclusters with a larger radii of gyration then those observed for CPN formed by other proteins such as the caseins. The results from DLS also indicate a single species forming rather then a wide distribution of sizes.

3.3.8 Osteopontin dynamics and the effects of calcium phosphate sequestration.

EINS data were recorded as a function of temperature against the angular dependency of scattering intensity, known as a fixed window EINS temperature scan. A Gaussian fit was applied to the data at low q from which MSDs were extracted from the slope of the straight line. An example of the Gaussian fit is shown for several different temperatures in Figure 3.16. At higher temperatures the linear dependency of the scattered intensity with q deviates more, caused by large amplitude motions. However at q values satisfying the condition in equation 3.5 the Gaussian approximation provides a satisfactory fit at higher temperatures q. From the linear fit the MSD was extracted for each temperature point, for both samples. The resulting data was plotted as a MSD against temperature, shown in Figure 3.17.

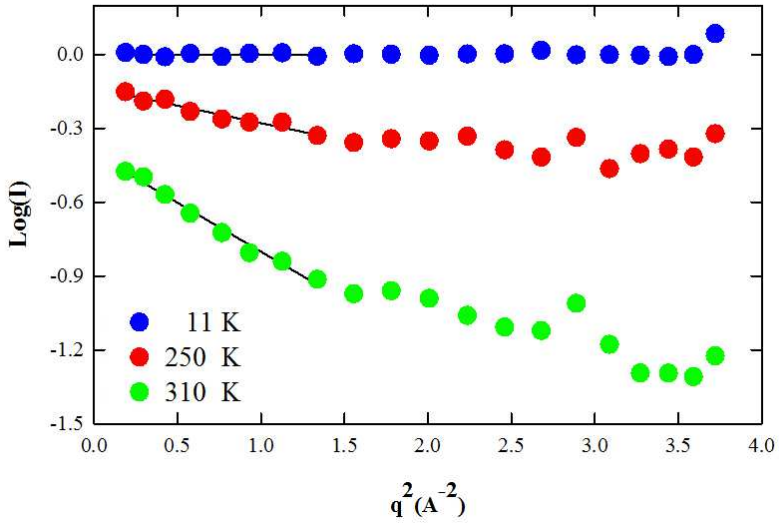

Figure 3.16: Elastic scattering intensity as a function of q measured for OPN1-149 shown for three temperatures, showing the linear fit from which the MSD is obtained.

The results obtained for both free OPN and OPN CPN respectively show typical two well model behaviour, previously observed in other hydrated protein samples examined by EINS [191]. Identical MSDs were obtained at temperatures below 240 K, where atoms are trapped in a low harmonic well. At higher temperatures both the OPN1-149 and OPN CPN start to sample the harmonic and anharmonic wells causing a rapid increase in measured MSDs values. At higher temperatures the MSD of OPN 1-149 is higher than that of OPN CPN, indicative of a higher flexibility of free OPN1-149. When comparing CPN-forming phosphopeptide dynamics, however, the number and distribution of clustered regions of phosphorylated residues (phosphate centres; PC) must be kept in mind. OPN1-149 interacts with the inorganic core of the CPN through phosphorylated regions, which act like anchor points to calcium phosphate, restricting the chain displacements.

The associated effective force constants (<k'>) were extracted from the data above 260K (Table 3.1). The effective force constant of OPN1-149 was determined to be 0.083 ± 0.009 which is in close agreement with the <k'> of Tau protein determined

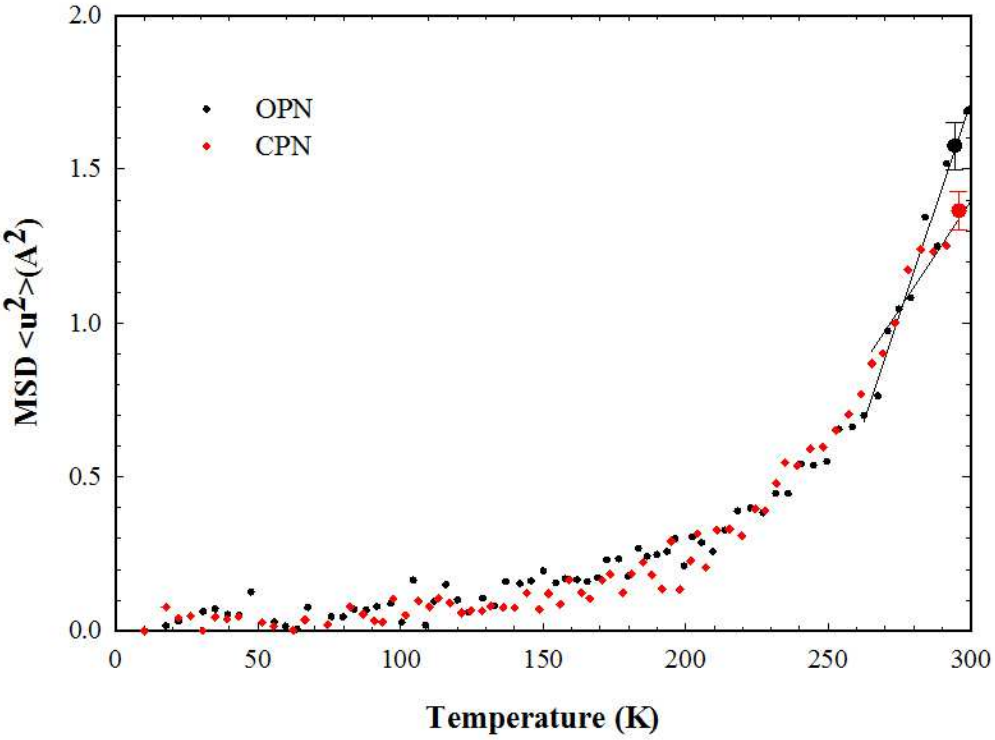

Figure 3.17: Disorder profiles predicted for hOPN1-157 and OPN1-149 by the IUPred method, a dashed red line shows the value above which the region is considered disordered [174].

to be 0.096 by Gallat *et al.* [157]. The EINS measurement from which this value was obtained are identical to the conditions used in this study. Tau protein is also an IDP, and the determined value of <k'> was compared to that obtained for MBP a folded protein of similar molecular mass. The determined effective force constant of MBP was almost twice as large as that determined for Tau Table 3.1 indicative of a stiffer protein. The effective force constant of OPN1-149 is almost twice as large compared to that obtained for OPN CPN. The observed increase in the effective force constant indicates that OPN1-149 becomes more resilient upon sequestration of calcium phosphate, which is combined with a decrease in flexibility. This is consistent with the crowded environment in the CPN, where the 3 PCs further anchor the protein chain. Furthermore table 1 shows that the resilience of hydrated osteopontin

Table 3.1: Values of the effective force constants (<k'>), obtained from the EINS temperature scans on IN16B. Values of Tau and MBP were taken from Gallat *et al.* [157]

Hydrated powder sample	**<k'>** (N/m)
OPN1-149	0.083 ± 0.009
CPN	0.156 ± 0.0190
Tau	0.096
MBP	0.185

is in close agreement with previously measured force constants of the disordered Tau protein. The flexibility of OPN CPN is somewhat reduced yet it is still larger then that observed in globular proteins. Therefore the peptides on the nanocluster shell are still able to occupy a comparably larger conformational space, compared to folded counterparts.

3.4 Discussion

Here a detailed description of the structure and dynamics of OPN1-149 is presented. Analysis of the amino acid sequence showed that OPN1-149 had several of the key characteristics of disordered proteins, including a high net negative charge and an excess of amino acid residues which have a low structural propensity [174]. Circular dichroism analysis showed a low propensity for secondary structure elements in OPN1-149, and other biophysical techniques indicated IDP characteristics of OPN1-149. This result is in close agreement with structural characterisation of other OPN fragments isolated from recombinant and natural sources. Using SAXS in combination with EOM analysis the wide range of conformations adopted by OPN1-149 in solution where determined, ranging from compact to more extended conformations. Similar compact and extended structures have previously been observed by NMR

in a different OPN fragment originating from Quail. Our results build on this data and provide low resolution structural models for the different conformations found in OPN1-149. Adopting a wide range of conformations in solution is a key principle of IDP function, allowing for rapid recognition and weak binding to multiple different partners. In the case of OPN1-149 if the protein is phosphorylated it is capable of sequestering calcium phosphate through highly phosphorylated regions. It has previously been determined that sequestration of calcium phosphate by OPN1-149 leads to the formation of calcium phosphate nanoclusters. It is now known that this sequestration of calcium phosphate is one of the key biological functions of OPN1-149 and similar proteins in biological fluids. In this study the ability of OPN1-149 to form monodisperse calcium phosphate nanoclusters was determined by DLS.

The EINS temperature scan was verified by the transitions observed in all hydrated protein powders studied at 240K. At lower temperatures contributions to the dynamics only arise from solid like vibrational motions. From the data at higher temperatures information about the resilience of the protein was obtained. This resilience is related to the softness of the protein, with a lower value corresponding to a more flexible protein. The resilience of OPN1-149 is lower then those found for globular proteins on the pico-second timescale. This information can be combined with the distribution of models obtained by the EOM SAXS data analysis, to show that in solution OPN1-149 is rapidly sampling a large area of conformational space. In the case of OPN1-149 it can be rationalised that sampling large areas of conformation space could allow the protein to quickly respond to local fluctuations in calcium phosphate concentration, allowing it to sequester excess calcium phosphate prior to nucleation. This rapid sampling of conformational space by IDPs such as Tau and OPN1-149 could be the reason that many proteins that control biomineralisation are IDPs. Upon sequestration of calcium phosphate by OPN1-149 <k'> decreases significantly, and the flexibility of the protein is more accurately represented by that of a globular

protein. This result shows that the specific dynamical and conformational properties of IDP not only depends on their amino acid composition or structure in solution but that the protein environment also has to be considered. This train of thought can be extended to the conditions experienced within the cellular environment or even in biological fluids which are often considerably more crowded then the conditions used to study proteins in the laboratory. For example disordered proteins can undergo disorder to order transition upon binding with their partners but can also retain substantial conformational flexibilities as shown here for OPN CPN. Given that previous work showed that the flexibility of OPN increases upon binding to heparin, structural and dynamical changes are strongly dependant on the molecular partner [192]. For both calcium phosphate and heparin, the order or disorder induced is not described by formation of secondary structural elements but rather OPN peptide occupation of conformational space. Such behaviour of IPDs has previously been simulated by computational studies, which showed that under crowding conditions the calculated R_G of unfolded proteins decrease, without induction of structural elements [173].

3.5 Conclusion

Several different biophysical characterisations showed that OPN1-149 behaves as an IDP in solution. This result was confirmed by solution SAXS studies and *ab initio* modelling. By using the ensemble optimisation method further information about the ensemble of OPN1-149 in solution was obtained. In solution the protein can adopt a large amount of different conformations, ranging from compact to highly extended. By combining the structural ensemble data and the data obtained by EINS it was shown that OPN1-149 more rapidly samples different conformations compared to globular proteins. And that these transient conformations can range from quite compact to highly extended. The ability of OPN1-149 to sequester calcium phosphate leading to

the formation of calcium phosphate nanoclusters was confirmed by DLS and SANS. Allowing the peptide to regulate biofluids supersaturated with respect to hydroxyapatite. The ability to sample large areas of conformational space could be vital in the process of sequestering excess calcium phosphate prior to irreversible nucleation and crystal growth. Data is also presented that shows an increase in the resilience of OPN1-149 upon formation of calcium phosphate nanoclusters and is similar to the resilience found in globular proteins. This increase in resilience could be caused by the tight packing of OPN1-149 around the particle core and could be of benefit for further CPN interactions [193]. This increase in resilience has important insights for the dynamics of intrinsically disordered proteins. Most structural studies of intrinsically disordered proteins are on monodisperse dilute samples in solution. The conditions under which biophysical experiments are performed make it difficult to understand the impact of binding partners and cellular conditions on protein dynamics and structure. It is for example well known that IDPs are more effected by the crowded protein conditions found in the cell, leading to adoption of more compact states.The complex conditions found in the cell and biological fluids could therefore also effect the dynamics of other intrinsically disordered proteins, not only in terms of induced folding but also in restriction of conformational space sampling. Recent neutron scattering experiments on highly concentrated globular protein solutions representing those found in cellular conditions show that that sidechain rotations are reduced at high crowding concentrations. Given their unique structural features, disordered proteins may respond to these factors differently, especially in terms of compaction.

Chapter 4

Calcium phosphate nanoclusters stabilised by osteopontin contain a core of amorphous calcium phosphate sequestered within a protein shell

Experimental work presented in the following chapter has been peer reviewed and published:

- Holt, C., **Lenton, S**., Nylander, T., Sørensen, E. S., Teixeira, S. C. M. (2014). Mineralisation of soft and hard tissues and the stability of biofluids. Journal of structural biology, 185(3), 383-396.

Abstract

In vertebrates insufficient availability of calcium and phosphate ions in biofluids leads to loss of density in mineralised tissue. Consequently concentrations of calcium and phosphate ions are above their respective levels of saturation. The same biofluids often permeate soft tissue without causing ectopic calcification. Several proteins have been shown to allow soft and hard tissue to coexist by sequestering calcium phosphate within a calcium phosphate nanocluster (CPN), thereby stabilising saturated biofluids. Osteopontin (OPN) is an example of a protein able to form CPN. The CPN stabilised by OPN are larger then those formed by other proteins. Here we present a detailed characterisation of the CPN formed by OPN proteins. Using small-angle neutron scattering we determine that the formed CPN have a core-shell type structure. Using neutron diffraction a distinct lack of crystallinity is observed within the core, indicating an amorphous structure of the sequestered calcium phosphate. The biological implications of the findings are discussed in context of the role of OPN in biomineralisation.

4.1 Introduction

In vertebrates mineralised tissue consists of a protein matrix within which crystalline calcium phosphate is deposited. Insufficient availability of calcium and phosphate ions in biological fluids that are in contact with mineralised tissue can lead to the loss of mineral density [9]. Consequently most biofluids contain calcium and phosphate ions above their soluble concentrations [194, 12]. At physiological pH a solution saturated with calcium and phosphate ions is expected to precipitate out of solution [14]. The precipitation and consequent formation of crystalline material is a thermodynamically driven process [74]. The process is initiated by the formation of a short lived amorphous calcium phosphate (ACP) phase, which is metastable at best, the fresh precipitate of ACP is then transformed through a number of transient calcium phosphate phases, before the final most stable phase of crystalline hydroxyapatite (HA) is formed [17, 195, 196]. In vertebrates HA is the mineral phase found in most mineralised tissue such as bone and tooth [4]. It would therefore be expected that biofluids would readily precipitate HA and readily mineralise soft tissue with which they are in contact. However precipitation of calcium phosphate in biofluids is not spontaneous and degenerate ectopic calcification is a relatively rare occurrence [1]. Several phosphorylated proteins have been shown to inhibit the precipitation of calcium phosphate from solution [23]. These proteins inhibit precipitation of calcium phosphate by forming equilibrium complexes, a vital requirement for the formation of such complexes is the presence of phosphate centres. A phosphate centre (PC) is defined as a short acidic subsequence within which at least three or more residues are phosphorylated [10]. If these phosphate centres are present in stoichiometric excess over the amount of initial precipitate of amorphous calcium phosphate the solution remains stable and no precipitate is formed [74]. The stability is afforded through the reaction of the phosphate centres with the formed precipitate of amorphous cal-

cium phosphate which leads to formation of calcium phosphate nanocluster (CPN) [74]. It is thought that calcium phosphate nanoclusters contain a core of amorphous calcium phosphate surrounded by a peptide shell [45]. The sequestration of calcium phosphate by PC containing peptides completely inhibits the maturation of the calcium phosphate and the core reaches equilibrium with the solution. In bovine milk the PC containing β-casein is responsible for stabilising the high concentrations of calcium phosphate through formation of CPN. The CPN formed by β-casein contains an amorphous core of calcium phosphate surrounded by a protein shell [67, 73]. Osteopontin (OPN) is another PC containing peptide found in bovine milk formed by the cleavage of full length OPN by the secreted plasmin protease, OPN1-149 contains three phosphate centres [45, 64, 197].

Table 4.1: Identified PC sequences formed by phosphorylation of secreted phosphoproteins by Golgi Kinase and Casein Kinase 2, with phosphorylated sites shown in bold. Table adapted from Holt *et al.* [72]

Protein	**Swiss-prot No**	**PC sequence**
OPN	P31096	6 - T**SS**G**SS**EEKQ - 15
Casein	P045656	19 - **SSSSSSS**EER - 28
Fetuin-A	P02765	307 - **S**LG**S**P**S**GEVS - 316

SANS experiments performed on the CPN stabilised by the single PC containing β-casein showed that the nanocluster can be described by a core shell model with a core radius of 2.30 $\pm$ 0.05 nm and a total radius of 4.04 $\pm$ 0.15 nm. SAXS experiments on OPN1-149 CPN showed a much larger radius of around 20 nm with a predicted core radius of 12.5 nm. However the SAXS results do not provide direct evidence of this core-shell structure, instead the core-shell model is imposed on the data [72]. The larger radius of the core was thought to be mainly due to hydration and the

increased sequestering ability of the OPN1-149 peptide afforded by a higher degree of phosphorylation [72]. The increased ability to sequester calcium phosphate is thought to be due to the higher concentration of hydrophobic sequences around the phosphate centre together. However, as of yet, there is no experimental evidence that the core of OPN1-149 CPN is amorphous, the larger core radius could also be due to crystal growth within the core.

4.1.1 Amorphous phases of calcium phosphate

Structures lacking in any medium or short range order are more closely related to the structures of liquid phases then those of crystalline material. They are named amorphous phases, from the Greek amorphos meaning without form or shapeless. The structures are highly irregular such that the concept of a reference crystal lattice is nonsensical. Instead an amorphous phase is more accurately described by a distribution of different structures, or ideally a perfectly random structure [198]. Amorphous calcium phosphate phases are the first phases of calcium phosphate to form from a supersaturated solution above pH 7. The formation of amorphous calcium phopshate is thermodynamically driven, however the formed amorphous structure is intrinsically unstable or metastable at best. The initial phase of ACP is spontaneously converted to more stable crystalline calcium phosphate phases. The transient nature of amorphous phases has been observed by electron diffraction patterns where over time the formation of a crystalline phase was observed originating from an amorphous diffraction pattern [199].

4.1.2 Crystalline phases of calcium phosphate

Crystalline solids are characterised by their regular, periodic atomic arrangement in three dimensional space. Periodic arrangement in structures can be described over different length scales. The crystalline state of different phases of calcium phosphate

can be described over the Ångstrom length scale (see Table 4.2). This length scale corresponds to shortest length scale used to describe biological material which typically lies around the order of Ångstroms which roughly corresponds to the intermolecular bond length. Nearly all liquids and solids can be said to contain order over these short length scales, so called short-range order. Crystalline solids contain structural order over much longer length scales, so called medium range order and also long-range order. The length scales of the three different scales of order have been defined in literature, Short-range consists of 2-5 Å, medium range order of 5-20 Å and long range order exceeds 20 Å [200, 201].

Table 4.2: Crystallographic data of different calcium phosphate phases. [202] MCPM = monocalcium phosphate monohydrate, MCPA = monocalcium phosphate anhydrate, DCPD = dicalcium phosphate dihydrate, OCP = octacalcium phosphate, α-TCP = α-tricalcium phosphate, β-TCP = β-tricalcium phosphate.

Compound	**Space Group**	**Unit cell parameters (Å)**
MCPM	Triclinic	$a = 5.3, b = 11.9, c = 6.5$
MCPA	Triclinic	$a = 7.6, b = 8.3, c = 5.6$
DCPD	Monoclinic	$a = 5.8, b = 15.2, c = 6.2$
OCP	Triclinic	$a = 19.6, b = 9.5, c = 6.8$
α-TCP	Monoclinic	$a = 12.9, b = 27.3, c = 15.2$
β-TCP	Rhombohedral	$a = b = 10.4, c = 6.9$
Hydroxyapatite	Monoclinic	$a = 9.8, b = 2a, c = 6.9$

It is this repetitiveness which is detected and detailed by the manner in which radiation is scattered by a crystalline sample. It is therefore possible to obtain information about the structure of the scattering crystal by investigation of the diffraction. In order to study structural features using radiation the incident beam has to probe

distances on the order of the periodicity, which for calcium phosphate phases on the the Ångstrom scale (see Table 4.2). Both X-rays and neutrons can be obtained on this order of magnitude. For X-rays the wavelength of the radiation is tuned using collimator's whereas for neutrons the corresponding wavelength is obtained by regulating the temperature of the neutron beam, for example by interaction with D_2O. X-ray diffraction has several benefits over neutron diffraction including, easier access to beam-time, faster data collection and smaller sample sizes. The interaction of high energy X-rays with soft matter such as protein and crystalline material can induce radiation damage in the sample. In some cases synchrotron radiation is known to alter the structure of the studied sample in both proteins and crystalline latices. Neutrons require much longer measurement times and larger samples, but the interaction of neutrons with the atomic nucleus is non destructive meaning that samples are not damaged. Diffraction studies using neutrons also allow the experimenter to take use contrast variation. This technique exploits the interaction of neutrons with different isotopes thereby allowing the match out of parts of the sample with different scattering length densities. Using this technique for example it is possible to study only the core of a calcium phosphate nanocluster without observing contributions from the protein shell.

In this chapter a combination of small-angle neutron scattering with contrast variation and static light scattering is used to show that OPN1-149-CPN formed in supersaturated solutions of calcium phospate are core shell particles. The core is then characterised using a combination of neutron diffraction and FT-IR spectroscopy. The implications of the results obtained are then discussed in terms of the biological function of OPN1-149 and OPN1-149-CPN.

4.2 Materials and methods

4.2.1 Protein purification and preparation

In milk OPN undergoes proteolytic processing by several different proteases, leading to a mixture (OPNmix) of different phosphopeptides [183]. For this study, naturally occurring OPNmix was isolated from bovine milk as previously described and used without further purification [45]. The principal components are N-terminal peptides ending between residues 145 and 153 of the mature protein sequence [197]. Previously this group of N-terminal peptides was considered equivalent to OPN 1-149. We have kept this nomenclature here for the sake of simplicity and continuity. Purification of OPN1-149 from OPNmix was completed per the method of Holt *et al*. [72]. From the OPNmix the OPN1-149 peptide was purified to homogeneity by size exclusion chromatography (SEC) as described by Holt *et al* [45]. In brief, osteopontin fragments were separated by size on a 120 ml volume column containing a Superdex 75 (G.E. Healthcare) stationary phase, connected to a Biologic dual flow purification system (Bio-Rad). Size exclusion was performed at room temperature, the column was equilibrated with 2 column volumes of elution buffer (50 mM phosphate, pH 7.0, 300 mM NaCl, 0.02% NaN_3) at a flow rate of 0.3 ml/min. 100 mg of OPNmix was then dissolved in the same elution buffer and loaded onto the column at a rate of 0.3 ml/min. Elution from the column was constantly monitored by U.V. absorption at $\lambda = 280$ nm, fractions were collected on volumes of 1 ml, elution was stopped after 2 column volumes of elution buffer.

