The self-assembly of proteins: Probing anisotropic protein-protein interactions using phase diagrams

A thesis submitted to the National University of Ireland in fulfilment of the requirements for the degree of

Doctor of Philosophy

By

Susan James, M.Sc.

Department of Chemistry

Maynooth University

October 2015

Research Supervisor: Dr. Jennifer McManus

Head of Department: Dr. John Stephens

Declaration

I hereby declare that this thesis leading to the award of the Degree of Doctor of Philosophy has not been submitted before, in whole or in part, to this or any other University for any degree and is except where otherwise cited and acknowledged within the text of my work.

Signed:

Date:

Acknowledgements

The journey I began in Dr. McManus's lab in the Chemistry department at Maynooth University six long years ago, it's quite hard to believe that it's all coming to an end. On this occasion, the first person I would like to thank is my supervisor, Dr. Jennifer McManus, her guidance and supervision has been instrumental in the successful completion of my project work and thesis report. I can never thank her enough for being an inspiration and for all the support and motivation she has offered me for the last few years. I would also like to thank Prof. John Lowry who had been kind enough to help me find a position within the department.

Special thanks to the Irish Research Council (IRC) whose funding helped me survive in Maynooth and to the Science Foundation Ireland for financially supporting the project work. I am also thankful to Prof. Amir Khan in the School of Biochemistry and Immunology, Trinity College Dublin for his patience and kind help with the crystal structures. To all the staff members within the chemistry department especially the administrative staff (Niamh, Donna and Carol), the technical staff (Ken, Noel, Ria, Barbara, Ollie, Anne, Orla, Walter) and all other staff members, I extend my sincerest gratitude.

The chemistry department has been my second home for the past few years and it would not have been a memorable experience if not for my big bunch of lab buddies. Ruthie, I will cherish the time spent with you forever. Ula (Urszula Migas), I will certainly miss the fights. You made our Portugal trip even more eventful. Alice, if not for you I would not have been convinced enough that 'it' turns rusty!! Michelle, my partner-in-crime and the first person I knew from the lab. I have the Rome trip and two desserts to remember you all forever!! Mark, thanks for all your patience and help. Fiona, it was nice knowing you, better late than never!! I wish you all the very best of luck with your future endeavours. I would also like to thank all the post-grads and post-docs in the past and present, especially to name a few, Aidy, you have been a great companion. To Gama, 'merci' for all the little French you have taught me and for all the pleasant drives back and forth to Clonee and 'shukriya' Muhib for all those corridor chats and paranthas.

Fondly remembering all my friends back home my Chennai group (Sruthi, Vidya, Guhan and Vedha), Bengal bunch (Urmi and Tamanna), Maynooth lot (Viral, Deepak,

Stephy, Amrita, Nasrin, Amala and Paula) and a few others specially Lovina, Om, Vineeth, Deepthi, Abhishek and Praveen who have supported and motivated me through the entire journey. Last but not the least heartfelt thanks and appreciation towards Mathew Abraham and Laly Mathew who have been kind enough to adopt me for the last few months, I couldn't have asked for a better home and family to be with.

The backbone of my existence, my family, I am blessed to have you all in my life. My father and mother, who have been my greatest strength, I owe everything to both of you. My husband, Febe, I can't thank you enough for putting up with me since the 28th of December, 2008 and for being there for our little man, Dhruv, so glad that I can have the two of you back soon. We, as a family have all the memories from Maynooth in our hearts for the rest of our lives. Heartfelt thanks to my in-laws who have been kind enough to support me for all the decisions that I have taken in life. Finally a word of thanks to the entire family my aunt Animol Ivan, Shobi, Elizabeth, Shery, Sobha, Vivek and baby Saira.

Abstract

Self-assembly is central to the formation of biological structures and has generated a number of synthetic structures, such as biomaterials and photonic devices with important physical properties. Self assembly is of fundamental importance to soft matter science and synthetic biology. For most colloidal and nanoscale particles, the self-assembly properties are defined by the isotropic interaction energy between the particles. However for biological macromolecules such as proteins, anisotropic interaction energies and multiple degrees of freedom dominate the behaviour, making predictability of solution behaviour and crystallization extremely complex. To probe the impact of anisotropic interactions on the self-assembly of proteins, we have created three double mutants of human γD-crystallin (an eye lens protein) for which the phase diagrams for the single mutants are known. In effect, these double mutants establish a competition between the influences of each individual amino acid substitution, each of which are known to significantly influence the phase behaviour of the protein. Hence, the degree to which the molecular anisotropy of the protein contributes to the behaviour of the double mutant can be robustly examined by measuring phase diagrams for the double mutant proteins. Protein phase diagrams establish the solution conditions in which protein condensed phases such as the dense liquid phase due to liquid-liquid phase separation, protein crystals, protein gels and physical states such as protein aggregates occur.