4.2.2 Formation of calcium phosphate nanoclusters

OPNmix and OPN1-149 CPN were prepared by the simple mixing method to give final concentrations of 22 mM calcium, 20 mM inorganic phosphate, an ionic strength of 63 mM and pH 7.0. The nanocluster equilibrium has a free calcium ion concentration

of 0.5 mM and, on average, each peptide is attached to calcium phosphate through only one of its 3 phosphate centres. Samples were allowed to equilibrate for at least 14 days prior to centrifugation. Two hydrated samples were produced from solutions of 10 ml of 30 mg/ml OPNmix CPN using 39% D_2O (contrast match point of the peptide) or 100% D_2O.

4.2.3 Static light scattering experiments

Experiments were performed on a CGS-3 Compact Goniometer System (ALV, Langen, Germany) using a 22 mW Helium-Neon laser operating at 632.8 nm. To prevent multiple scattering the sample was diluted 50 fold with a buffer that matched the ultrafiltrate to maintain sample integrity. Prior to measurement, the sample and dilution buffer were filtered through a Sartorius syringe filter of 0.2 μm porosity into a QS cylindrical cell with an optical path length of 8 mm (Hellma, Mullheim, Germany). Measurements were made at scattering angles between 30° and 150° in intervals of 10° at a temperature of 24°C. After subtracting the dilution buffer scattering, the radius of gyration (R_G) of the OPNmix nanoclusters in 0 % and 39 % D_2O was recovered from the slope of the Guinier plot.

4.2.4 Small-angle neutron scattering experiments

Neutron scattering measurements were made at 24 °C on the small-angle neutron diffractometer D33 [152] at the Institut Laue-Langevin, Grenoble. Prior to each measurement, samples containing 30 mg/ml OPNmix were filtered through a 0.2 μm Sartorius syringe filter into Hellma quartz cells of 1 mm optical path length. Scattering curves were recorded using a fixed incident neutron wavelength of 0.6 nm and wavelength spread $\left(\frac{\Delta\lambda}{\lambda}\right)$ of 10 %. A Guinier plot was made of the small-angle neutron scattering in the scattering wave vector (q) range 0.00408 - 0.01281 Å. Data satisfying the condition $q \cdot R_G \leq 1$ were fitted to a straight line to recover the value of

the radius of gyration.

4.2.5 Model calculations

The core-shell model used originates from the work of Holt *et al.* [67]. Total scattering measured P arises from the difference between excess scattering of the core and shell P_1 and P_2 above the scattering of the solvent P_s.

$$P = P_1 + P_2 + P_S \tag{4.1}$$

The non solvent scattering material consists of contributions from the core and shell which have radii R_1 and R_2, respectively. According to Holt *et al.* [67] the R_G for a core-shell particle is given by:

$$R_G^2 = \left(\frac{3}{5}\right) \cdot \left(\frac{[(P_2 - P_S)\, R_2^5 + (P_1 - P_S)\, R_1^5]}{[(P_2 - P_S)\, R_2^3 + (P_1 - P_S)\, R_1^3]}\right) \tag{4.2}$$

For a neutron scattering experiment the scattering length densities of the core and shell contributions can be calculated, and depend upon the concentration of D_2O, β, in the solvent. Assuming full exchange of the buffer

$$P_s = -0.562 + 6.962\beta \tag{4.3}$$

$$P_1 = 2.019 + 4.043\beta \tag{4.4}$$

$$P_2 = 1.901 + 1.011\beta \tag{4.5}$$

From which it can be calculated that $P_1 = P_S$ at $\beta = 0.882$, $P_2 = P_S$ at $\beta = 0.413$ and $P_1 = P2$ at $\beta = -0.039$.

4.2.6 Preparation of micellar calcium phosphate

The preparation of micellar calcium phosphate from the casein micelle of bovine milk followed the method described by Holt *et al.* [203]. Casein micelles were recovered from 40 ml of commercial skimmed milk by centrifugation at 60,000 g for one hour, the resulting supernatant was disposed and the pellet re-suspended in the same volume of milk. Sodium azide was added to the resulting solution to a final concentration of 1.5 mM. The solution was left to stir for 24 hours at 4°C until the solution was fully re-dispersed. 100 mg of proteinase K was added to the solution and placed in a Visking dialysis tube. The resulting solution was then dialyzed for 24 hours against 2 litres of skimmed milk also containing 1.5 mM sodium azide at 4°C. The resulting solution in the dialysis bag cleared. Casein micelles where then recovered by centrifugation (85,000 g, 16 hours at 8°C) which yielded a clear brown pellet. Previous work has shown that the pellet comprises the micellar calcium phosphate and associated casein phosphopeptides resistant to digestion by the proteolytic enzyme [203, 204].

4.2.7 Preparation of calcium phosphate samples

Amorphous calcium phosphate was prepared using a modified method of Holt *et al.*, by including the tryptic digest of β-casein, β-casein 1-25, which inhibits the maturation of the formed amorphous calcium phosphate [10, 205]. Briefly, 500 ml of a phosphate solution was mixed with 500 ml of a calcium solution at room temperature. The phosphate solution contained 100 mM NaH_2PO_4 and 100 mM Na_2HPO_4, the calcium solution contained 4 mM $CaCl_2$, 135 mM NaCl, 14 mM $MgCl_2$ and 10 mg of the bovine casein tryptic digest β-casein 1-25. The pH of the resulting solution was adjusted and maintained to 7 by addition of NaOH whilst stirring. The solution was filtered through a 0.44 μm Sartorius membrane after 60 minutes of stirring, the obtained solid was then freeze dried and kept over silica powder overnight and finally

stored at -20°C. Dicalcium phosphate dihydrate (DCPD) was prepared according to a previously published protocol. In brief, 0.825 g of KH_2PO_4 and 3 g Na_2HPO_4, to which 200 ml of a solution containing 4 g of $CaCl_2 \cdot 2H_2O$ was added whilst stirring . The solution was left to stir at room temperature for 60 minutes resulting in the formation of a precipitate. The solid was recovered by filtration through a 0.44 μm Sartorius membrane and washed several times with de-ionised water. The obtained solid was left to dry overnight on silica gel before storage at -20°C.

4.2.8 Simulated neutron and X-ray diffraction patterns

The detection of crystalline material within a sample depends upon many parameters including the number, purity and size of crystalline material present as well as instrumental parameters which are often quite complex. Many of these different parameters are taken into account in the FullProf suite thereby allowing for an accurate prediction of neutron scattering by different crystalline material [206]. Diffraction patterns were calculated for different crystalline sizes of HA and DCPD, within a range corresponding to the sizes present in different calcium phosphate nanoclusters. Neutron and X-ray diffraction patterns where also predicted using the Inorganic Crystal Structure Database (ICSD-for WWW) hosted at the Institut Laue-Langevin [207].

4.2.9 FT-IR experiments

FT-IR analysis was performed with a Bruker VERTEX 70 (Rheinstetten, Germany) infrared spectrometer equipped with a Platinum ATR device. Data was recorded with a resolution of 2 cm^{-1} in the range 4000 - 400 cm^{-1}, with a total of 32 scans. The spectra were recorded and stored using spectroscopic software (OPUS, Bruker Inc.).

4.2.10 X-ray diffraction experiments

X-ray diffraction data were collected using a GeniX Cu high flux generator (1.5418 Å), average flux for this configuration was 400x106 photons.s^{-}1, and recorded on a Mar345 detector with exposure times of 20 minutes per sample. Each dried or hydrated sample was mounted in a glass capillary of 0.5 mm diameter (beam size at focus 200x200 μm^2 with a divergence of 5 mrad).

4.2.11 Neutron diffraction experiments

Diffraction measurements were collected at the D16 diffractometer at the Institut Laue-Langevin, Grenoble France [208]. Data were recorded over the q range of 0.01-2.5 Å^{-1} by rotation of the detector around the sample position. Where q is the scattering wave vector described by $q = (4\pi/\lambda)sin\theta$ where λ is the wavelength of the incident neutron beam in this case 4.76 Å($\Delta\lambda/\lambda = 10\%$). Measurements were made corresponding to a 2θ range of 30 - 115° at this range the D16 instrumental neutron diffraction peak broadening corresponds to a maximum of 0.83° at the wider angles (where 0.5° is the maximum beam bandwidth in the range). Diffraction was recorded on a millimetre resolution large area 3He detector (MILAND, 320 × 320mm^2). The incident beam had a diameter of 3mm, the divergence was calculated from the image of a attenuated beam and determined to be 0.33°. During data collection the incident flux of neutrons on the sample was 1×10^6 neutrons cm^{-2}s^{-1}. Samples were deposited in sample containers with a 5mm path-length. Empty cell scattering was recorded and subtracted from the sample scattering pattern. Data was collected on OPNmix-CPN samples prepared in H_2O, 100% D_2O and 41% D_2O. The samples were pelleted after 3 weeks of formation to allow them to reach equilibrium size [45]. Diffraction measurements were repeated 4 weeks later, no changes in diffraction were observed. While the hydrated samples were pelleted to obtain a higher concentration within

the beam the dry powdered samples were deposited in the sample container without compaction. Data was collected at room temperature with exposure time of 120 minutes per sample. Data reduction was carried out using the Large Array Manipulation Programme (LAMP) of the ILL.

4.3 Results

4.3.1 Match-point of OPN1-149

The match-point of OPN1-149 was calculated to be around 40% D_2O. This was verified experimentally by measurement of the protein in buffer containing 40% D_2O, shown in Figure 4.1

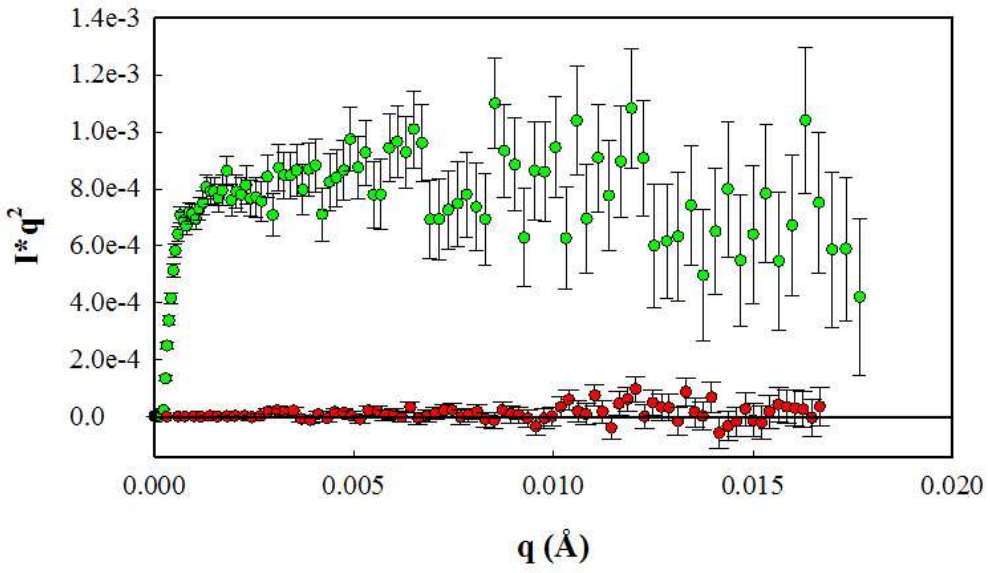

Figure 4.1: Kratky plot of OPN1-149 measured in 100% D_2O (green circles) and 40% D_2O (red circles). In the 40% D_2O there is no scattering above noise, verifying that OPN1-149 is matched out.

4.3.2 CPN stabilised by OPN have a core-shell structure

The formation of calcium phosphate nanoclusters by OPNmix, in both 0% and 39% D_2O, was observed in the Guinier plot of the static light scattering data. The R_G was determined from the slope of a straight line fit for OPNmix-CPN in 0% and 39% D_2O, to be 24.5 ± 0.3 nm and 24.5 ± 0.48 nm respectively. The results indicate that the increased levels of D_2O do not affect the overall size of the formed CPN and that the CPN can still form even under higher percentages of D_2O. The R_G determined from the SANS for OPNmix-CPN in 0% D_2O was 19.5 ± 0.36 nm, which compares well with the R_G of 20.5 ± 0.1 nm, previously determined by small-angle X-ray scattering [45]. The slight differences observed between the presented light scattering and small-angle neutron scattering data are expected, because for each method the scattering depends on a range of physical properties. The small-angle neutron scattering data in 39% D_2O, close to the protein match-point, yielded a lower R_G of 17.5 ± 0.51 nm, confirming that the nanoclusters have a core-shell structure. The data was then modelled according to the core-shell structure model (See quation:4.2), from which the core and shell radii can be calculated from the experimentally determined R_G. In 0% D_2O the radius of the whole calcium phosphate nanocluster is then 25.2 ± 0.44 nm and the radius of the core determined from the 39% D_2O sample is 22.6 ± 0.68 nm, corresponding to a shell thickness of 2.6 nm (Table 4.3).

Further studies were completed using purified OPN1-149 for which a more accurate calculation of the core shell model can be completed. The Guinier plot of the SANS data of OPN1-149-CPN is shown in Figure 4.3. The R_G obtained for the 0% D_2O and 39% D_2O sample was 18.1 ± 0.29 nm and 19.5 ± 0.36 nm, respectively. The core-shell model calculation was also applied to the the data and is shown in table Table 4.3. From the model a shell thickness of 2 nm is obtained, slightly smaller compared to the shell thickness of OPNmix-CPN. The Kratky plot of both the 0 % and 39 % OPN1-149-CPN is shown in Figure 4.4. Both Kratky plots show a well

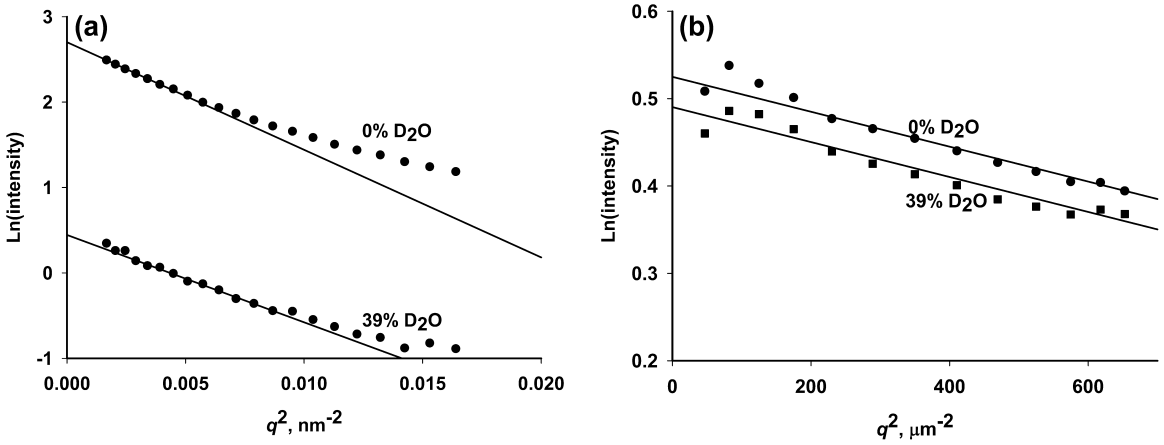

Figure 4.2: A) Guinier plot of SANS from OPNmix-CPN in 39% and 0% D_2O **B)A Guinier plot of Elastic light scattering data from OPNmix-CPN in 39% and 0% D_2O**. The radius of gyration was recovered from the fit of a straight line to the points.

defined maximum, which is at lower q for the non matched out sample, corresponding to a larger particle. This type of behaviour where the core is smaller then the whole particle corresponds to the core-shell mode.

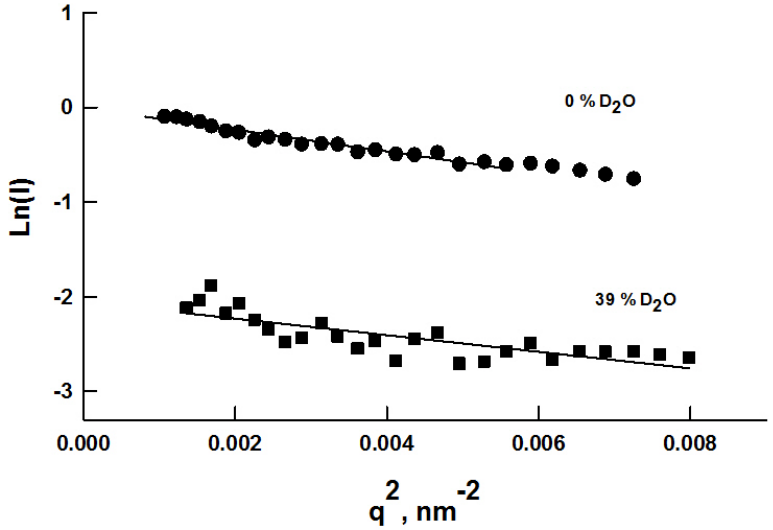

Figure 4.3: Guinier plot of the small-angle neutron scattering data from OPN1-149 CPN in 0 % and 39% D_2O. The obtained R_G where 18.1±0.29 nm and 16.5±0.60 nm respectively.

Table 4.3: Radius of gyration obtained for OPNmix and OPN1-149 CPN by SANS. Core R_G measurements were performed at the match point of protein (41% D_2O). CPN R_G measurements were obtained from SANS measurements in 0% D_2O. The calculations of the Core and shell radius are based on the model presented in equation 4.2. All units are nm unless stated otherwise

	Core R_G	**Core radius**	**CPN R_G**	**CPN radius**
OPNmix	17.5 ± 0.51	22.6	19.5 ± 0.36	25.1
OPN1-149	16.5 ± 0.60	21.4	18.1 ± 0.29	23.4

4.3.3 Characterisation of calcium phosphate phases

In order to confirm successful synthesis of the control samples, ACP and DCPD, were initially characterised by FT-IR. The experimentally determined FTIR spectra of the control samples are shown in Figure 4.5 and are in close agreement with previous experimental observations [209, 210]. The structure of DCPD is indicated by broad bands centred around 2850 cm^{-1} corresponding to the hydrogen phosphate anions is present in the spectra [211]. A pair of doublets can be observed around 3480 cm^{-1} and 3200 cm^{-1} caused by the water stretching mode. Neither of these peaks are observed in the ACP or dry CPN sample (Figure 4.5). The characteristic water bending mode at 1649 cm and it's subsidiary peak at 1730cm $^{-1}$, indicative of DCPD are also present. The expected lifting of the degeneracy of the broad phosphate asymmetric stretch (1100 cm^{-1}) for crystalline calcium orthophosphates, due to site symmetry of factor group splitting is observed in the DCPD sample but is not present in either the ACP or CPN spectra. The measured spectra of ACP is also in close agreement with previously published results for ACP prepared in a similar manner (Figure 4.5) [212]. A large part of the measured spectra is obscured by the presence of water, specifically in the region ranging from 2500-3650 cm^{-1}. The strong peak observed at around

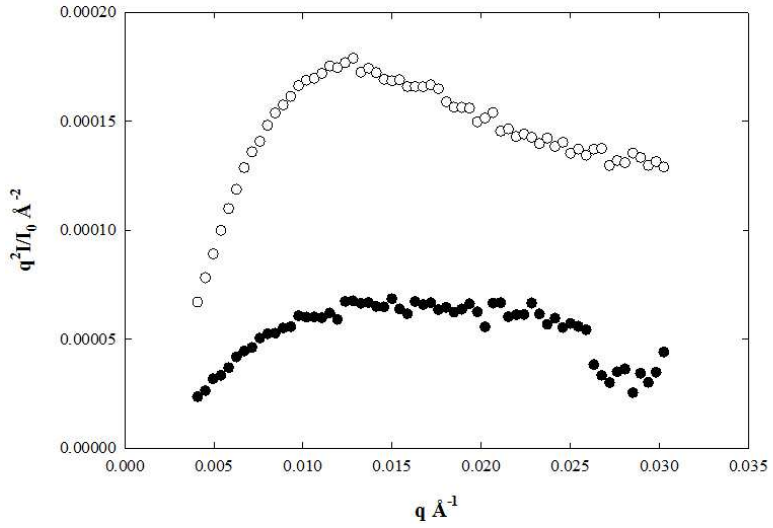

Figure 4.4: Kratky plot of the small-angle neutron scattering data from OPN1-149 nanoclusters. Filled circles shows OPN1-149 CPN in 39% D_2O and empty circles shows OPN1-149 CPN in 0% D_2O

850 cm^{-1} is caused by the stretching mode of protonated phosphate groups, which is absent in the dry CPN sample (Figure 4.5) this is probably due to it being obscured by other absorption bands as previously observed in FTIR of CPN samples [73]. The FT-IR results therefore indicate formation of an amorphous phase of calcium phosphate and successfully synthesis of DCPD. Moreover the dried OPN-CPN sample shows a FT-IR spectrum similar to those obtained for amorphous calcium phosphate.

4.3.4 Calculation of neutron and X-ray diffraction patterns

Neutron diffraction patterns were simulated over the q-range measured on D16 for different calcium phosphate phases (Figure 4.7). All calcium phosphate phases simulated show diffraction under the measurement conditions used on D16. The simulated neutron diffraction patterns of DCPD and HA for different possible crystalite sizes are also shown in Figure 4.7. The crystallite sizes were chosen to cover a range from the dimension equivalent to just one or two stacked unit cells, up to a size that would occupy the full diameter of the core of CPN. DCPD and HA were chosen as represen-

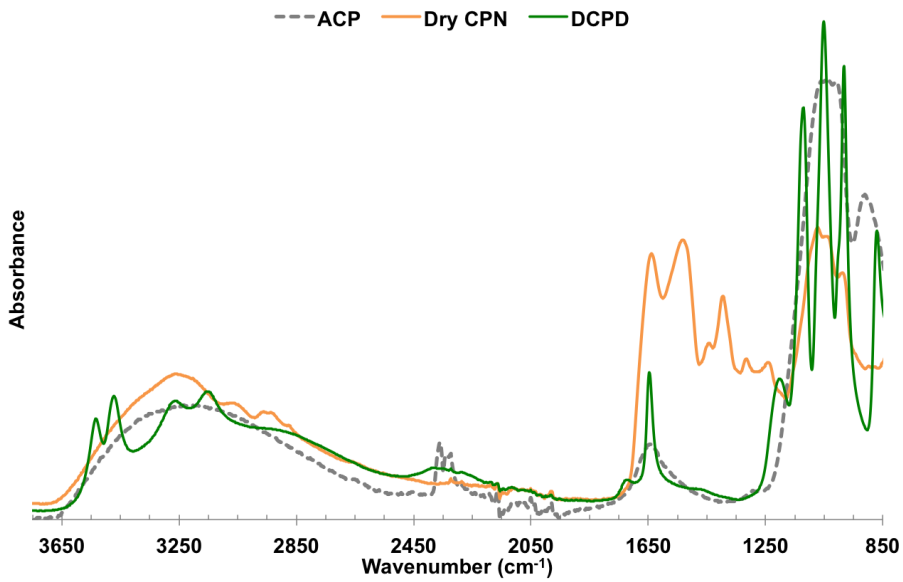

Figure 4.5: FT-IR spectra recorded for the synthesised ACP, DCPD and the dried CPN samples. The dried CPN samples show a similar absorbance as ACP, lacking any of the characteristic crystalline peaks as observed for DCPD between 850 and 1250 cm^{-1}

tative potential structures that crystallise in small unit cells. If pure HA crystallites were present (Figure 4.7b), and given that the experiments carried out allowed for data to be collected up to a 2θ of 115°, the simulations suggest that crystallites of 18.9 Å- large enough to include two or more unit cells stacked in the 'b' or 'c' unit cell edge directions - would cause a broad peak centred at ~34°. A second broad peak would be outside the experimental range but a tail would cause intensities to increase significantly with the scattering angle from a 2θ of 60°. A similar profile would already be expected for a 13.8 Å crystallite, which can contain two unit cells stacked along the 'c' direction and would fill 2% of the casein CPN core, or only 3% of the OPN core. For DCPD, the simulations suggest that experimental resolution would not allow resolution of the presence of crystallites formed by up to 3 unit cells stacked in the 'a' direction, nor any made of a single monoclinic unit cell (as shown by the broad intensity profile for 5.8 Å or 15.2 Å). Crystallites of DCPD larger than 15.2 Å would however be resolved through the presence of at least two distinct peaks within the experimental range observed.

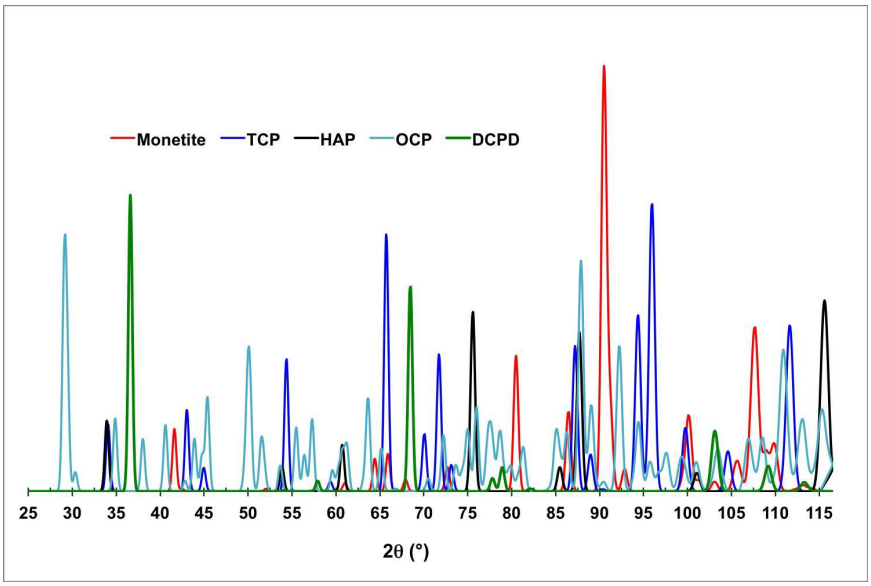

Figure 4.6: Neutron diffraction patterns for different calcium phosphate phases calculated under conditions measured on D16 at the Institut Laue-Langevin. The diffraction patterns are based on the crystal structures of anhydrous calcium phosphate, tri-calcium phosphate, hexagonal HAP, DCPD and OCP. All calcium phosphate phases are expected to diffract within the range measured on D16

4.3.5 Neutron diffraction control measurements

Known phases of calcium phosphates were measured on the D16 diffractometer as control measurements. Figure 4.8 A, shows the diffraction pattern recorded on a control sample of DCPD. The measured diffraction pattern is noisy, however it is possible to discern the two major diffraction peaks caused by DCPD crystalline material. Only peaks 15 times above the standard deviation were considered significant. The measured peaks agree well with the position from the simulated data, and indicate that it is possible to detect crystalline phases of calcium phosphate on the D16 diffractometer.

Figure 4.8 B shows measurements completed on the synthesised amorphous calcium phosphate, previously characterised by FT-IR. The synthesised amorphous phase stabilised by magnesium was also measured and showed no clear Bragg diffraction, as would be expected for a sample lacking in crystalline order. Upon hydration of

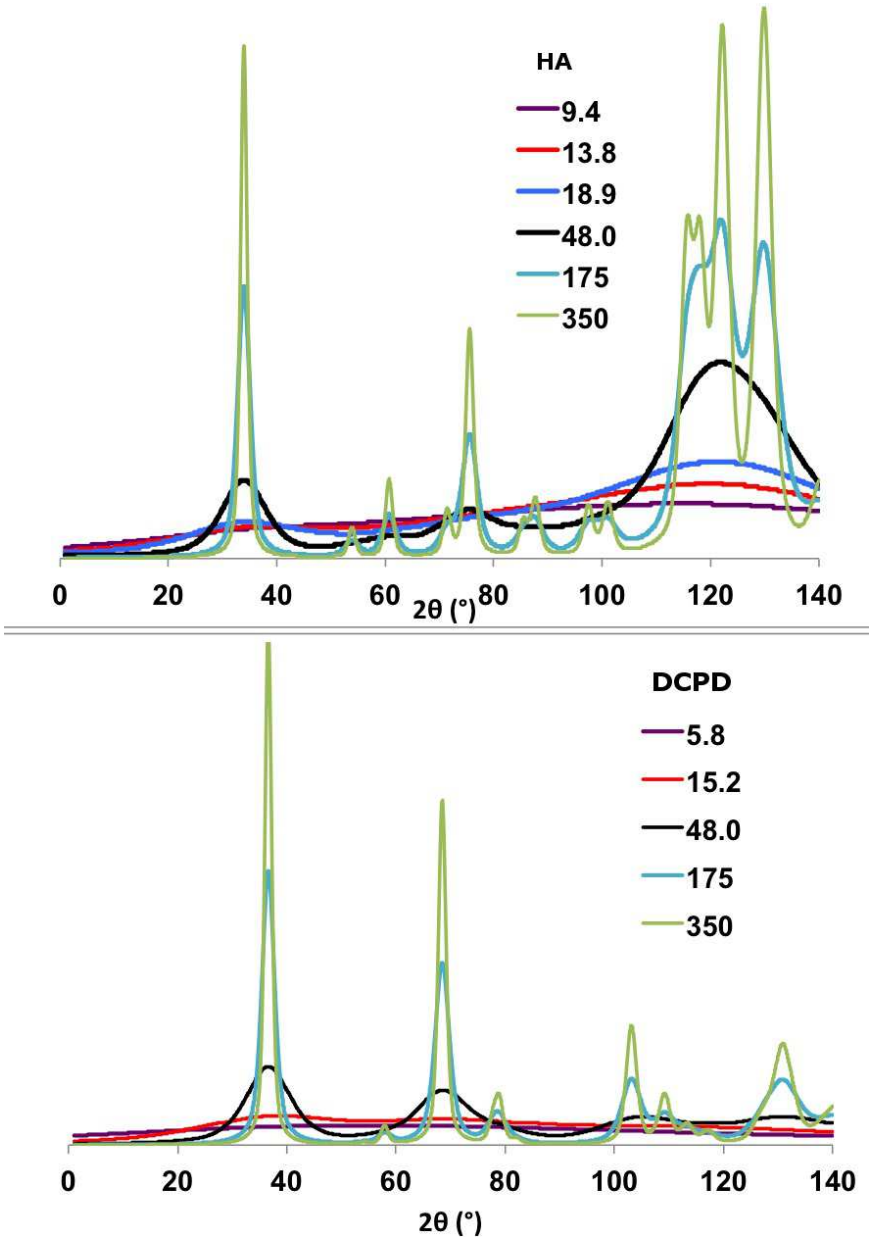

Figure 4.7: Simulations of neutron diffraction patterns expected at D16 using the sample measurement conditions. Patterns for Hydroxyapatite (top) assumed a pure hexagonal crystal with the unit cell: a = b = 9.432, c = 6.881 (Å). DCPD was also assumed to take the form of a pure crystalite with the unit cell: a = 5.812, b = 15.18, c = 6.239 (Å). CPN stabilised by OPN have the largest core diameter of 350Å whereas CPN stabilised by β casein have a core of 48 Å, corresponding to the predicted sizes indicated.

the amorphous phase crystallinity is introduced to the system, which has a diffraction pattern similar to that observed in DCPD, indicating a transformation from an amorphous to crystalline phase Figure 4.8 B.

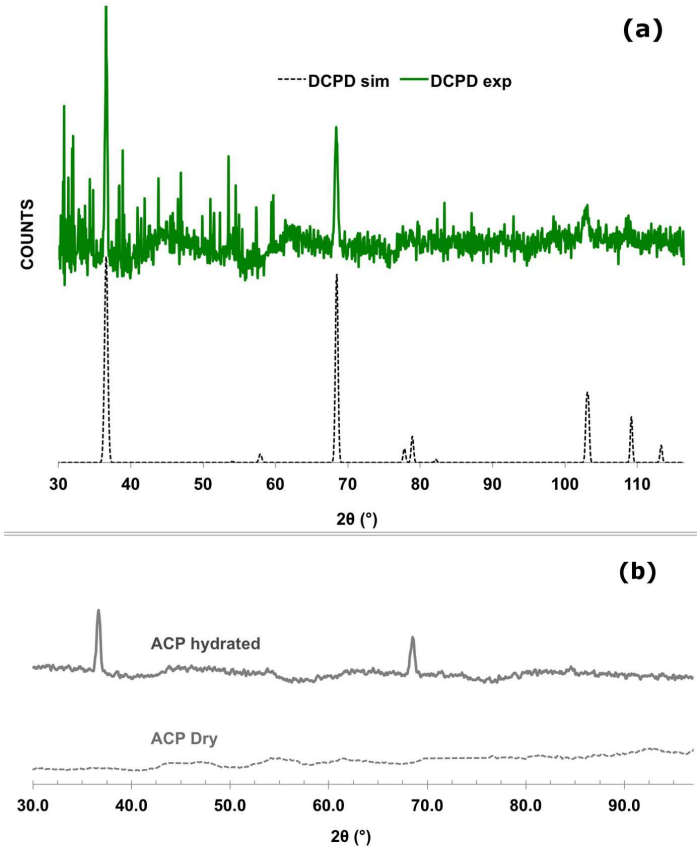

Figure 4.8: (a) Simulated and experimental neutron diffraction pattern of DCPD. Experimental data was collected on D16 at the ILL. Sharp Bragg peaks are observed as expected at $2\theta = 36.3°$ and $2\theta = 67.8°$, closely matching the simulated diffraction pattern. (b) Maturation of ACP as observed by neutron diffraction. The dry sample lacks any sharp Bragg peaks however after being left in a hydrated, or a atmosphere sharp peaks corresponding to DCPD appear as the ACP matures. Evidence that hydrated ACP does indeed eventually form DCPD, or a more crystalline phase.

4.3.6 Neutron and X-ray scattering indicate a lack of crystallinity within the OPN-CPN

Neutron diffraction measurements were carried out on several OPN-CPN samples, including a dry sample, a matched out sample, and a sample in 100 % D_2O(Figure 4.9A). Initially it was thought that the peptide may form fibres or other ordered structures which would interfere with the measured diffraction pattern. In 41% D_2O these contributions are hidden because this is the match-point of protein. However none of the measured samples showed any Bragg diffraction by neutron scattering.

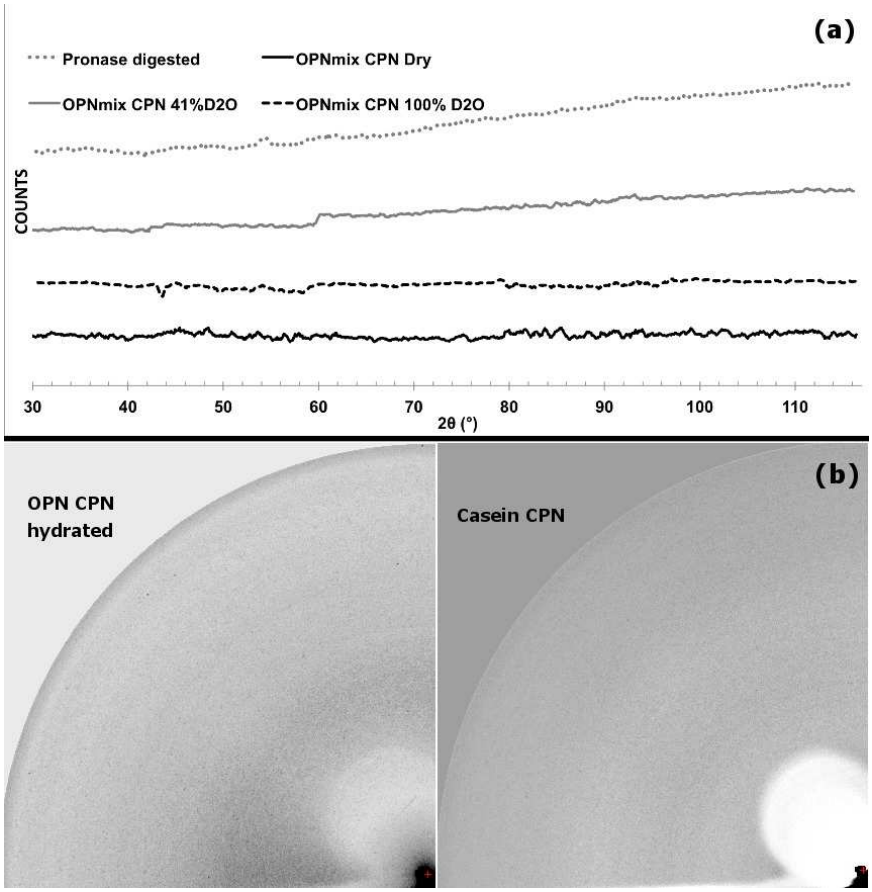

Figure 4.9: (A) Neutron diffraction measurements collected on D16 for different samples containing calcium phosphate nanoclusters. The observed patterns are similar to those collected for dry ACP, and no unambiguously diffraction rings can be detected. **(B)** X-ray diffraction patterns for the same samples as shown in (A), covering a lower 2θ range. At this range no Bragg peaks are observed for either sample.

The same samples were also measured by X-ray scattering in order to detect any Bragg diffraction at lower values of 2θ Figure 4.9 B. Once again the diffraction pattern is featureless and corresponds well with the neutron diffraction data and in agreement with the data obtained by FT-IR.

4.4 Discussion

4.4.1 OPN forms core shell CPN

Previously the formation of CPN after sequestration of calcium phosphate by OPN1-149 was observed by SAXS, Holt *et al.* [45] modelled a core-shell structure to the collected SAXS data. Using small-angle neutron scattering with contrast variation it is possible to hide contributions of the shell from the scattering, which is not possible by SAXS. Here we report SANS data with contributions from the shell hidden using contrast variation. Our results yield a similar size of the CPN stabilised by OPN1-149 23.4 nm, compared to the result of Holt *et al.* of 21.2 nm [45]. Further more we also obtain information about the size of the core and whole CPN (Table 4.3). For OPN1-149 CPN we obtain a shell thickness of 2 nm which compares well with the model obtained by Holt *et al.* which yields a shell thickness of 1.6 nm [67]. The shell thickness of OPN1-149 CPN is 2 nm, slightly smaller compared to the shell thickness of OPNmix-CPN of 2.5 nm. This is probably due to presence of un-cleaved OPN fragments in the OPNmix [213]. Previously it was shown that CPN stabilised by β-casein have a radius of 4.04 nm [67]. This radius is much smaller compared to our results. Using SANS it was determined that the core of a β-casein-CPN has a radius of 2.3 nm, yielding a shell thickness of 1.7 nm. The shell thickness of OPN1-149 CPN is only slightly larger at 2 nm. If we consider the length of the amino acid sequence β-casein consists of only 42 amino acids compared to the 149 amino acid residues of OPN1-149. The similarity of the shell thickness can be understood once it is considered that OPN1-149 contains 3 phosphate centres through which it can interact with the core, compared to only one phosphate centre present on β-casein [67, 213]. The core of β-casein-CPN is considerably smaller then the core of OPN1-149-CPN, 2.305 nm and 21.4 nm respectively. That the core radius is significantly

larger for OPN CPN can be explained by the apparent increased sequestration ability of this peptide. Previously it was determined that the core of OPN1-149 is more basic compared to that found in β-casein-CPN [67]. Holt *et al.* proposed that the increased size of the core of OPN1-149-CPN is probably due to the decreased surface area of OPN1-149 phosphate centres compared to those found in β-casein. This would allow more phosphate centres to interact and hence stabilise a larger area of calcium phosphate in the core. The decreased surface area per phosphate centre is dependent upon the properties of the phosphate centre in the protein. Therefore the power of sequestration depends upon several different factors not least the extent of phosphorylation in the phosphate centre but also the flanking regions alongside each phosphate centre. This greater ability to sequester calcium phosphate can be explained by the higher negative charge amino acid residues present in osteopontin compared to casein. Even though the phosphate centres are roughly similar the peptide as a whole is more electronegative, favouring the interaction with calcium phosphate. The higher net negative charge increases the affinity of the PC and aids in the act of sequestration.

4.4.2 The core of CPN

The calcium phosphate core of β-casein-CPN has been studied using a range of techniques. X-ray absorption experiments yielded X-ray radial distribution functions similar to those found in biological and synthetic amorphous calcium phosphates [214, 215, 216]. Other biophysical methods have also been used to determine any order on several length-scales in calcium phosphate associated with the casein micelle. High resolution electron diffraction did not observe any periodic lattice spacing ruling out the presence of medium and long range order (15 Å plus) [217]. The presence of short and medium range order in the inorganic part of the casein micelle has been studied by extended X-ray fine absorption fine structure (EXAFS). EXAFS absorp-

tion spectra of the β-casein-CPN sample and a control sample of amorphous calcium phosphate were modelled and ACP provided a good match for the spectra obtained for the β-casein-CPN. The results further rule out the effects of medium range order, concluding that β-casein-CPN has only short-range order in the crystalline phase [69]. Other structures for the inorganic CPN calcium phosphate have been proposed including a structure consisting of 9 unit cells of DCPD in the form of a tetrahydra [218]. This proposition is not based on any experimental evidence, furthermore only monoclinic of DCPD has been observed, which has to lower symmetry to form a tetrahydra structure [219]. Other workers proposed that the micellar calcium phosphate contains order over the medium range scale in the form of single $Ca_9(PO_4)_6$ units with a 0.95 nm diameter [220]. Order over the medium range length scale has been ruled out by several extensive EXAFS studies [221, 69].

Both the X-ray and neutron scattering data obtained for OPN1-149-CPN presented here did not show any diffraction, indicating a lack of structural order in the sample. Using match-out neutron diffraction contributions from the protein shell were hidden. No diffraction patterns were observed. According to the neutron diffraction simulations of different crystalline size it should be possible to observe crystalline material equivalent to the core size of OPN-CPN. We can conclude that the core has an amorphous structure over the long and medium range. With no well defined crystal packing.

4.4.3 Implications of an amorphous phase

Amorphous phases are inherently metastable at best, formation of a crystalline phase is thermodynamically desirable due to the negative energy of formation of the crystalline lattice. The supersaturation level of a fluid can be calculated with respect to different crystalline phases. Holt *et al.* have completed this calculation on several biological fluids with respect to different calcium phosphate phases shown in

(Figure 4.10) [10].

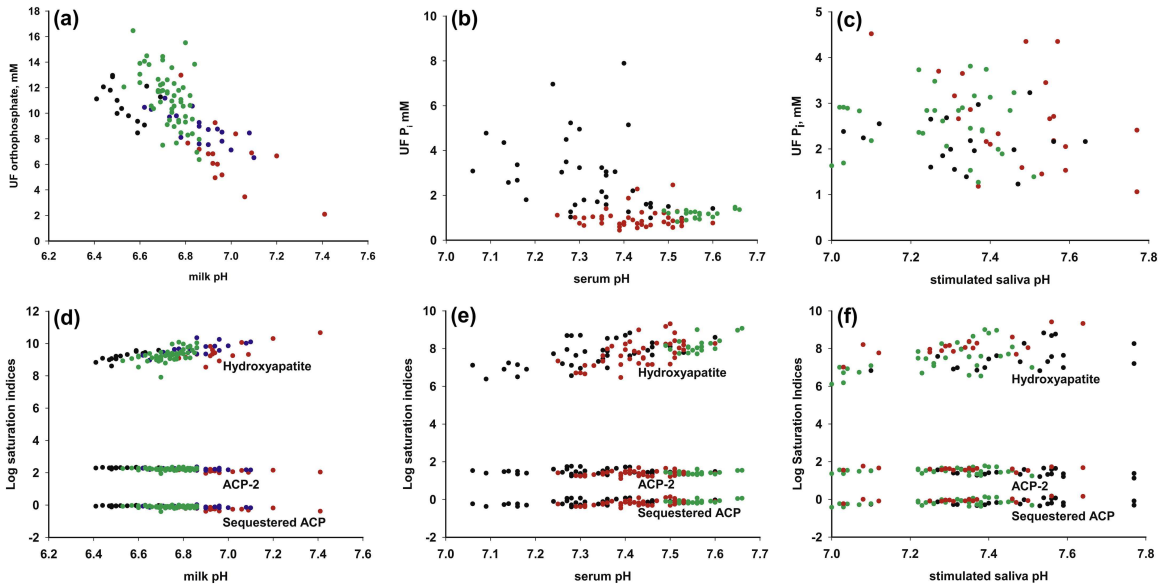

Figure 4.10: Compositional variation and Log(saturation indices) in ultrafiltrates of bovine milk, human blood and stimulated saliva taken from Holt *et al.* [10]

.

The calculation is based on concentrations of calcium and phosphate taken from real biological fluids. The calculated saturated is shown assuming different structures of calcium phosphate in the biofluid [10]. The calculation shows that milk, blood and saliva are all supersaturated with respect to the thermodynamically stable hydroxyapatite. However if the structure of calcium phosphate is assumed to be amorphous the saturation index lies below 0 for all three biofluid [10]. That the core of OPN1-149-CPN is amorphous agrees well with the thermodynamic description of nanocluster formation [72]. Resulting in the stabilisation of biofluids supersaturated with respect to hydroxyapatite by sequestration of amorphous calcium phosphate within CPN. In vertebrates the mineralisation process proceeds via the deposition of mineral within an organic matrix [222]. Often the organic matrix consists of protein such as collagen. Although the exact process is not well understood several theories are under development [222, 223]. Most of these theories share in common that they require calcium phosphate precursor is an amorphous phase so that the properly ordered hard mineral

can grow [222, 224]. This is because the maturation of calcium phosphate into it's lower energy crystalline phases is an irreversible process. An amorphous phase therefore allows for more control over crystal growth and the resulting properties of the mineralised tissue [223]. The amorphous nature of calcium phosphate within OPN-CPN is indicative of the possibility that OPN-CPN are of general importance, not only as an inhibitor of mineralisation in biological fluid but also that they can act as a stable reservoir of amorphous calcium phosphate, allowing mineral phases to grow at the expensive of the calcium phosphate present within the core [10]. This corresponds with several independent observations of OPN peptides. Including the up-regulation of protein expression upon an increase in calcium concentration in different biological fluids, up-regulation upon trauma to bone and other mineralised tissues as well as the observed concentration of osteopontin on the mineral front of bone and teeth [225, 226, 227]. The amorphous phase precipitates first during biomineralisation and acts as the precursor for consequent mineral formation. OPN-CPN and similar CPN found *in vivo* may therefore not only stabilise biofluids supersaturated with respect to HA but also provide bio-available calcium phosphate in the form of ACP. This behaviour has been observed *in vitro* where an OPN-CPN containing solution was shown to remineralise an artificial collagen fibril, forming ordered HA from the amorphous calcium phosphate present in the CPN [61]. Calcium phosphate nanoclusters stabilised by β-casein have also been shown to be able to remineralise tooth enamel when included within a chewing gum [228]. Several potential applications for such a technology have been imagined including the creation of mimetic biofluids which more accurately represent the ion concentrations found *in vivo*. Remineralisation of collagen and enamel scaffolds [61, 228] as well as increasing bio-availability of low solubility compounds [229]. Therefore the observations that different peptides can form different size CPN whilst maintaining an amorphous core is vital for the development of artificial peptides. The characterisation presented here is a first step in creating

artificial phosphate centre containing peptides. By engineering the peptide and more specifically the amino acids present in the phosphate centre different peptides with different size CPN can be formed, with potential applications in drug delivery and remineralisation of protein scaffolds *in vivo*.

4.5 Conclusion

It has been shown that nanoclusters stabilised by OPN1-149 and OPNmix have a core-shell structure, the formed core is larger then the core of CPN stabilised by β-casein, with a similar shell thickness. The difference in size is due to the increased sequestration ability of OPN1-149, caused by the PC and flanking amino acids. The reported diffraction studies observed no crystalline material in the core of OPN-CPN despite the larger radius. That the core is amorphous is in agreement with the theory of nanocluster formation and stabilisation of biofluids. Furthermore, amorphous calcium phosphate is biologically available and given the abundance of OPN in the extracellular matrix CPN formed by OPN peptides may play an important role in maintaining and growing mineralised tissue. There are several possible applications of the science described in this chapter, and an understanding of the size and structure of CPN is a requirement for future biotechnological developments. Possible applications include, bio-mimetic materials for the repair of damaged bone and re-mineralisation of teeth, artificial biofluids that remain stable over long periods of time and for the development of growth medium in tissue engineering applications that can support the growth of mineralised tissues.

Chapter 5

Effects of phosphorylation on hOPN1-157 structure and dynamics

A manuscript containing the experimental findings presented in this chapter is under preparation:

- **Lenton, S.,** Nylander, T., Holt, C., Haertlein, M., Zaccai, G. and Teixeira, S. Structure and dynamics of the human Osteopontin protein and the effects of phosphorylation.

Abstract

In humans osteopontin is known to play important roles in mineralisation, cellular signalling and inflammation. The function of osteopontin in many of these processes is highly dependent upon the location and amount of post-translational modifications of osteopontin. In the control of mineralisation phosphorylation is of particular importance for determining osteopontin structure. Phosphorylated amino acid residues alter the way in which the protein interacts with mineral interfaces and facilitates the sequestration of amorphous calcium phosphate in solutions supersaturated with respect to hydroxypapatite. In this chapter the structure, and effects of phosphorylation thereon, of human osteopontin is characterised using small-angle X-ray scattering. Information about the effects of phosphorylation on protein dynamics is obtained by elastic incoherent neutron scattering. The results demonstrate that human osteopontin is a disordered protein that does not gain structural features upon phosphorylation. Elastic incoherent neutron scattering indicates that the resilience of the protein on the nano to pico second timescale is independent of phosphorylation. Finally the importance of phosphorylation with respect to the role of osteopontin in the control of biomineralisation is discussed in terms of the results obtained.

5.1 Introduction

Several proteins that are involved in bone growth and mineral regulation undergo significant amounts of phosphorylation including osteopontin (OPN), bone sialoprotein and Fetuin A [66, 230, 59]. OPN is a highly conserved multifunctional protein expressed in many mineralised tissues, soft tissues and biological fluids [89, 161]. OPN is a member of the SIBLING (small integrin-binding ligand, N-linked glycoprotein) family of proteins, it binds integrin through a short Arg-Gly-Asp sequence [84]. In mammals OPN is a versatile protein that is highly modified by post translational modifications. Post-translational modification (PTM) of proteins is an important, often reversible, biochemical event involved in the regulation of various cellular functions [231, 124]. There exists over 300 different types of modifications regulated by PTM enzymes, from the introduction of phosphate groups to the cleavage of the amino acid chain, for a recent review see Witze *et al.* [125]. PTMs significantly increase the range of chemically different amino acids from 20 to 140 and thereby significantly diversify both the structure and function of proteins after translation [124, 35]. OPN is found in high levels in milk of different species, typically in a highly phosphorylated state [99, 86]. For example human breast milk contains around 100 - 300 mg/mL OPN corresponding to 1-2 % wt/wt of the total protein concentration [232]. Other human biofluids including urine and blood also contain significant concentrations of OPN [81]. High levels of OPN are also found on the matrix of mineralised tissue including bone and teeth [84]. OPN plays an extensive role in the formation and regulation of mineralisation, and is also associated with several different pathologies including inflammation and cancer [233, 234, 90, 91]. The Kinase responsible for phosphorylating several residues of OOPN in humans has recently been identified as the FAM20C Golgi Kinase [97, 96]. In globular proteins phosphorylation can alter the conformational space occupied by the amino acid residues, causing for example the opening of a binding

pocket allowing for substrate access [235]. Other post-translational modifications are not reflected in the conformation of the protein but instead alter functional changes in protein-protein or protein-substrate interactions [126]. Due to the lack of secondary and tertiary structure in intrinsically disordered protein the latter type of regulation by post-translational modification is considered more important. The function of IDPs such as osteopontin are widely regulated by many types of post-translational modifications, the reversible nature of many post-translational modifications enable IDP to adopt both different functional and conformational states in a rapid manner [236, 237]. A sub-class of intrinsically disordered proteins, known as hub proteins, convey many roles *in vivo*. Hub proteins require a large amount of flexibility at either a local or global level and therefore they have a propensity to contain intrinsically disordered regions or are fully disordered [131]. On average mammalian proteins undergo at least three post-translational modifications [238]. Proteins that undergo a large amount of post-translational modifications have a tendency to be intrinsically disordered [120, 122, 121]. This is due to the increased accessibility of disordered regions to enzymes that undertake post-translational modifications [121, 239, 240, 241]. The reversible nature of many post-translational modifications and the wide prevalence in intrinsically disordered proteins PTM are thought to allow IDPs to undertake their often multifunctional role *in vivo*. [124, 238]. Phosphorylation is one of the most common post-translational modifications found in intrinsically disordered proteins, in eukaryotes at least a third of intrinsically disordered proteins are post-translationally phosphorylated [242, 238, 243]. Due to the inherent quantity and amount of phosphorylation found *in vivo* the study of the effects of phosphorylation on IDPs such as OPN1-157 is vital to fully understand the regulation and effects of PTMs on the protein structure and function.

As previously mentioned one of the most important aspects of PTMs is their ability to modulate protein structure, as is often observed in globular proteins [236]. In

intrinsically disordered proteins and intrinsically disordered protein regions several examples exist where phosphorylation of amino acid residues induces folding in, causing so called disorder to order transitions [244, 245]. For example the intrinsically disordered tau protein is largely processed by kinase enzymes. Upon phosphorylation of specific threonine and serine residue the tau protein remained globally intrinsically disordered but some local structure becomes ordered, including the stabilisation of an α helix [246, 247, 248]. Induction of secondary and tertiary structure elements in IDPs by PTMs is, however, uncommon. Instead, modifications of IDPs by PTMs often modify protein-protein and protein-substrate interactions instead.

Experimental evidence of the effects of phosphorylation on protein-substrate interaction have been observed for the casein class of proteins. Whilst phosphorylated caseins can sequester amorphous calcium phosphate, unphosphorylated caseins can not [79]. This is due to the charge-charge interaction that stabilises the amorphous phase [79]. Similarly it is also known that phosphorylation is required for OPN to sequester amorphous calcium phosphate thereby forming calcium phosphate nanoclusters [45]. Certain fragments of OPN can inhibit the growth of calcium phosphate mineral phases by binding to the surface, the binding and hence inhibition of mineral growth is also phosphorylation dependent [59, 249, 94, 250].

So far no studies on the effects of phosphorylation on the dynamics of intrinsically disordered proteins have been completed. Even if the average conformational space occupied by a protein is not effected the flexibility of a protein may be. This is especially the case for intrinsically disordered protein where the protein exists as an ensemble of structures. Therefore the characterisation of both structure and dynamics is important for a full understanding of the effects of protein phosphorylation. IDPs are often involved in the development of different diseases [251]. Aggregates of Tau protein found on the brains of alzheimer patients are found in a hyperphosphorylated state although the relationship between protein phosphorylation and plaque formation

is not well understood [246, 252].

In bovine milk the OPN fragment OPN1-149 is easily isolated, and is the result of cleavage by the plasmin protein. The analogous peptide in humans is hOPN1-157. The sequences similarity between human and bovine OPN is high for the cleavage fragments (Figure 5.1). In this chapter the effects of phosphorylation on this novel fragment of human osteopontin, believed to play an important role in biomineralisation, is studied. The hOPN1-157 protein is produced using recombinant methods and produced in phosphorylated form by co-expression with casein kinase. The structure of the protein is characterised by a range of biophysical techniques. Information about the conformational ensemble present in solution is obtained by small-angle X-ray scattering using the ensemble optimisation method. The dynamics of the protein in both phosphorylated and unphosphorylated states are examined in the hydrated powder state by elastic incoherent neutron scattering. The results show that hOPN1-157 is a disordered protein independent of phosphorylation state which samples a large area of conformational space. On the pico-second timescale no effects of phosphorylation are observed, however it is shown that formation of CPN is phosphorylation dependent.

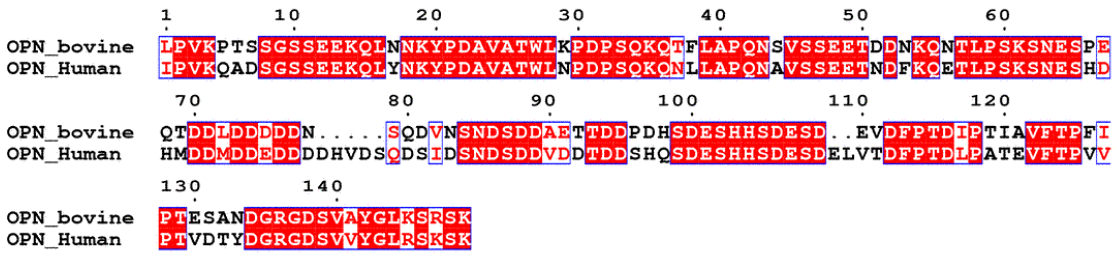

Figure 5.1: Amino acid sequence alignment of human OPN1-157 (Uniprot: P10451) and bovine OPN1-149 (Uniprot: P31096) osteopontin fragments. The alignment was performed using clustalW [253], residues highlighted in red are conserved amino acids.

5.2 Materials and methods

5.2.1 Bioinformatics

PONDR predictions where made using the VL-XT and VSL2 algorithms on the hOPN sequence. Detailed descriptions of the differences between these algorithms can be found elsewhere [175, 254, 255, 256]. Charge-Hydropathy analysis was completed using the PONDR server (www.pondr.com), which included charge hydropathy calculation for known folded and unfolded proteins [175].

5.2.2 Over-expression and purification of human osteoptontin

Recombinant over-expression of phosphorylated hOPN1-157 in *E.coli* was based on the co-expression system developed by Clegg and Holt [79]. The method relies on the simultaneous co-expression of two different vectors with a low and high copy number, expressing a protein kinase and substrate protein respectively. The hOPN1-157 gene containing a 6 histidine tag on the N-terminus (527 base pairs) was inserted into the pET 28a(+) vector. The chosen expression vector carries genes which encode for ampicillin resistance and a T7 promoter. Plasmid containing the gene encoding for the casein kinase (CK2α) was kindly provided by Dr R. Clegg (Glasgow University) on a pACYC-Duet vector (Novagen). In the pACYC-Duet 1 vector protein expression is also controlled by a T7 promoter, the vector also contains a gene encoding for chloramphenicol resistance. This plasmid maintains a lower copy number in *E.coli* compared to the pET 28(+) plasmid. By combining two different antibiotic resistance genes the suppression of non transformed bacterial cells is maintained by the addition of both chloramphenicol and ampicillin, allowing for tighter control of the expression system.

The host organism used for recombinant protein over-expression was BL21-DE3. This

strain of *E.coli* incorporates elements from the T7 bacteriophage, and includes the DE3 lysogen which carries the gene encoding for T7 RNA polymerase. The expression of DE3 lysogen is under the control of a lacUV promoter, thereby allowing for induction of the T7 polymerase and consequent protein over-expression on both plasmids via the addition of isopropyl β-D-1-thiogalactopryanoside (IPTG).

The pET28 plasmid containing the hOPN gene was introduced into chemically competent BL21-DE3 *E.coli* cells by heat shock. The transformed bacteria were propagated on plates in the presence of ampicillin to select for pET28 induced ampicillin resistance and and were incubated overnight at 37°C. Colonies from the overnight plates were used to inoculate 10 ml Luria-Bertani (LB) medium containing ampcicillin, which was left to grow overnight at 37°C [257]. 500 ml of LB containing ampicillin was inoculated with 5 % V/V of the overnight cultures. The cells were grown at 37°C with vigorous stirring until the culture had reached mid-exponential phase at an optical density (OD) at $\lambda = 600$ nm of 0.7. Transcription of the OPN1-157 gene from the plasmid encoded T7 promoter in pET28 was then induced by addition of IPTG to a final concentration of 1 mM. Cells where grown for a further 3 hours at 37°C to allow OPN1-157 production, after which the bacteria were harvested by centrifugation (10.000 g, 20 minutes) the resulting cell pellet was stored at -22°C until purification. In order to achieve *in vivo* phosphorylation the pET28 plasmid containing the hOPN1-157 gene and a PaCYC-Duet vector containing the casein kinase were simultaneously introduced into chemically competent BL21-DE3. The pACYC duet contains a gene expressing for chloramphenicol resistance. The transformed bacteria were propagated on LB containing plates in the presence of both ampicillin and chloramphenicol to select for bacteria transformed with both vectors. The rest of the preparation was identical to the unphosphorylated preparation, with the inclusion of chloramphenicol at every step of bacterial growth. Aliquots taken before induction and after induction were stained with coomasie blue and resolved on

a 12% SDS-PAGE gel. A western blot was performed by transferring aliquots onto a polyvinylidene fluoride (PVDF) membrane. The membrane was blocked with 5% (w/v) bovine serum albumin in tris buffered saline (50 mM Tris-Cl, pH 7.5, 150 mM NaCl) and probed with a Monoclonal anti-polyHistidine antibody produced in mouse (Sigma) for 90 minutes at room temperature. Blots were developed by addition of goat anti-mouse alkaline phosphatase antibody (Sigma), and visualised by addition of an alkaline phosphatase substrate. The frozen cell pellet was re-suspended in ice cold lysis buffer (20 mM tris, 0.5 M NaCl, pH 8) (10 ml per gram of cell paste) containing EDTA free protease inhibitors (Roche), cells were lysed by sonication on ice. Insoluble material was released from the lysate by centrifuging at 14,000 g for 30 minutes. The cleared protein extract was then loaded onto a HisTrap 5 ml nickle affinity column (G.E. Healthcare) using an Akta Prime system. The column was washed with lysis buffer containing 25 mM imidazole to release weakly bound proteins (1 ml/minute, 30 minutes). The His-tagged protein was eluted by increasing imidazole concentration in the buffer from 50 mM to 250 mM, whilst constantly collecting fractions of 1 ml. The fractions were analysed by SDS-PAGE and consequently transferred onto a polyvinylidene fluoride (PVDF) membrane which was developed as before. Samples containing the expressed fraction were further purified by analytical size exclusion chromatography. 250 μg of each sample in buffer containing 50 mM phosphate, 300 mM NaCl, 0.02% NaN_3 at pH 7), were loaded onto a Superdex 75 pg 10/300 GL column (GE Healthcare) connected to a Bio-rad system (Bio-rad). The column was calibrated using molecular weight standards from Bio-rad (Bio-rad) dissolved in the sample buffer. The elution from the column was monitored by U.V. absorption at 280 nm, and fractions of 0.5 ml collected. The fractions were analysed by SDS-PAGE for purity, samples containing the expressed protein were pooled, and extensively dialyzed against H_2O and consequently freeze dried and stored in a desiccator containing phosphopentoxide. Successful expression of the purified protein was further confirmed

by N-terminal sequencing of the first 5 amino acids.

5.2.3 Mass spectrometry analysis

Samples were dialysed against a buffer containing 20 mM tris pH 8, 250 mM NaCl and concentrated to 0.8 mg/ml (50 nMol). 4 μL of the sample diluted to 5μM in 0.03 % TFA were injected for analysis. Loaded samples were desalted using a Zorbax 300sb-C8 (Agilent technologies) on-line protein trap for 3 minutes with a flow rate of 100 μ l/min of 0.03% TFA. The sample was then eluted at 50 μL/min with 5% acetonitrile. The desalted samples were analysed on a 6210 LC-TOF mass spectrometer interfaced with an LC pump system (Agilent Technologies).

5.2.4 Proteinase K sensitivity assay

Proteolytic sensitivity analysis of hOPN1-157 in the phosphorlyated and unphosphorylated state and a control of myoglobin were completed by the method of Denning *et al*. [258]. Briefly, 5 μg of each protein was incubated with 100 ng of proteinase K (Sigma) at 37 °C controlled by a water bath. Aliquots of each sample were taken at the indicated time points, the reaction was stopped by boiling the aliquots in SDS loading buffer. The samples where then loaded onto a 12 % SDS-PAGE gel run at 150 V for 50 minutes, with molecular Pagerule plus prestained protein ladder (Fermentas, USA).

5.2.5 1D HET-SOFAST NMR

NMR spectra were measured at the NMR facility at the Institut de Biologie Structurale (Grenoble, France). All NMR experiments were performed at a protein concentration of 150 μM in a buffer of 50 mM phosphate, 200 mM NaCl, at pH 7. NMR experiments were performed at room temperature on a Varian 600 MHz spectrometer (Agilent technologies, USA), as described by Schanda *et al.* [259].

5.2.6 FPLC-SAXS Experiment

Measurements were performed using an on-line high pressure liquid chromatography (FPLC) system (Viscotek GPCmax, Malvern intstruments) directly attached to the sample capillary on the BM29 Bio-SAXS beamline at the European Synchrotron Radiation Facility (Grenoble, France). A 2.4 ml Superdex 200 Increase 5/150 GL column (G.E. Healthcare) was attached; prior to each run the column was equilibrated with 2 column volumes of the corresponding buffer. Sample volumes of 100 μl (5 mg/ml) were directly loaded onto the column using the FPLC system, at a flow rate of 0.1 ml per minute at room temperature. Elution from the column was measured by U.V. absorption at 280 nm^{-1} . X-ray scattering data was continuously collected on the column elution at a wavelength of 0.99 Å using a sample-to-detector distance of 2.81 m corresponding to a measured q range of 0.08 - 4 nm^{-1}. Small angle X-ray scattering data was collected at the rate of 1 frame every 2 seconds leading to the collection of roughly 1000 frames for the whole run. Primary data reduction leading to radially integrated, calibrated and normalised one dimensional scattering profiles was completed by the EDNA software pipeline [260]. Frames were compared manually and investigated for radiation damage, non damaged frames were averaged, and buffer frames were averaged for background subtraction. Calibration of the superdex 200 (GE Healthcare) column was completed using the same method with molecular weight standards from Bio-Rad.

5.2.7 FPLC-SAXS data analysis

Model independent

The experimental radius of gyration, R_G, and the forward scattering intensity, I_0, were calculated from data at low q values in the range of $qR_G < 1.3$, using the Guinier approximation: $lnI(q) = Ln(I_{(0)}) - R_G^2 q^2/3$. The pair distribution function p(r)

in real space was calculated by indirect Fourier transform using GNOM from the ATSAS software package [147]. Analysis of the p(r) function allowed the estimation of the maximum inter-particle distance (D_{max}) as well as R_G and I_0. The Kratky plot (q^2I against q) provides further model independent information about the scattering particle. Globular particles typically show a well defined maximum when plotted in this way whereas for a disordered protein the Kratky plot should plateau at high values of q.

Ensemble Optimisation Method (EOM)

Flexible proteins are best described by an ensemble of structures rather than one well defined structure as is the case for most globular proteins. The ensemble optimisation method (EOM) software was used which takes into account protein flexibility by allowing the co-existence of several different conformational isomers in the scattering solution. Using the EOM software a pool of 10,000 random models based on the amino acid sequence of the hOPN1-157 protein were generated. A genetic algorithm was then employed in order to narrow the pool down to 50 models that best fitted the scattering curve which were then selected for further analysis. The R_G and D_{max} of these 50 models were plotted as a histogram and 5 models were selected to represent the ensemble.

5.2.8 Elastic incoherent neutron scattering experiments

Neutron scattering experiments on phosphorylated and unphosphorylated hOPN D_2O hydrated powder sample were carried out on the IRIS spectrometer at the ISIS spallation neutron source, Chilton, UK. By exploiting the Pyrolitic Graphite PG(002) analyser configuration and the Beryllium filter, the covered momentum transfer range was between 0.3 and 1.8 [1] with an energy resolution of 17 μeV (corresponding to a 50-ps time scale). Each sample contained 100 mg of protein powder with 0.44 g D_2O/g protein held in a flat aluminium cell with internal spacing of 0.3 mm. The sample was

placed at an angle of 135° with respect to the incident beam. A elastic temperature scan was performed by continuously counting the scattered neutrons whilst increasing the temperature at a rate of 1 K/min from 10 K to 310 K. The MSD were obtained from the analysis of the integrated intensity over a frequency window corresponding to instrumental resolution around the elastic peak.

5.2.9 Dynamic Light Scattering experiments

Samples (5 mg/ml) were prepared in phosphate buffer as per the method of Holt. Samples were measured on a Malvern Zetasizer instrument with an incident wavelength of 633 nm. At least 10 consecutive measurements of the correlation time of the scattered light intensity were performed and averaged for each sample. The data were fitted according to the algorithm present in the Malvern analysis software. Briefly the correlation function ($G(t)$) was fitted to the equation

$$G(t) = B + A2^{-2q^2}Dt \tag{5.1}$$

Where B is the baseline measurement, A is the amplitude, q is the scattering vector, t is the delay time and D is the diffusion coefficient. The hydrodynamic radius of the particle (R_h) is inversely proportional to the diffusion coefficient D and the solvent viscosity η as per the Stokes-Einstein equation

$$R_H = \frac{k_B T}{6\pi\eta D} \tag{5.2}$$

Such that the radius of hydration ($\mathrm{R_H}$) can be determined experimentally.

5.3 Results

5.3.1 Amino acid composition analysis of hOPN1-157

OPN belongs to the Small integrin binding-ligand n-linked glycoproteins (SIBLING) family of proteins, known to have degrees of structural disorder [84]. The hOPN1-157 fragment studied here is significantly shorter and lacks several of the binding and interaction domains present in full length OPN. This raises the question as to whether the resulting fragment retains intrinsically disordered features or obtains secondary structural elements. Amino acid analysis of the hOPN1-157 protein was completed using several different computational techniques, that focus on predicting intrinsic disorder. The Uversky plot analyses protein sequences based on their charge and hydropathy. A combination of low overall hydrophobicity and a large net charge are key characteristics of IDP responsible for preventing the hydrophobic collapse into a globular structure. As a result they generally occupy different region to globular proteins in charge-hydrophobicity plots [184]. A plot of the mean net charges and hydrophobicities of hOPN1-157 and a set of reference proteins is shown in Figure 5.2. A boundary is observed between reference proteins known to be globular and those that are intrinsically disordered. The distance of hOPN1-157 from the folded unfolded boundary is shown in the figure, and the protein lies well within the intrinsically disordered region. The charge-hydrophobicity analysis of hOPN1-157 indicates that the chemical properties of hOPN1-157 are similar to those found in IDPs.

PONDR is one of the most reliable and consistently used predictors of disorder in proteins [254]. Analysis by the PONDR predictor using the VSL2 and VLXT algorithm provides information about the propensity of folding on different regions of the protein. According to the PONDR predictions the hOPN1-157 sequence is mainly disordered, and is lacking any kind of secondary structure elements (Figure 5.3). The

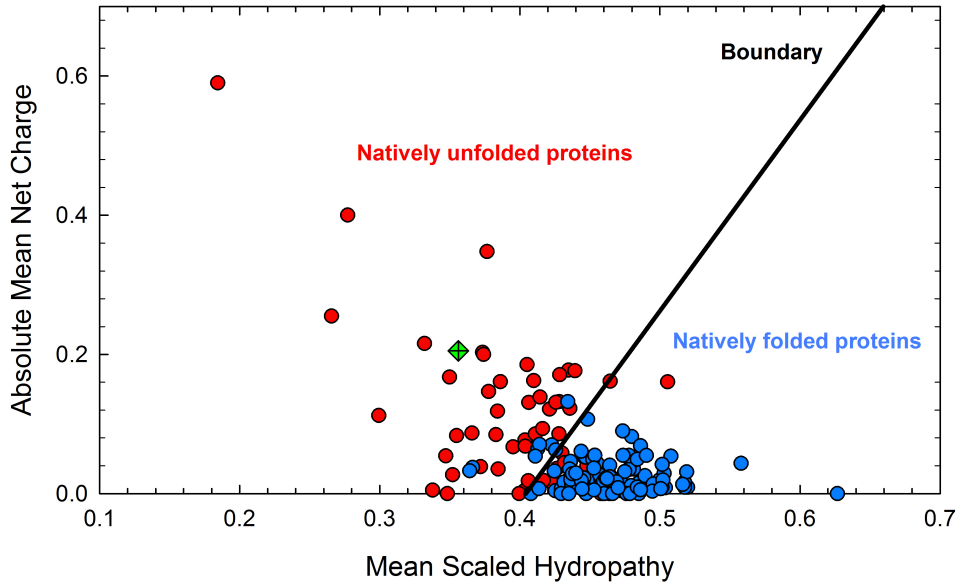

Figure 5.2: Comparison of the mean net charge and the mean hydrophobicity indicated for a set of folded proteins (blue circles) and intrinsically disordered proteins (red circles) data taken from Uversky *et al.* [175]. The boundary between natively folded and unfolded proteins is indicated by a solid black line. hOPN1-157 (Green diamond) lies within the region expected of a unfolded protein.

predictions made by PONDR are in close agreement with the position of hOPN1-157 in the charge-hydropathy plot which also indicated a disordered structure.

5.3.2 Production, purification and phosphorylation of hOPN1-157

Expression of of hOPN1-157 was monitored after induction with IPTG by SDS PAGE analysis. Cells where lysed three hours after induction and the protein was identified in the cell lysate which was used for further protein purification. The first step of purification was completed by the affinity of the his-tagged recombinant protein to a nickel column, eluted fractions were pooled and loaded onto a size exclusion column in order to remove any contaminating species. The collected fractions were then analysed by SDS-page gel as shown in Figure 5.4, the analysis indicates the successful purification of a single species. The identity of the purified species was confirmed by

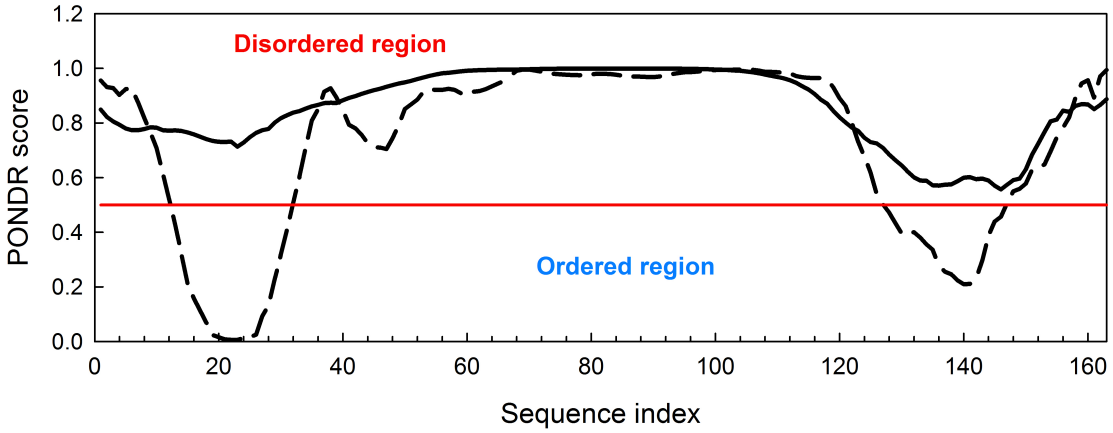

Figure 5.3: Sequence based structural analysis of hOPN1-157using PONDR. The prediction of hOPN1-157 disorder tendency using two different prediction algorithms. Solid line shows the VSL2 algorithm and the dashed line shows the VLXT algorithm. Scores above 0.5 are considered disordered regions (indicated by the red line).

Western blot analysis with antibody specific of the 6 histidine tag, which is not present in any *E.coli* proteins (Figure 5.4). In order to further confirm the identified purified protein the N-terminal 5 residues were sequenced and correctly identified as those expected for hOPN1-157. A combination of N-terminal sequencing, SDS-PAGE gel and Western blot indicate successful production, isolation and purification of recombinant hOPN1-157.

Protein phosphorylation involves the addition of phosphate groups to specific protein side chains, and hence the extent of phosphorylation can be identified by measuring the mass of the protein. In order to determine whether the recombinant expression system produced phosphorylated protein both the HIS tagged proteins were analysed by electron spray ionisation (ESI) mass spectrometry. Unphosphorylated hOPN1-157 showed one species corresponding to the expected molecular mass of the protein in an unphosphorylated state, as would be expected due to the lack of kinase enzymes within *E.coli*. The de-convoluted mass spectrum of phosphorylated hOPN1-157 contained several major species corresponding to different levels of phosphorylation (Table 5.1). The levels of phosphorylation appear to be non specific and correspond to

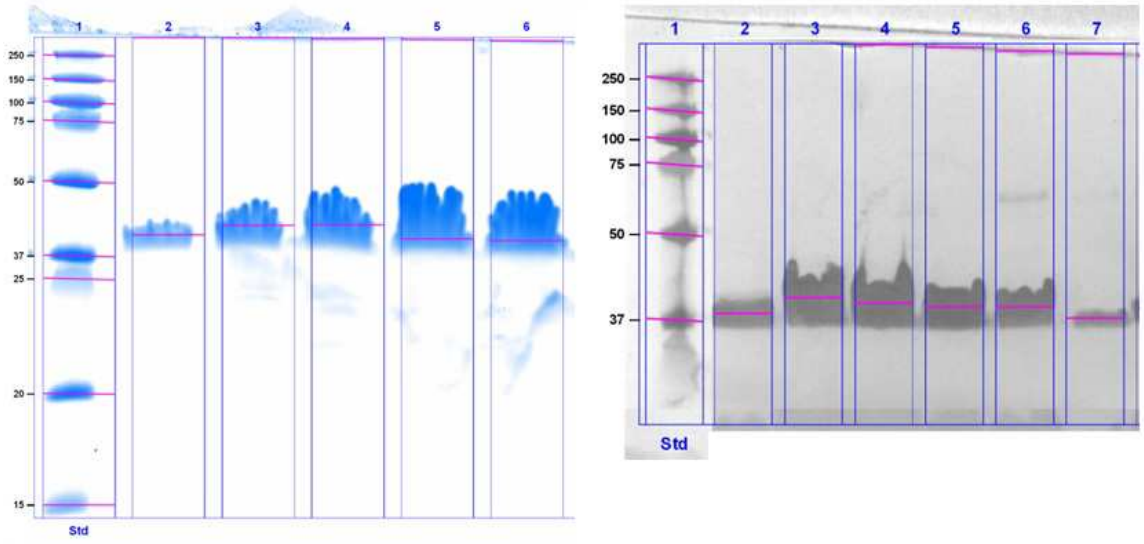

Figure 5.4: Left: SDS-PAGE of hOPN1-157 after size exclusion chromatography. Right: corresponding Western Blot against 6HIS antibody. Molecular masses are indicated in kDa.

levels higher then expected by the inclusion of a single phosphate centre. .

5.3.3 Biophysical characterisation of hOPN1-157

To further investigate the aforementioned predictions experimentally a combination of limited proteolysis, SDS-PAGE gel migration, 1-D HET-SOFAST NMR and analytical size exclusion chromatography (SEC) was used. Intrinsically disordered proteins exhibit larger hydrodynamic radii compared to globular proteins of the same molecular weight [119]. This is caused by their low compactness in solution. Analytical gel filtration was used to calculate the apparent molecular mass of hOPN1-157 according to its migration on a SEC column. The apparent molecular mass can be calculated by measuring the retention time of proteins with a known molecular weight. In this case a reference set of globular proteins was used. In order to determine the effects of phosphorylation on the hydrodynamic radius both phosphorylated and unphosphorylated hOPN1-157 were analysed by this technique. The results are shown in Figure 5.5 A. The retention time of hOPN1-157 is phosphorylation independent, indicating no significant changes in the radius of hydration upon phosphorylation. Based on the

Table 5.1: Species observed by MALDI-TOF measurements of phosphorylated hOPN1-157 and the corresponding number of phosphorylated residues.

Observed mass (Da)	Phosporylated residues
18356.00	3
18435.89	4
18515.86	5
18596.03	6
18675.88	7
18756.04	8
18835.82	9
18915.79	10

elution volume observed for hOPN11-157 the apparent molecular mass was calculated to be 59kDa, which is independent of phosphorylation state and larger then the mass determined by mass spectrometry, and the mass expected based on the amino acid sequence. Migration at a higher apparent molecular weight on a size exclusion column is a typical property of IDP [119].

It is known that intrinsically disordered proteins typically migrate on SDS-PAGE gels at higher molecular masses due to their lack of a hydrophobic core, which leads to binding of SDS in an atypical manner leading to reduced mobility by electrophoresis. The migration of both phosphorylated and un phosphorylated hOPN1-157 on a 12 % SDS-PAGE gel was investigated (Figure 5.5 B). A reference set of globular proteins was used to determine the apparent molecular mass. In a similar manner to the analytical SEC results there seems to be no effect of phosphorlyation on the apparent molecular mass. However the apparent molecular mass at which hOPN1-157 migrates (37 kDa) is much higher then the expected molecular mass. Once again this is another key characteristic of intrinsically disordered proteins, and further indi-

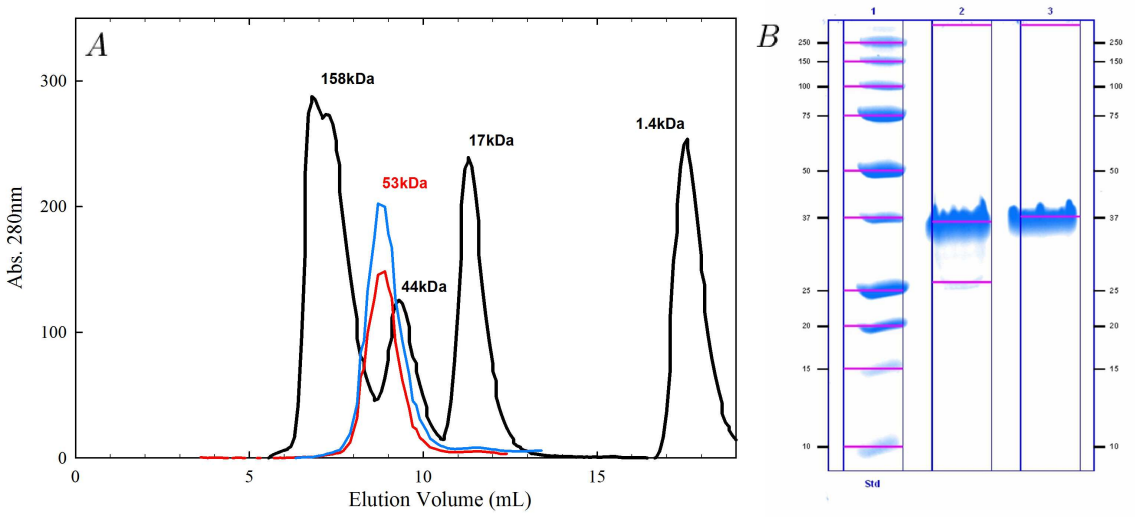

Figure 5.5: (A) Analytical size exclusion chromatography indicates a single peak which is the same for phosphorylated and unphosphorylated hOPN1-157 (green and red curves respectively). Both samples eluted at a higher molecular weight then would be expected for folded proteins. Molecular weight standards are indicated by the blue curve and consist of a set of globular reference proteins. **(B)** Both phosphorylated and unphosphorylated samples migrate at the same rate on a SDS-PAGE gel, at a higher apparent mass then expected

cates a disordered structure of hOPN1-157 [245]. Due to their low compactness and lack of secondary structure, disordered proteins are more susceptible to proteolysis compared to globular proteins, where residues are inaccessible because for example they are buried in the hydrophobic core [119]. The susceptibility to proteolysis is therefore often used to identify and characterise IDP and disordered protein regions [261]. In certain cases phosphorylation has shown to inhibit the degradation of IDP by protease enzymes [262]. To asses the extent of intrinsic disorder in hOPN1-157 the purified protein was exposed in the phosphorylated and unphosphorylated state and myoglobin, a reference protein of similar molecular mass, to low concentrations of proteinase K. In low concentrations and short periods of time proteinase K is capable of rapidly digesting disordered regions, but not globular regions [258]. The ability of proteinase K to digest hOPN1-157 in both the phosphorylated and unphosphorylated state and the reference myoglobin is shown in Figure 5.6.

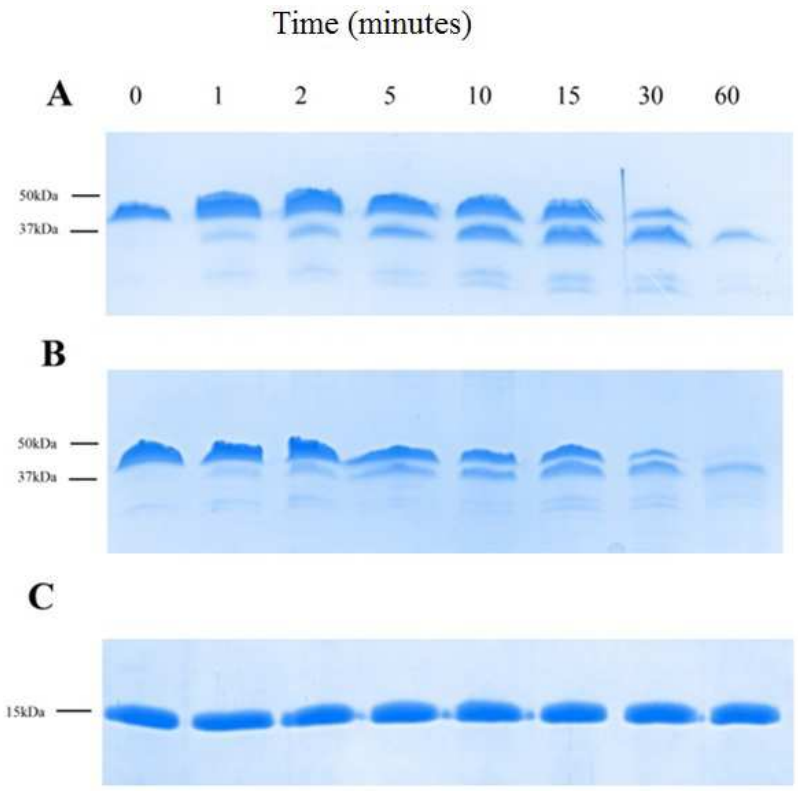

Figure 5.6: Proteinase K induced limited proteolysis at 37 °C of **(A)** Phosphorylated hOPN1-157 **(B)** Unphosporylated hOPN1-157 **(C)** myoglobin. Exposure time for each lane is indicated at the top in minutes.

The intensities of the different bands directly correspond to protein concentration. For both phosphorylated and unphosphorylated hOPN1-157 the intensity decreases rapidly so that after several minutes the protein is heavily digested. Myoglobin on the other hand does not undergo digestion on the same timescale. After 2 hour incubation with protease K both hOPN1-157 samples are fully digested as are the smaller fragments, the myoglobin sample still shows a single band at the correct molecular mass indicating that the protein wasn't digested to the same extent as either hOPN1-157 proteins. After two hours all bands in the hOPN1-157 samples disappear indicate that there is an almost total lack of order in any region of the structure. The high rate of degradation of hOPN1-157 compared to the globular myoglobin protein indicates that hOPN1-157 is heavily exposed to the solvent, independent of phosphorylation state, and hence in both the phosphorylated and unphosphorylated state hOPN1-157

is an IDP. These results agree with the previously presented prediction of intrinsic protein disorder and the high levels of phosphorylation detected within the protein. Using 1-D HET-SOFAST NMR it is possible to get an estimate of the degree of folding within a protein, and hence an indication about the extent of protein flexibility. Samples of both phosphorylated and unphosphorylated hOPN1-157 were further analysed using this technique, which yields a quantitative parameter known as λNOE. Phosphorylated and unphosphorylated hOPN1-157 displayed λNOE of 0.84 and 0.77, respectively (Figure 5.7). The measured λNOE for both samples lie within the region expected of disordered proteins (Figure 5.7), and lies far away from those expected for globular proteins. Furthermore the spectra show an almost total lack of any folded secondary structure elements. There is a slight change in λNOE observed upon phosphorylation of the protein which could correspond to a reduction of mobility upon phosphorylation, however the figure still lies well within the intrinsically disordered region. The above mentioned biophysical analysis provide strong evidence that the hOPN1-157 adopts a nonglobular conformation in solution, which is not phosphorylation dependent.

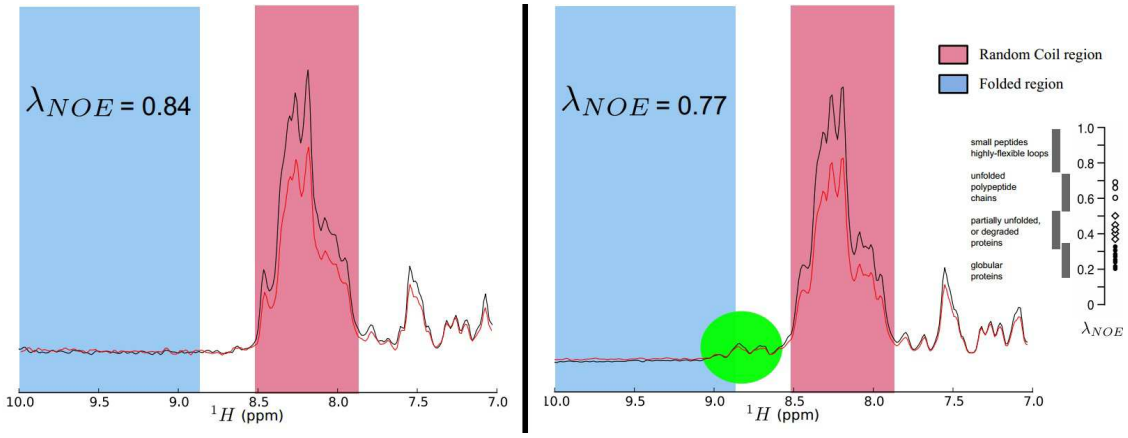

Figure 5.7: One dimensional 1H NMR spectra of the reference and saturated intensities of phosphorylated (left) and unphosphorylated hOPN1-157 (right). The corresponding λ NOE are indicated for both samples.

5.3.4 Structural analysis by FPLC-SAXS

The U.V. traces of the elution from the online FPLC experiments of hOPN1-157 in the phosphorylated and unphosphorylated state are shown in Figure 5.8. Both samples migrate at the same rate observed in the preparation step Figure 5.5 and the retention time is independent of phosphorylation state.

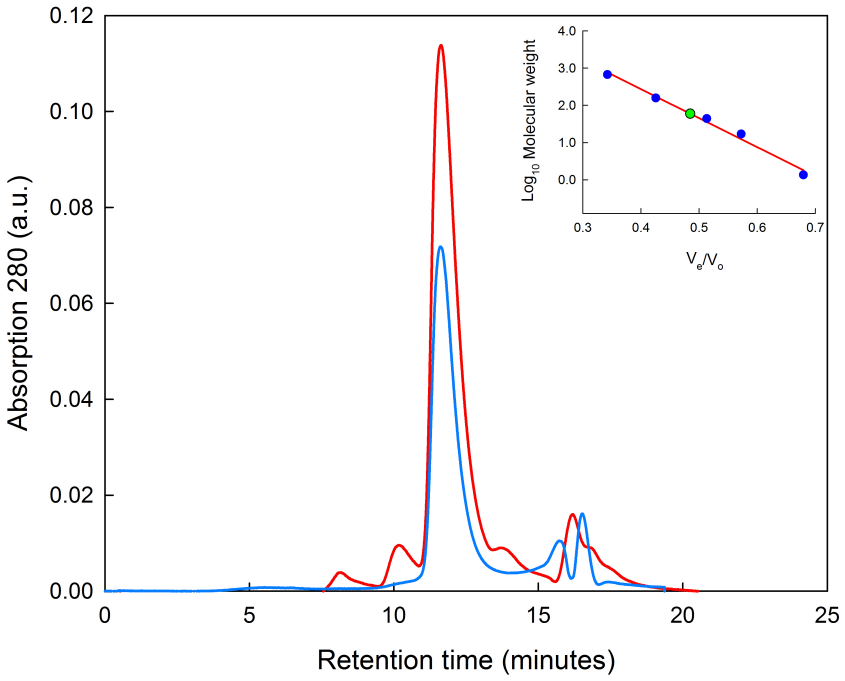

Figure 5.8: Hydrodynamic characterisation of phosphorylated (red) and unphosphorylated hOPN (blue) by analytical size exclusion chromatography, both proteins migrate at a mW corresponding to 59kDa. Inset: Calibration curve using molecular weight standards, indicated by blue dots, Thyroglobulin (670kDa),y-globulin (158kDa), Ovalbumin (44kDa), Myoglobin(17kDa), Vitamin B12(1.35kDa), elution volume of the measured samples are indicated by the green dot.

During the online FPLC-SAXS experiments elution from the column was constantly measured by SAXS. Elution of protein was identified by the U.V. absorption. The EDNA pipeline automatically calculated the R_G and I(0) from the Guinier plot, the results of this are shown in Figure 5.9. The eluted peak corresponds to the peak as measured by U.V. absorption at 280 nm (Figure 5.8). Each frame was then inspected manually for radiation damage and quality of data, for both samples multiple frames

where picked and averaged, providing the final data set. The averaged frames are shown in Figure 5.10 for both samples, the error in radii of gyration is relatively large due to the short exposure of each individual frame. Upon averaging the individual frames (shown in Figure 5.10) the error in data is reduced. Frames from the beginning and end of the elution peak were compared by overlaying the scattering curves. Buffer for background frames containing only buffer were averaged and subtracted from peaks containing protein sample.

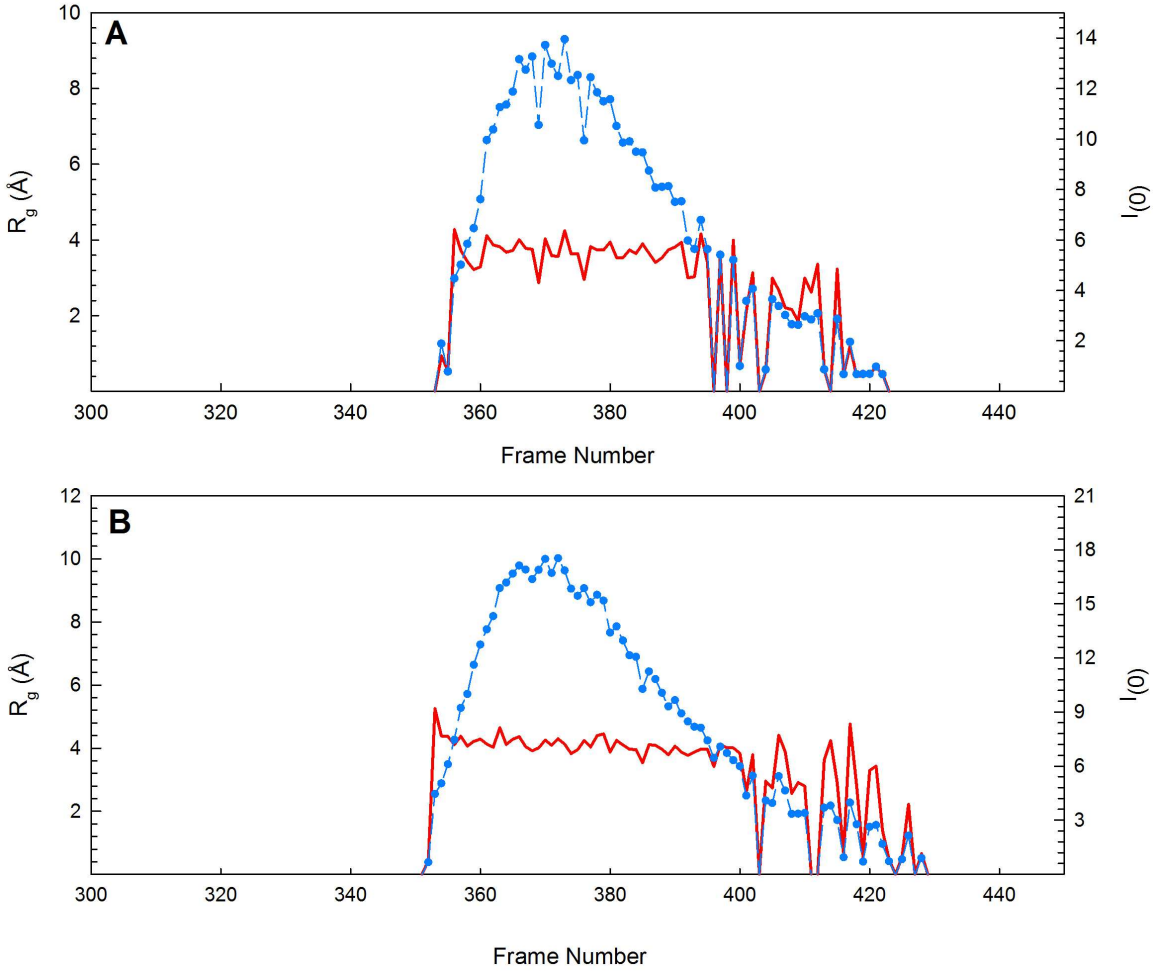

Figure 5.9: Trace of I(0) (blue) and R_G (Red). for **(A)** unphosphorylated hOPN1-157 and **(B)** phosphorylated hOPN1-157, frames as determined by the Guinier plot of the measured FPLC-SAXS. With an exposure time of 2 seconds per frame whilst flowing through a quartz capillary.

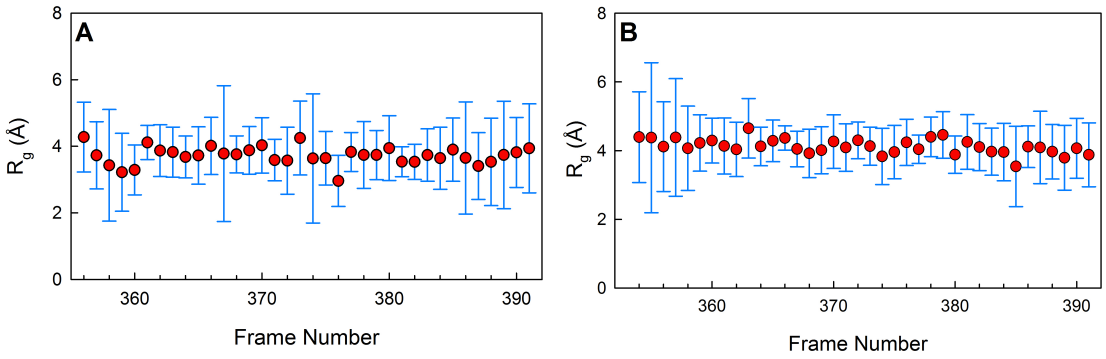

Figure 5.10: The radius of gyration determined for each frame by the Guinier plot with the error shown in blue bars. For **(A)** Unphosphorylated hOPN1-157 and **(B)** Phosphorylated hOPN1-157. For both samples all frames where averaged to provide the final data set.

The R_G determined from the Guinier plots and calculated from the p(r) distributions are shown in Table 5.2 and graphically in Figure 5.11 . The determined R_G are similar between phosphorylated and unphosphorylated hOPN1-157. The expected R_G of an IDP can be calculated based on the amino acid length of a protein using the Flory equation, as can the R_G expected for globular proteins [263]. The calculated RG for globular and IDP are shown in Figure 5.11 C. hOPN1-157 in phosphorylated state and unphosphorylated state lie within the region expected for IDPs, consistent with a highly flexible structur. The calculated p(r) distributions of phosphorylated and unphosphorylated hOPN1-157 are shown in Figure 5.11 D. The distributions are similar and the asymmetrical shape observed in both samples is consistent with that expected for IDPs. From the p(r) distribution the R_G were also obtained in real-space, and lie in close agreement with those obtained from the Guinier plot (Table 5.2). The Guinier plot was also used to determine the forward scattering intensity at $q = 0$, which is dependent upon protein concentration. This value was used to produce a normalised Kratky plot for both samples shown in Figure 5.11E. For both samples the shape of the Kratky plots are identical, with a plateau at high q which corresponds to Gaussian chain, behaviour characteristic of IDPs and flexible proteins [264].

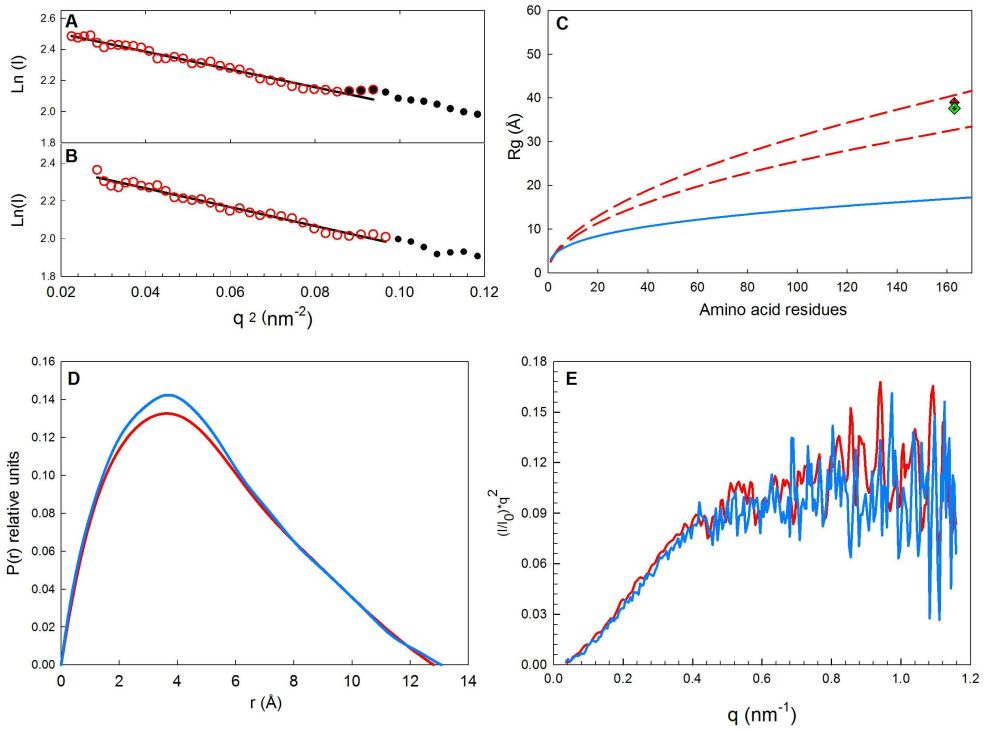

Figure 5.11: Guinier plots of phosphorylated **(A)** and unphosphorylated **(B)** hOPN1-157. The R_G and I(0) were recovered from the fit of the straight line where $q \cdot R_G < 1.3$. **(C)** Comparison of the experimentally determined R_G of phosphorylated (red diamond) and unphosphorylated (Green diamond) with those expected for a disordered protein according to the Flory equation (red dash lines) and the R_G of a globular protein with the same molecular mass. **(D)** The p(r) distribution of phosphorylated and unphosphorylated hOPN1-157. **(E)** Normalized Kratky plot of phosphorylated hOPN1-157 (Red line) and unphosphorylated hOPN1-157 (Blue line).

Table 5.2: SAXS parameters of hOPN1-157 in the unphosphorylated and phosphorylated state

	Phosphorylated	**Unphosphorylated**
R_G-Guinier	39.0 ± 0.59 Å	37.6 ± 0.60 Å
R_G-p(r)	39.8 ± 0.20 Å	39.4 ± 0.26 Å
D_{max}-p(r)	129.4 Å	130.6 Å

SAXS data is often interpreted to obtain information about the size and shape of a protein. Model independent parameters obtained include the shape of the Kratky plot, the p(r) distribution and the R_G. In the case of SAXS the experimentally determined R_G is a size estimate averaged over all conformations in solution, and therefore for proteins with transient structure which display a large range of conformations it is not a very accurate description of the protein in solution. Using the ensemble optimisation method (EOM) a more detailed description of the heterogeneous conformations present in solution can be obtained. The R_G and D_{max} were obtained for the ensemble that best represents the different flexible conformations present in solution for both phosphorylated and unphosphorylated hOPN1-157 (Figure 5.12). The resulting ensemble fits the SAXS data well (Figure 5.12 C and D). Both proteins are represented by a broad ensemble of structures with a R_G ranging from 30 Å up to approximately 50 Å (Figure 5.12 A and B). A similar distribution of sizes is refledted in the distribution of D_{max} ranging from 90 Å to 130 Å . The obtained size distributions are similar for both phosphorylated and unphosphorylated hOPN1-157, indicating that the protein occupies a similar set of transient structures which are not greatly affected by extensive protein phosphorylation. The obtained distribution of sizes can be compared to an ensemble randomly generated based on the amino acid sequence of the protein. This ensemble is shown as a black line in Figure 5.12A and

B. The distributions of hOPN1-157 are broadly similar to the random ensemble and it samples nearly all the same conformations. The random pool sample conformations at higher values of D_{max} and R_G compared to the experimental data, indicating that hOPN1-157 possibly is not as extended as a completely random structure. This tendency does not appear to be affected by phosphorylation of hOPN1-157.

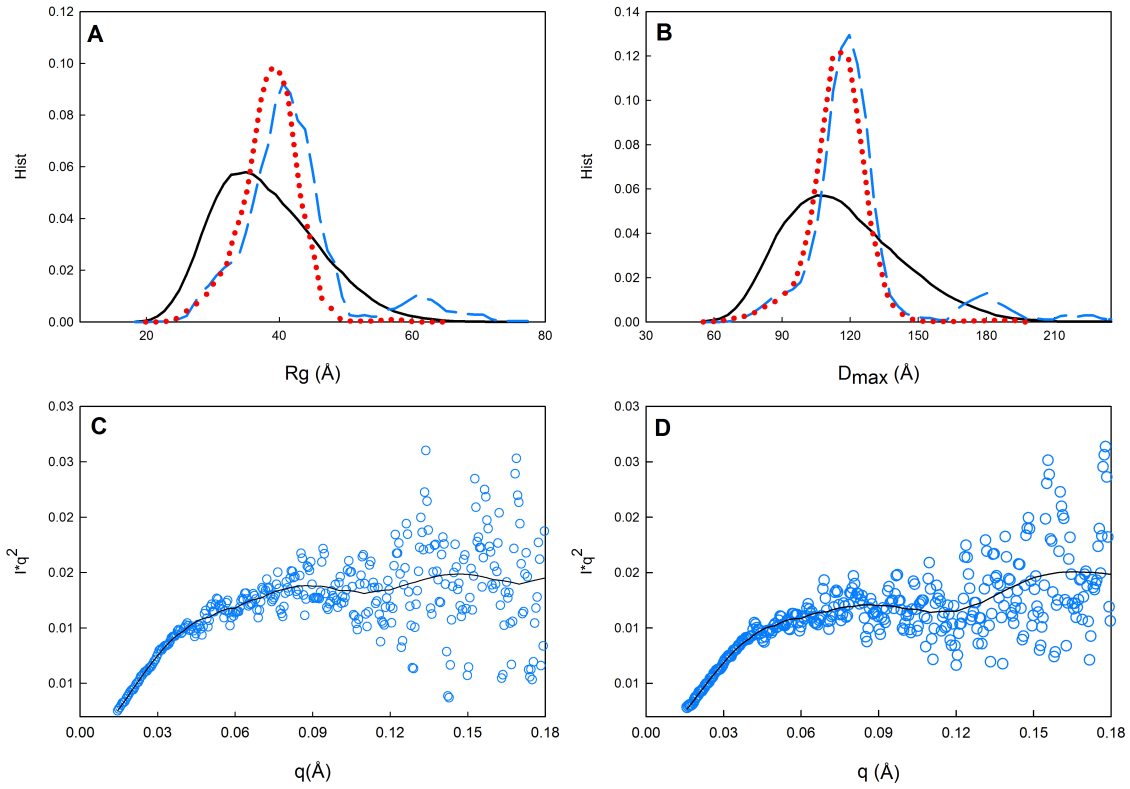

Figure 5.12: Ensemble optimisation analysis of the FPLC-SAXS profile measured for hOPN in the phosphorylated and unphosphorylated state. **(A)** R_G distribution of the samples and the random ensemble (black line) (phosphorylated hOPN1-157 blue line, unphosphorylated hOPN1-157 red line). **(B)** D_{max} distribution shown for phosphorylated hOPN1-157 blue line, unphosphorylated hOPN1-157 red line and the pool of random structures black line. **(C)** Fit of the selected pool to the hOPN1-157 phosohporylated scattering curve ($\chi^2 = 1.090$) **(D)** Fit of the selected pool to the unphosphorylated hOPN1-157 scattering curve $\chi^2 = 1.064$).

5.3.5 Effects of phosphorylation on protein dynamics

The mean squared displacement (MSD) of the samples were determined on the nano-second timescale using elastic incoherent neutron scattering on the IRIS instrument at ISIS. The MSDs were extracted as a function of temperature and are shown in Figure 5.13. The large incoherent scattering cross section of hydrogen dominates the scattering by the hydrated sample. Therefore the obtained MSDs can be related to the average movement of hydrogen which are spread across the sidechains of hOPN1-157. According to the obtained results there does not seem to be a significant effect of phosphorylation upon the MSD on the nano second timescale.

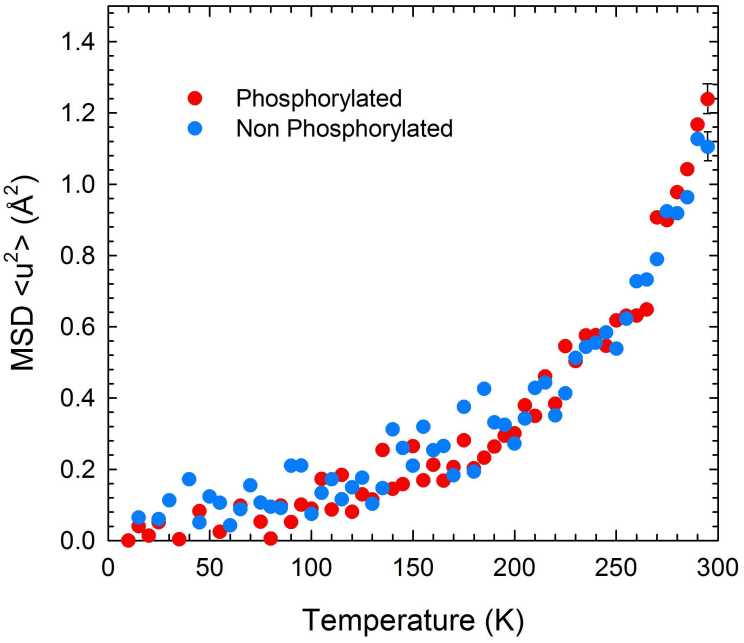

Figure 5.13: MSDs of phosphorylated (Red circles) and unphosphorylated (Blue circles) hOPN1-157. Error bar is shown at the highest temperature and decreases with decreasing temperature. Data was measured on the IRIS spectrometer at ISIS

A break in the increase of the MSD at around 200 K can be seen for both samples. This has been observed in several hydrated protein powders studied by neutron scattesring and is called the glass transition temperature [265]. The glass transition temperature

can be interpreted in terms of the two well model, below the transition temperature the atoms do not have enough energy to sample all structural fluctuations and only samples the lower energy well, the atoms are harmonic [266]. At temperatures above the glass transition the atoms within the sample have enough energy to sample all structural configurations, the atoms become anharmonic [266]. This anharmonic transition is present in both samples at similar points. At temperatures above the glass transition no significant differences are observed in the measured MSDs. The results indicated that on the nano-second timescale the motions of hOPN1-157 are not influenced by the state of phosphorylation of the protein.

5.4 Discussion

The hOPN1-157 protein was readily expressed by the host organism with a yield of around 10 mg per litre of culture for both the phosphorylated and unphosphorylated form. By inclusion of both the hOPN1-157 vector and the vector containing CK2α in the same bacteria, expression of phosphorylated hOPN1-157 was achieved. The hOPN1-157 vector contained 3 serines in a row placed within the recognition sequence of CK2α. Mass spectrometric analysis of the purified hOPN1-157 protein showed however that phosphorylation of hOPN1-157 by CK2α was non specific. Instead of phosphorylation of only the desired 3 residues several other side chains were also phosphorylated to some degree. 1 D HET-SOFAST NMR provides a method to semi-quantitatively asses the structural compactness and heterogeneity of proteins in solution. The measurements performed on hOPN1-157 in the phosphorylated and unphosphorylated state and λNOE values were calculated to be 0.84 and 0.77 respectively, indicative of highly flexible proteins in solution. The results obtained by NMR were highly similar to those previously observed for another OPN fragment [84]. The structural studies that followed indicated that hOPN1-157 has a disordered

and highly flexible structure. The nonspecific phosphorylation observed in the recombinant hOPN 1-157 can be understood by a distinct lack of structural elements. The intrinsic flexibility of the hOPN1-157 peptide allows for wider recognition by kinases due to the increased exposure of the protein chain to solvent. Indeed this was observed by the sensitivity of hOPN1-157 by proteinase K, which was higher then that observed in the globular myoglobin protein (see Figure 5.6). Recently it has been observed that the kinase used in this study (CK2α) is not as specific as previously thought. Previously it was thought that CK2α recognised a specific motif our study shows that this is not the case. It may instead be that the specific motif increases affinity to CK2α and increases the rate of phosphorylaton. In the case of OPN1-157 the increased exposure of all the sidechains due to intrinsic disorder results in phosphorylation of multiple residues outside of the CK2α recognition sequence.
Calculation of the apparent molecular mass of the protein by both SDS-PAGE and analytical SEC yielded a much higher molecular weight then would be expected based on the amino acid sequence (Figure 5.5). According to the amino acid analysis hOPN1-157 contains a high degree of negatively charged amino acids, compared to globular proteins Figure 5.2. The aberrant migration of hOPN1-157 by SDS-PAGE is due to this high content of negatively charged amino acids which inhibit the binding of SDS and in doing so reduces the migration of the protein on the gel. This behaviour is typical of proteins that adopt an intrinsically disordered conformation as observed for a number of IDPs [267, 193, 263]. Analytical size exclusion chromatography provides a estimate of the molecular mass depending on the hydrodynamic radius of the sample. For hOPN1-157 the calibrated chromatogram indicates that the protein elutes at a molecular mass which corresponds to that of a globular protein of a much higher weight. Uversky explained the difference in the theoretical molecular mass and that observed by analytical SEC as due to the larger hydrodynamic radii of intrinsically disordered proteins compared to globular proteins.

In order to obtain good quality data by small-angle X-ray scattering, particularly for visualising small conformational changes in heterogeneous structures, sample of high purity lacking in aggregates is important. Protein aggregation is common at high concentrations but the formation of aggregates is not necessarily concentration dependent [268]. SAXS data analysis can lead to false interpretations if the sample contains only a small amount of aggregate. Error introduced by measurement of aggregates particularly affects data at low resolution and is typically observed in the Guinier region, thereby influencing the model independent determined radius of gyration and I(0) [148]. To overcome the problem of protein aggregation and in order to determine even small changes in the R_G online FPLC-SAXS has been developed at several SAXS beamlines [269, 270, 150]. Measurement of the protein or a complex of proteins immediately after size exclusion purification improves the sample homogeneity and increases the quality of data. Several experiments have been performed that have identified small changes in the size of the macro-molecule by the FPLC-SAXS method [270, 271]. FPLC-SAXS is therefore ideally suited for proteins which display heterogeneous conformations allowing the separation of different structural conformers based on their size and migration rate through the size exclusion column. This technique is therefore ideally suited to investigate whether the broad band eluting for hOPN1-157 from a SEC column is due to different conformations or purely the structural inhomogeneity caused by hOPN1-157. The measured R_G was consistent across all frames measured during the FPLC-SAXS experiment indicative of only a singular species, the frames were therefore averaged to provide the final data. Model independent data analysis of the FPLC-SAXS data by means of the Kratky and Guinier plot indicate that hOPN1-157 adopts an unfolded conformation in both phosphorylated and unphosphorylated states. The ensemble optimisation method was used to characterise the heterogeneous ensemble of structures present in solution. The generated ensemble of both phosphorylated and unphosphorylated

hOPN1-157 displays a highly extended and flexible structure which is identical for both. From the selected ensemble of structures it can be observed that in solution hOPN1-157 adopts a wide range of conformations from more compact structures to highly extended conformations. The result is in agreement with both the 1D-HET-SOFAST NMR and the model independent SAXS data. Interestingly the ensemble of heterogeneous structures cannot be separated out by FPLC SAXS, indicating that the protein is a heterogeneous ensemble of structures that is transient in solution.

From the SAXS data information can be obtained about the structural characteristics; however this data does not yield any information about the dynamical properties of the protein in solution. In order to assess the effects of phosphorylation on the picosecond timescale we studied the mean squared displacements of the samples using elastic incoherent neutron scattering on IRIS at ISIS. From both the data the mean square displacements were calculated as a function of temperature. The results shown in Figure 5.13 indicate no significant change in protein dynamics upon phosphorylation of hOPN1-157 on the picosecond timescale.

The experimental evidence presented here indicates no significant alteration in either protein structure or dynamics on the timescales measured upon phosphorylation of the protein. Phosphorlyation of hOPN does not affect structure or dynamics but in other osteopontin fragments phosphorylation does affect function; of primary importance is the altered electrostatic effects of the protein which are greatly altered by phosphorylation [59]. The size of the core of a calcium phosphate nanocluster depends on the sequestering power of the peptide [79]. In the case of CPN formation, phosphorylation is required to provide sufficient sequestering power to the phosphate centres affording sufficient free energy for sequestration. The importance of phosphorylation in hOPN1-157 is therefore not structural or dynamical but rather phosphorylation alters the electrostatic properties of the protein. Multiple phosphorylation of the protein in regions previously defined as phosphate centers leads to the formation of a

so called net charge or mean field [272]. The disordered state of hOPN1-157 in solution and the resulting rapid interconversion (observed by EINS) through a multitude of transient conformations means that the protein does not present discrete charges through the individual phosphorylated amino acids but rather a mean electrostatic field, which represents the sum of all the charges in the phosphate centre, is presented [272, 273] . It has previously also been suggested that the adoption of an intrinsically disordered conformation is a requirement for calcium phosphate sequestration because it allows the protein to be concentrated to a high degree on the core surface of a calcium phosphate nanocluster [23]. This combination of properties between the mean field and the ability to pack densely on the core of a calcium phosphate nanocluster may be the most important properties of OPN in terms of sequestration rather than the exact conformations adopted in solution. The unfolded conformation of hOPN1-157 may not only be important for the act of sequestration but also to allow the protein to undertake its multifunctional roles *in vivo*. As observed here, by the susceptibility of proteinase K and the high degree of phosphorylation of hOPN1-157 by CK2α, a disordered conformation allows easier access to amino acid residues by enzymes which undertake post-translational modifications. In the case of CPN formation regulation of the mean electrostatic field and hence protein interaction can be more rapidly controlled as well as post-translational modifications that regulate other functions *in vivo*.

5.5 Conclusion

hOPN1-157 samples a transient ensemble of conformations in solution independent of phosphorylation state. The rate at which this ensemble is sampled is also not effected by the phosphorylated state of the protein. The nonspecific phosphorylation observed in the recombinant hOPN1-157 can be understood by structural elements. Firstly the

intrinsic flexibility of the hOPN1-157 peptide allows for wider recognition not only by kinases but also by other enzymes which undertake post-translational modification due to the increased exposure of the protein chain to solvent. Secondly nonspecific phosphorylation of hOPN by CK2α could be due to the apparent non-specificity of the CK2A enzyme itself, which has been proposed to phosphorylate and regulate a wide range of biological processes. This high propensity to interact with other proteins could explain how intrinsically disordered proteins can participate in many different cellular processes. The combination of disorder and regions which contain a high amount of phosphorylation allows OPN1-157 to present a mean electrostatic field. It is through the mean charge of the electrostatic field that OPN peptides can sequester amorphous calcium phosphate and pack densely around the core leading to the formation of CPN.

Chapter 6

Summary & conclusions

The main results from this work related to osteopontin phosphor-peptides and the calcium phosphate nanoclusters formed upon sequestration of calcium phosphate. The key mechanism of calcium phosphate sequestration requires phosphate centres within the sequestering protein. Formation of calcium phosphate nanoclusters leads to the stabilisation of biofluids and allows the concentration of calcium and phosphate ions to exceed their saturation level. It is believed that the calcium phosphate nanoclusters also provide a biologically available source of calcium phosphate for the growth of mineralised tissue *in vivo*.

In this work several techniques were used to study the structure of OPN1-149 isolated from bovine milk. Biophysical characterisation showed that OPN1-149 displays features of a disordered protein. Consequently SAXS was used to study the conformation of OPN1-149 in solution. Model free interpretation of the SAXS data showed properties consistent with those expected of intrinsically disordered protein. The data were further interpreted using the ensemble optimisation method which showed that OPN1-149 adopts a range of transient structures in solution ranging from extended to compact. Elastic incoherent neutron scattering was then used to obtain information about the rate at which the OPN1-149 peptide samples the heterogeneous conformations present in solution. The mean force constant was extracted from a EINS

temperature scan. Comparison of the obtained mean force constant with previous results showed that the value obtained for OPN1-149 closely matches those observed in another IDP. The resilience of OPN1-149 was also studied after sequestration of calcium phosphate, the obtained force constant showed an increase in rigidity, showing a difference in behaviour upon sequestration.

Formed OPN1-149-CPNs were studied using small-angle neutron scattering in combination with contrast variation. Static light scattering showed no change in the size of OPN1-149 CPN upon variation of D_2O in the buffer. The size determined by static light scattering closely matched the size previously obtained by SAXS which are far larger than previously observed in any other CPN system. The match-point of OPN1-149 was confirmed. SANS studies of the formed CPN showed properties of core-shell type particles. From the SANS data the radius of the core and shell were calculated and the thickness of the shell obtained. The shell thickness was comparable to those observed in other CPN. The core radius was, however, larger than those observed in other CPN. The core structure was studied using a range of techniques to rule out formation of crystalline material, including neutron and X-ray diffraction. No crystallinity was observed. The increased size of the core and whole CPN particle was therefore due to the increased power of sequestration of OPN1-149.

In order to elucidate the role of osteopontin in human a biofluids, a recombinant *E.coli* expression system was used to produce a phosphorylated human OPN fragment. The system relied on two plasmids, one which expressed the protein and one which expressed CK2α. Mass spectrometry experiments of the phosphorylated protein showed phosphorylation in a non specific manner. This was related to the intrinsic disorder observed in the protein by SAXS and other biophysical techniques. Intrinsic disorder in proteins leaves the protein more susceptible to solvent and hence interaction with

other proteins. It is thought that this intrinsic disorder allows unfolded proteins such as OPN to rapidly sequester excess calcium phosphate prior to precipitate formation. The dynamics of hOPN1-157 were also studied using EINS, however no effects of phosphorylation were observed. Phosphorylation of hOPN1-157 has no detectable effect on the structure or dynamics of the protein, however introducing large phosphorylated areas on the protein does effect the electrostatic properties. Thus it can be concluded that phosphorylation regulates hOPN1-157 function merely by modifying its electrostatic properties rather than its conformational properties.

References

[1] C M Giachelli. Ectopic calcification: gathering hard facts about soft tissue mineralization. *The American journal of pathology*, 154(3):671–5, 1999.

[2] Maximilien E. Launey, Markus J. Buehler, and Robert O. Ritchie. On the Mechanistic Origins of Toughness in Bone. *Annual Review of Materials Research*, 40(1):25–53, 2010.

[3] Arun K Nair, Alfonso Gautieri, Shu-Wei Chang, and Markus J Buehler. Molecular mechanics of mineralized collagen fibrils in bone. *Nature communications*, 4:1724, 2013.

[4] Matthew J. Olszta, Xingguo Cheng, Sang Soo Jee, Rajendra Kumar, Yi-Yeoun Kim, Michael J. Kaufman, Elliot P. Douglas, and Laurie B. Gower. Bone structure and formation: A new perspective. *Materials Science and Engineering: R: Reports*, 58(3-5):77–116, 2007.

[5] Ego Seeman and Pierre D Delmas. Bone quality–the material and structural basis of bone strength and fragility. *The New England journal of medicine*, 354(21):2250–61, 2006.

[6] F. C. M. Driessens. The mineral in bone, dentin and tooth enamel. *Bulletin des Sociétés Chimiques Belges*, 89(8):663–689, 2010.

[7] M. Larsen and B. Nyvad. Enamel Erosion by Some Soft Drinks and Orange Juices Relative to Their pH, Buffering Effect and Contents of Calcium Phosphate. *Caries Research*, 33(1):81–87, 1998.

[8] A. Bardow, B. Nyvad, and B. Nauntofte. Relationships between medication intake, complaints of dry mouth, salivary flow rate and composition, and the rate of tooth demineralization in situ. *Archives of Oral Biology*, 46(5):413–423, 2001.

[9] Guy H Carpenter. The secretion, components, and properties of saliva. *Annual review of food science and technology*, 4:267–76, 2013.

[10] Carl Holt, Samuel Lenton, Tommy Nylander, Esben S Sø rensen, and Susana C M Teixeira. Mineralisation of soft and hard tissues and the stability of biofluids. *Journal of structural biology*, 185(3):383–396, 2014.

[11] C. Holt and JA Carver. Darwinian transformation of a 'scarcely nutritious fluid'into milk. *Journal of evolutionary biology*, 25(7):1253–1263, 2012.

[12] Carl Holt, Douglas G. Dalgleish, and Robert Jenness. Calculation of the ion equilibria in milk diffusate and comparison with experiment. *Analytical Biochemistry*, 113(1):154–163, 1981.

[13] Hans-Göran Tiselius, Bengt Lindbäck, Anne-Marie Fornander, and Mari-Anne Nilsson. Studies on the role of calcium phosphate in the process of calcium oxalate crystal formation. *Urological research*, 37(4):181–92, 2009.

[14] MSA S Johnsson and GH H Nancollas. The role of brushite and octacalcium phosphate in apatite formation. *Critical reviews in oral biology and medicine*, 3(1-2):61–82, 1992.

[15] C. Holt, J.A. Carver, H. Ecroyd, and D.C. Thorn. Invited review: Caseins and the casein micelle: Their biological functions, structures, and behavior in foods. *Journal of dairy science*, 96(10):6127–6146, 2013.

[16] Jaroslav Nývlt. The Ostwald Rule of Stages. *Crystal Research and Technology*, 30(4):443–449, 1995.

[17] JL Meyer and ED Eanes. A thermodynamic analysis of the amorphous to crystalline calcium phosphate transformation. *Calcified Tissue International*, 68:59–68, 1978.

[18] Yue Liu and George H. Nancollas. Crystallization and Colloidal Stability of Calcium Phosphate Phases. *The Journal of Physical Chemistry B*, 101(18):3464–3468, 1997.

[19] A. Bigi, E. Foresti, A. Incerti, N. Roveri, P.A. Borea, and G. Zavagli. Chemical and structural study of the mineral phase associated with a human subcutaneous ectopic calcification. *Inorganica Chimica Acta*, 46:271–274, 1980.

[20] Leslie S. Burstein, Adele L. Boskey, Paul J. Tannenbaum, Aaron S. Posner, and Irwin D. Mandel. The crystal chemistry of submandibular and parotid salivary gland stones. *Journal of Oral Pathology and Medicine*, 8(5):284–291, 1979.

[21] Georg Schlieper, Anke Aretz, Steven C Verberckmoes, Thilo Krüger, Geert J Behets, Reza Ghadimi, Thomas E Weirich, Dorothea Rohrmann, Stephan Langer, Jan H Tordoir, Kerstin Amann, Ralf Westenfeld, Vincent M Brandenburg, Patrick C D'Haese, Joachim Mayer, Markus Ketteler, Marc D McKee, and Jürgen Floege. Ultrastructural analysis of vascular calcifications in uremia. *Journal of the American Society of Nephrology*, 21(4):689–96, 2010.

[22] Jae Sam Lee, Joel D Morrisett, and Ching-Hsuan Tung. Detection of hydroxyapatite in calcified cardiovascular tissues. *Atherosclerosis*, 224(2):340–7, 2012.

[23] Carl Holt. Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. *Current opinion in structural biology*, 23(3):420–5, 2013.

[24] H A Krebs. Chemical composition of blood plasma and serum. *Annual review of biochemistry*, 19:409–30, 1950.

[25] Peter Lanzer, Manfred Boehm, Victor Sorribas, Marc Thiriet, Jan Janzen, Thomas Zeller, Cynthia St Hilaire, and Catherine Shanahan. Medial vascular calcification revisited: review and perspectives. *European heart journal*, 14(35):1515–1525, 2014.

[26] Yen-Pin Lin and Chih-Hao Chen. Huge tumoral calcinosis of the buttock. *Formosan Journal of Surgery*, 47(1):23–27, 2014.

[27] E. D. Chan, D.V. Morales, C.H. Welsh, M.T. McDermott, and Schwarz. M.I. Calcium Deposition with or without Bone Formation in the Lung. *American Journal of Respiratory and Critical Care Medicine*, 165(12):1654–1669, 2002.

[28] F T Hammad, M Kaya, and E Kazim. Bladder calculi: did the clinical picture change? *Urology*, 67(6):1154–8, 2006.

[29] Jordan D Miller, Robert M Weiss, and Donald D Heistad. Calcific aortic valve stenosis: methods, models, and mechanisms. *Circulation research*, 108(11):1392–1412, 2011.

[30] Qiaoli Li, Qiujie Jiang, and Jouni Uitto. Ectopic mineralization disorders of the extracellular matrix of connective tissue: Molecular genetics and pathomechanisms of aberrant calcification. *Matrix biology*, 33C:23–28, 2014.

[31] Meiting Wu, Cameron Rementer, and Cecilia M Giachelli. Vascular calcification: an update on mechanisms and challenges in treatment. *Calcified tissue international*, 93(4):365–73, 2013.

[32] V. Benigno. Hypomagnesaemia-hypercalciuria-nephrocalcinosis: a report of nine cases and a review. *Nephrology Dialysis Transplantation*, 15(5):605–610, 2000.

[33] Sean Coffey, Brian Cox, and MJA Williams. The prevalence, incidence, progression, and risks of aortic valve sclerosis: a systematic review and meta-analysis. *Journal of the American ...*, 63(25):2852–2861, 2014.

[34] Jeremy J Thaden, Vuyisile T Nkomo, and Maurice Enriquez-Sarano. The global burden of aortic stenosis. *Progress in cardiovascular diseases*, 56(6):565–71, 2014.

[35] annurev-food-022811-101214 (Objet application/pdf).

[36] Nalini M Rajamannan, Robert O Bonow, and Shahbudin H Rahimtoola. Calcific aortic stenosis: an update. *Nature clinical practice. Cardiovascular medicine*, 4(5):254–62, 2007.

[37] B F Stewart, D Siscovick, B K Lind, J M Gardin, J S Gottdiener, V E Smith, D W Kitzman, and C M Otto. Clinical factors associated with calcific aortic valve disease. Cardiovascular Health Study. *Journal of the American College of Cardiology*, 29(3):630–4, 1997.

[38] Blase A Carabello and Walter J Paulus. Aortic stenosis. *Lancet*, 373(9667):956–66, 2009.

[39] S. Yusuf, S. Reddy, S. Ounpuu, and S. Anand. Global Burden of Cardiovascular Diseases: Part I: General Considerations, the Epidemiologic Transition, Risk Factors, and Impact of Urbanization. *Circulation*, 104(22):2746–2753, 2001.

[40] Doddametikurke Ramegowda Basavaraj, Chandra Shekhar Biyani, Anthony J. Browning, and Jon J. Cartledge. The Role of Urinary Kidney Stone Inhibitors and Promoters in the Pathogenesis of Calcium Containing Renal Stones. *EAU-EBU Update Series*, 5(3):126–136, June 2007.

[41] David C. Thorn, Heath Ecroyd, John A. Carver, and Carl Holt. Casein structures in the context of unfolded proteins. *International Dairy Journal*.

[42] Shuichi Ito, Takashi Saito, and Kaori Amano. In vitro apatite induction by osteopontin: interfacial energy for hydroxyapatite nucleation on osteopontin. *Journal of biomedical materials research. Part A*, 69(1):11–6, 2004.

[43] David a Pampena, Karen a Robertson, Olga Litvinova, Gilles Lajoie, Harvey a Goldberg, and Graeme K Hunter. Inhibition of hydroxyapatite formation by osteopontin phosphopeptides. *The Biochemical journal*, 378(Pt 3):1083–1087, 2004.

[44] John R de Bruyn, Maria Goiko, Maryam Mozaffari, Daniel Bator, Ron L Dauphinee, Yinyin Liao, Roberta L Flemming, Michael S Bramble, Graeme K Hunter, and Harvey a Goldberg. Dynamic light scattering study of inhibition of nucleation and growth of hydroxyapatite crystals by osteopontin. *PloS one*, 8(2):e56764, 2013.

[45] Carl Holt, Esben S Sø rensen, and Roger A Clegg. Role of calcium phosphate nanoclusters in the control of calcification. *The FEBS journal*, 276(8):2308–2323, 2009.

[46] W G Robertson, M Peacock, and B E Nordin. Activity products in stone-forming and non-stone-forming urine. *Clinical science*, 34(3):579–94, 1968.

[47] T. Brooks and C.W. Keevil. A simple artificial urine for the growth of urinary pathogens. *Letters in Applied Microbiology*, 24(3):203–206, 1997.

[48] V W Leung and B W Darvell. Artificial salivas for in vitro studies of dental materials. *Journal of dentistry*, 25(6):475–84, 1997.

[49] N J Cochrane and E C Reynolds. Calcium phosphopeptides – mechanisms of action and evidence for clinical efficacy. *Advances in dental research*, 24(2):41–7, 2012.

[50] Graeme K Hunter, Jason O'Young, Bernd Grohe, Mikko Karttunen, and Harvey a Goldberg. The flexible polyelectrolyte hypothesis of protein-biomineral interaction. *Langmuir*, 26(24):18639–18646, 2010.

[51] Keith Alvares. The role of acidic phosphoproteins in biomineralization. *Connective tissue research*, 55(1):34–40, 2014.

[52] Anne George and Arthur Veis. Phosphorylated proteins and control over apatite nucleation, crystal growth, and inhibition. *Chemical reviews*, 108(11):4670–4693, 2008.

[53] Charles Sfeir, Ping-An Fang, Jayaraman Thottala, Aparna Raman, Zhang Xiaoyuan, and Elia Beniash. Synthesis of bone-like nanocomposites using multiphosphorylated peptides. *Acta biomaterialia*, 10(5):2241–2249, 2014.

[54] Alexander Heiss, Vitaliy Pipich, Willi Jahnen-Dechent, and Dietmar Schwahn. Fetuin-A is a mineral carrier protein: small angle neutron scattering provides new insight on Fetuin-A controlled calcification inhibition. *Biophysical journal*, 99(12):3986–95, 2010.

[55] Paul A Price and Joo Eun Lim. The inhibition of calcium phosphate precipitation by fetuin is accompanied by the formation of a fetuin-mineral complex. *The Journal of biological chemistry*, 278(24):22144–22152, 2003.

[56] Yung-Ching Chien, David L Masica, Jeffrey J Gray, Sarah Nguyen, Hojatollah Vali, and Marc D McKee. Modulation of calcium oxalate dihydrate growth by selective crystal-face binding of phosphorylated osteopontin and polyaspartate peptide showing occlusion by sectoral (compositional) zoning. *The Journal of biological chemistry*, 284(35):23491–501, 2009.

[57] Laura Brylka and Willi Jahnen-Dechent. The role of fetuin-A in physiological and pathological mineralization. *Calcified tissue international*, 93(4):355–64, 2013.

[58] C Holt. The milk salts and their interaction with casein. *Advanced Dairy Chemistry*, Volume 3:917–924, 1997.

[59] Adele L Boskey, Brian Christensen, Hayat Taleb, and Esben S Sø rensen. Post-translational modification of osteopontin: Effects on in vitro hydroxyapatite formation and growth. *Biochemical and biophysical research communications*, 9(419):333–338, 2012.

[60] Kevin M Zurick, Chunlin Qin, and Matthew T Bernards. Mineralization induction effects of osteopontin, bone sialoprotein, and dentin phosphoprotein on a biomimetic collagen substrate. *Journal of biomedical materials research. Part A*, 101(6):1571–81, 2013.

[61] Douglas E Rodriguez, Taili Thula-Mata, Edgardo J Toro, Ya-Wen Yeh, Carl Holt, L Shannon Holliday, and Laurie B Gower. Multifunctional role of osteopontin in directing intrafibrillar mineralization of collagen and activation of osteoclasts. *Acta biomaterialia*, 10(1):494–507, 2014.

[62] Willi Jahnen-Dechent, Cora Schäfer, Markus Ketteler, and Marc D McKee. Mineral chaperones: a role for fetuin-A and osteopontin in the inhibition and regression of pathologic calcification. *Journal of molecular medicine*, 86(4):379–89, 2008.

[63] Lijun Wang, Xiangying Guan, Ruikang Tang, and John R Hoyer. Phosphorylation of Osteopontin Is Required for Inhibition of Calcium Oxalate Crystallization. *Journal de Physique*, 112(30):9151–9157, 2009.

[64] Susan A Steitz, Mei Y Speer, Marc D McKee, Lucy Liaw, Manuela Almeida, Hsueh Yang, and Cecilia M Giachelli. Osteopontin inhibits mineral deposition and promotes regression of ectopic calcification. *The American journal of pathology*, 161(6):2035–46, 2002.

[65] T. Schinke, C. Amendt, A. Trindl, O. Poschke, W. Muller-Esterl, and W. Jahnen-Dechent. The Serum Protein 2-HS Glycoprotein/Fetuin Inhibits Apatite Formation in Vitro and in Mineralizing Calvaria Cells. *Journal of Biological Chemistry*, 271(34):20789–20796, 1996.

[66] Alexander Heiss, Thomas Eckert, Anke Aretz, Walter Richtering, Wim van Dorp, Cora Schäfer, and Willi Jahnen-Dechent. Hierarchical role of fetuin-A and acidic serum proteins in the formation and stabilization of calcium phosphate particles. *The Journal of biological chemistry*, 283(21):14815–25, 2008.

[67] C Holt, P a Timmins, N Errington, and J Leaver. A core-shell model of calcium phosphate nanoclusters stabilized by beta-casein phosphopeptides, derived from sedimentation equilibrium and small-angle X-ray and neutron-scattering measurements. *European journal of biochemistry*, 252(1):73–78, 1998.

[68] Samuel Lenton, Tommy Nylander, Susana C. M. Teixeira, and Carl Holt. A review of the biology of calcium phosphate sequestration with special reference to milk. *Dairy Science & Technology*, 95(1):3–14, 2015.

[69] C. Holt and D.W.L. Hukins. Structural analysis of the environment of calcium ions in crystalline and amorphous calcium phosphates by X-ray absorption spectroscopy and a hypothesis concerning the biological function of the casein micelle. *International Dairy Journal*, 1(3):151–165, 1991.

[70] C. Holt. Structure and Stability of Bovine Casein Micelles. *Advances in Protein Chemistry*, 43:63–151, 1992.

[71] E.S Smyth, R.A Clegg, and C. Holt. A biological perspective on the structure and function of caseins and casein micelles. *International Journal of Dairy Technology*, 57(2):121–126, 2004.

[72] Carl Holt. An equilibrium thermodynamic model of the sequestration of calcium phosphate by casein micelles and its application to the calculation of the partition of salts in milk. *European biophysics journal*, 33(5):421–434, 2004.

[73] Carl Holt, NM Wahlgren, and T Drakenberg. Ability of a beta-casein phosphopeptide to modulate the precipitation of calcium phosphate by forming amorphous dicalcium phosphate nanoclusters. *Biochemical Journal*, 314:1035–1039, 1996.

[74] Elaine M Little and Carl Holt. An equilibrium thermodynamic model of the sequestration of calcium phosphate by casein phosphopeptides. *European biophysics journal*, 33(5):435–47, 2004.

[75] Carl Holt and Lindsay Sawyer. Caseins as Rheomorphic Proteins. *Journal of the Chemical Society, Faraday Transactions*, 89(15):2683–2692, 1993.

[76] Christopher D Syme, Ewan W Blanch, Carl Holt, Ross Jakes, Michel Goedert, Lutz Hecht, and Laurence D Barron. A Raman optical activity study of rheomorphism in caseins, synucleins and tau. New insight into the structure and behaviour of natively unfolded proteins. *European journal of biochemistry*, 269(1):148–156, 2002.

[77] M.L. Groves and W.G. Gordon. The major component of human casein: A protein phosphorylated at different levels. *Archives of Biochemistry and Biophysics*, 140(1):47–51, 1970.

[78] P Ferranti, A Malorni, G Nitti, P Laezza, R Pizzano, L Chianese, and F Addeo. Primary structure of bovine alpha s1-caseins: localization of phosphorylation sites and characterization of genetic variants A, C and D. *The Journal of dairy research*, 62(2):281–96, 1995.

[79] Roger A Clegg and Carl Holt. An E. coli over-expression system for multiply-phosphorylated proteins and its use in a study of calcium phosphate sequestration by novel recombinant phosphopeptides. *Protein expression and purification*, 67(1):23–34, 2009.

[80] D T Denhardt, C A Lopez, E E Rollo, S M Hwang, X R An, and S E Walther. Osteopontin-induced modifications of cellular functions. *Annals of the New York Academy of Sciences*, 760:127–42, 1995.

[81] J. Sodek, B. Ganss, and M.D. McKee. Osteopontin. *Critical Reviews in Oral Biology & Medicine*, 11(3):279–303, 2000.

[82] Patrizia Lanteri, Giovanni Lombardi, Alessandra Colombini, Dalila Grasso, and Giuseppe Banfi. Stability of osteopontin in plasma and serum. *Clinical chemistry and laboratory medicine*, 50(11):1979–84, 2012.

[83] Juan B Grau, Paolo Poggio, Rachana Sainger, William J Vernick, William F Seefried, Emanuela Branchetti, Benjamin C Field, Joseph E Bavaria, Michael A Acker, and Giovanni Ferrari. Analysis of osteopontin levels for the identification of asymptomatic patients with calcific aortic valve disease. *The Annals of thoracic surgery*, 93(1):79–86, 2012.

[84] L W Fisher, D a Torchia, B Fohr, M F Young, and N S Fedarko. Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochemical and biophysical research communications*, 280(2):460–465, 2001.

[85] C. Qin, O. Baba, and W.T. Butler. Post-Translational Modifications of Sibling Proteins and Their Roles in Osteogenesis and Dentinogenesis. *Critical Reviews in Oral Biology & Medicine*, 15(3):126–136, 2004.

[86] Brian Christensen, Mette S Nielsen, Kim F Haselmann, Torben E Petersen, and Esben S Sø rensen. Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. *The Biochemical journal*, 390(Pt 1):285–292, 2005.

[87] Nilana M T Barros, Betty Hoac, Raquel L Neves, William N Addison, Diego M Assis, Monzur Murshed, Adriana K Carmona, and Marc D McKee. Proteolytic processing of osteopontin by PHEX and accumulation of osteopontin fragments in Hyp mouse bone, the murine model of X-linked hypophosphatemia. *Journal of bone and mineral research*, 28(3):688–99, 2013.

[88] Brian Christensen, Christian C Kazanecki, Torben E Petersen, Susan R Rittling, David T Denhardt, and Esben S Sorensen. Cell type-specific post-translational modifications of mouse osteopontin are associated with different adhesive properties. *The Journal of biological chemistry*, 282(27):19463–19472, 2007.

[89] D T Denhardt and M Noda. Osteopontin expression and function: role in bone remodeling. *Journal of cellular biochemistry. Supplement*, 30-31:92–102, 1998.

[90] Hema Rangaswami, Anuradha Bulbule, and Gopal C Kundu. Osteopontin: role in cell signaling and cancer progression. *Trends in cell biology*, 16(2):79–87, 2006.

[91] Min Yang, Anup Ramachandran, Hui-Min Yan, Benjamin L Woolbright, Bryan L Copple, Peter Fickert, Michael Trauner, and Hartmut Jaeschke. Osteopontin is an initial mediator of inflammation and liver injury during obstructive cholestasis after bile duct ligation in mice. *Toxicology letters*, 224(2):186–95, 2014.

[92] Y Katayama, C M House, N Udagawa, J J Kazama, R J McFarland, T J Martin, and D M Findlay. Casein kinase 2 phosphorylation of recombinant rat osteopontin enhances adhesion of osteoclasts but not osteoblasts. *Journal of cellular physiology*, 176(1):179–87, 1998.

[93] J R Hoyer, R A Pietrzyk, H Liu, and P A Whitson. Effects of microgravity on urinary osteopontin. *Journal of the American Society of Nephrology*, 10 Suppl 1:S389–393, 1999.

[94] J R Hoyer, J R Asplin, and L Otvos. Phosphorylated osteopontin peptides suppress crystallization by inhibiting the growth of calcium oxalate crystals. *Kidney international*, 60(1):77–82, 2001.

[95] S Jono, C Peinado, and C M Giachelli. Phosphorylation of osteopontin is required for inhibition of vascular smooth muscle cell calcification. *The Journal of biological chemistry*, 275(26):20197–20203, 2000.

[96] Vincent S Tagliabracci, James L Engel, Jianzhong Wen, Sandra E Wiley, Carolyn a Worby, Lisa N Kinch, Junyu Xiao, Nick V Grishin, and Jack E Dixon. Secreted Kinase Phosphorylates Extracellular Proteins that Regulate Biomineralization. *Science*, 336(6085):1150–1153, 2012.

[97] Hiroyuki O Ishikawa, Aiguo Xu, Eri Ogura, Gerard Manning, and Kenneth D Irvine. The Raine syndrome protein FAM20C is a Golgi kinase that phosphorylates bio-mineralization proteins. *PloS one*, 7(8):e42988, 2012.

[98] S Ashkar, J L Schaffer, E Salih, L C Gerstenfeld, and M J Glimcher. Phosphorylation of osteopontin by Golgi kinases. *Annals of the New York Academy of Sciences*, 760:296–298, 1995.

[99] E S Sorensen, P Ho jrup, and T E Petersen. Posttranslational modifications of bovine osteopontin: identification of twenty-eight phosphorylation and three O-glycosylation sites. *Protein science*, 4(10):2040–2049, 1995.

[100] Eva Kläning, Brian Christensen, Esben S Sorensen, Thomas Vorup-Jensen, and Jan K Jensen. Osteopontin binds multiple calcium ions with high affinity and independently of phosphorylation status. *Bone*, 66:90–5, 2014.

[101] Dennis Kurzbach, Gerald Platzer, Thomas C Schwarz, Morkos A Henen, Robert Konrat, and Dariush Hinderberger. Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry*, 52(31):5167–5175, 2013.

[102] Lajos Kalmar, Daniel Homola, Gabor Varga, and Peter Tompa. Structural disorder in proteins brings order to crystal growth in biomineralization. *Bone*, 51(3):528–34, 2012.

[103] Franck Peysselon, Bin Xue, Vladimir N Uversky, and Sylvie Ricard-Blum. Intrinsic disorder of the extracellular matrix. *Molecular bioSystems*, 7(12):3353–65, 2011.

[104] P E Wright and H J Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, 293(2):321–31, 1999.

[105] Robin van der Lee, Marija Buljan, Benjamin Lang, Robert J Weatheritt, Gary W Daughdrill, A Keith Dunker, Monika Fuxreiter, Julian Gough, Joerg Gsponer, David T Jones, Philip M Kim, Richard W Kriwacki, Christopher J Oldfield, Rohit V Pappu, Peter Tompa, Vladimir N Uversky, Peter E Wright, and M Madan Babu. Classification of intrinsically disordered regions and proteins. *Chemical reviews*, 114(13):6589–631, 2014.

[106] Christopher J Oldfield and A Keith Dunker. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual review of biochemistry*, 83:553–84, 2014.

[107] Matt E Oates, Pedro Romero, Takashi Ishida, Mohamed Ghalwash, Marcin J Mizianty, Bin Xue, Zsuzsanna Dosztányi, Vladimir N Uversky, Zoran Obradovic, Lukasz Kurgan, A Keith Dunker, and Julian Gough. D^2P^2 : $database of disordered protein predictions. Nucleic acids research, 41(Database issue)$: $D508 - -16, 2013$.

[108] Predrag Radivojac, Lilia M Iakoucheva, Christopher J Oldfield, Zoran Obradovic, Vladimir N Uversky, and A Keith Dunker. Intrinsic disorder and functional proteomics. *Biophysical journal*, 92(5):1439–56, 2007.

[109] Celeste J Brown, Audra K Johnson, A Keith Dunker, and Gary W Daughdrill. Evolution and disorder. *Current opinion in structural biology*, 21(3):441–6, 2011.

[110] Celeste J Brown, Sachiko Takayama, Andrew M Campen, Pam Vise, Thomas W Marshall, Christopher J Oldfield, Christopher J Williams, and A Keith Dunker. Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of molecular evolution*, 55(1):104–10, 2002.

[111] Gary W Daughdrill, Pranesh Narayanaswami, Sara H Gilmore, Agniezka Belczyk, and Celeste J Brown. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *Journal of molecular evolution*, 65(3):277–88, 2007.

[112] Harry Amri Moesa, Shunichi Wakabayashi, Kenta Nakai, and Ashwini Patil. Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. *Molecular bioSystems*, 8(12):3262–73, 2012.

[113] S Lise and D T Jones. Sequence patterns associated with disordered regions in proteins. *Proteins*, 58(1):144–50, 2005.

[114] Xin Deng, Jesse Eickholt, and Jianlin Cheng. A comprehensive overview of computational protein disorder prediction methods. *Molecular bioSystems*, 8(1):114–21, 2012.

[115] Zsuzsanna Dosztányi, Veronika Csizmók, Péter Tompa, and István Simon. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of molecular biology*, 347(4):827–39, 2005.

[116] Paul C Whitford, Karissa Y Sanbonmatsu, and José N Onuchic. Biomolecular dynamics: order-disorder transitions and energy landscapes. *Reports on progress in physics*, 75(7):076601, 2012.

[117] Malene Ringkjø bing Jensen, Markus Zweckstetter, Jie-rong Huang, and Martin Blackledge. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chemical reviews*, 114(13):6632–60, 2014.

[118] Konstantin K Turoverov, Irina M Kuznetsova, and Vladimir N Uversky. The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Progress in biophysics and molecular biology*, 102(2-3):73–84, 2010.

[119] Vladimir N Uversky. Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta*, 1834(5):932–951, 2013.

[120] Vikas Pejaver, Wei-Lun Hsu, Fuxiao Xin, A Keith Dunker, Vladimir N Uversky, and Predrag Radivojac. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein science*, 23(8):1077–93, 2014.

[121] Atsushi Kurotani, Alexander A Tokmakov, Yutaka Kuroda, Yasuo Fukami, Kazuo Shinozaki, and Tetsuya Sakurai. Correlations between predicted protein disorder and post-translational modifications in plants. *Bioinformatics*, 30(8):btt762–, 2014.

[122] Lilia M Iakoucheva, Predrag Radivojac, Celeste J Brown, Timothy R O'Connor, Jason G Sikes, Zoran Obradovic, and A Keith Dunker. The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research*, 32(3):1037–49, 2004.

[123] Predrag Radivojac, Vladimir Vacic, Chad Haynes, Ross R Cocklin, Amrita Mohan, Joshua W Heyen, Mark G Goebl, and Lilia M Iakoucheva. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins*, 78(2):365–80, 2010.

[124] Christopher T Walsh, Sylvie Garneau-Tsodikova, and Gregory J Gatto. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie*, 44(45):7342–72, 2005.

[125] Eric S Witze, William M Old, Katheryn A Resing, and Natalie G Ahn. Mapping protein post-translational modifications with mass spectrometry. *Nature methods*, 4(10):798–806, 2007.

[126] Bruce T Seet, Ivan Dikic, Ming-Ming Zhou, and Tony Pawson. Reading protein modifications with interaction domains. *Nature reviews. Molecular cell biology*, 7(7):473–83, 2006.

[127] Lilia M Iakoucheva, Celeste J Brown, J David Lawson, Zoran Obradović, and A Keith Dunker. Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of molecular biology*, 323(3):573–84, 2002.

[128] Arun Mehra, Christopher L Baker, Jennifer J Loros, and Jay C Dunlap. Post-translational modifications in circadian rhythms. *Trends in biochemical sciences*, 34(10):483–90, 2009.

[129] Mi-Kyung Yoon, Diana M Mitrea, Li Ou, and Richard W Kriwacki. Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochemical Society transactions*, 40(5):981–8, 2012.

[130] Hongbo Xie, Slobodan Vucetic, Lilia M Iakoucheva, Christopher J Oldfield, A Keith Dunker, Vladimir N Uversky, and Zoran Obradovic. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *Journal of proteome research*, 6(5):1882–98, 2007.

[131] Wei-Lun Hsu, Christopher J Oldfield, Bin Xue, Jingwei Meng, Fei Huang, Pedro Romero, Vladimir N Uversky, and A Keith Dunker. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein science*, 22(3):258–73, 2013.

[132] Zsuzsanna Dosztányi, Jake Chen, A Keith Dunker, István Simon, and Peter Tompa. Disorder and sequence repeats in hub proteins and their implications for network evolution. *Journal of proteome research*, 5(11):2985–95, 2006.

[133] Vladimir N Uversky. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chemical Society reviews*, 40(3):1623–34, 2011.

[134] Vladimir N Uversky. Intrinsic disorder-based protein interactions and their modulators. *Current pharmaceutical design*, 19(23):4191–213, 2013.

[135] C J Tsai, S Kumar, B Ma, and R Nussinov. Folding funnels, binding funnels, and protein function. *Protein science : a publication of the Protein Society*, 8(6):1181–90, 1999.

[136] C J Tsai, B Ma, Y Y Sham, S Kumar, and R Nussinov. Structured disorder and conformational selection. *Proteins*, 44(4):418–27, 2001.

[137] David F Lowry, Amber Stancik, Ranjay Mann Shrestha, and Gary W Daughdrill. Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. *Proteins*, 71(2):587–98, 2008.

[138] Jennifer M Bui and Jörg Gsponer. Phosphorylation of an intrinsically disordered segment in Ets1 shifts conformational sampling toward binding-competent substates. *Structure*, 22(8):1196–203, 2014.

[139] Jakob Dogan, Stefano Gianni, and Per Jemth. The binding mechanisms of intrinsically disordered proteins. *Physical chemistry chemical physics*, 16(14):6323–6331, 2014.

[140] Peter E Wright and H Jane Dyson. Linking folding and binding. *Current opinion in structural biology*, 19(1):31–8, 2009.

[141] Christopher J Oldfield, Yugong Cheng, Marc S Cortese, Pedro Romero, Vladimir N Uversky, and A Keith Dunker. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, 44(37):12454–70, 2005.

[142] J. M. Rogers, V. Oleinikovas, S. L. Shammas, C. T. Wong, D. De Sancho, C. M. Baker, and J. Clarke. Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proceedings of the National Academy of Sciences*, 111(43):15420–15425, 2014.

[143] Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in proteinâĂŞprotein interactions. *Trends in Biochemical Sciences*, 33(1):2–8, 2008.

[144] A Guinier and G Fournet. *Small-angle scattering of X-rays*. Wiley, New York, 1955.

[145] Efstratios Mylonas and Dmitri I. Svergun. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography*, 40(s1):s245–s249, 2007.

[146] Pau Bernadó, Efstratios Mylonas, Maxim V Petoukhov, Martin Blackledge, and Dmitri I Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, 129(17):5656–64, 2007.

[147] A. V. Semenyuk and D. I. Svergun. GNOM, a program package for small-angle scattering data processing. *Journal of Applied Crystallography*, 24(5):537–540, 1991.