In a series of double mutants, two distinct behaviours associated with each of the two single amino acids substitutions were incorporated into the same protein (1) propensity to crystallize and (2) inverted solubility. Each single mutant differs in either the position of equilibrium phase boundary or the kinetics associated with either aggregate or crystal formation.

The first double mutant was the P23VR36S variant of HGD. The phase diagram for this double mutant revealed that the protein formed crystals with normal solubility at very low protein concentrations (less than 1mg/ml). These crystals had an equilibrium phase boundary identical to the R36S single mutant. Given that the inverted solubility associated with the P23V substitution only occurs for aggregates formed at significantly higher concentrations (greater than 10mg/ml), the behaviour of the double mutant in this case is consistent with the behaviour expected from these substitutions.

The second double mutant, P23TR36S, again incorporating an amino acid substitution associated with crystallization (R36S) and another associated with inverted solubility (P23T) had a more dramatic and complex phase behaviour. It formed two different crystal types with different morphologies and very significantly, each crystal type displayed different solubility behaviour. One crystal had normal solubility (rod-shaped crystals) and the second crystal had inverted solubility (rhombic-shaped crystals). This type of polymorphism in proteins is a rare occurrence, where the only solution condition change required to interchange between the two crystal forms is temperature. Since the P23T single mutant occurs at significantly lower concentrations than for the P23V single mutant, it was possible to obtain a mutant protein which displayed the two single mutant behaviours within the same protein. The phase diagram also revealed that the phase boundary lines for the two different crystal types were qualitatively consistent with the phase lines for the individual single mutants. Furthermore, the point at which the solubility lines for the two single mutant proteins overlap, co-existence of the two crystal forms is observed. High resolution X-ray crystallography data (2.2Å and 1.2 Å respectively) confirmed that the two crystal forms are polymorphs.

Finally, the third double mutant, P23VR58H, exhibits condensed phases associated with both single mutant proteins. The R58H mutant is associated with increased propensity to crystallize, but the kinetics of crystal growth is much slower than for the R36S mutant. Therefore, it was possible to form large aggregates (like those for the P23V single mutant) before crystallization occurred. When crystallization did occur, it was on the surface of the aggregates and not in the bulk solution. The equilibrium solubility line measured for this protein represents the coexistence between protein monomers and the protein condensed phases (crystals + aggregates). This suggests, contrary to the accepted view, that the protein aggregates are as thermodynamically stable as the protein crystal. Kinetic studies on the growth of the protein aggregates and crystals show that the growth of the aggregates occurs by first order growth kinetics and that crystallization occurs independently of the initial aggregate growth. The formation of both condensed phases is independent, but kinetically controlled.

These observations, using three double mutants of HGD are striking and are the first time that this approach has been used to examine the role that molecular anisotropy of the protein surface has on the formation of condensed protein phases. While HGD is an effective model to examine these interactions, the work has important implications for protein crystallization efforts in general and for industrial processes where high

concentration protein solutions are required such as in the food and biopharmaceutical industries, since anisotropic protein-protein interactions are particularly important in understanding these systems.

List of abbreviations

μ – Electrophoretic mobility
μ^c – Chemical potential
$\Delta\mu_{trans}$ – Transfer chemical potential
η_0 – Solvent viscosity
φ – Volume fraction
Aβ – β-amyloid peptide
APP – Amyloid precursor protein
B_{22}/B_2 – Osmotic second virial coefficient
cDNA – Complementary DNA
CD – Circular dichroism
D_0 – Diffusion coefficient
DTT – Dithiothreitol
dsDNA – Double-stranded DNA
DNase I – Deoxyribonuclease I
EDTA – Ethylene diamine tetraacetic acid
F – Mobile phase flow
F^{att} – Attractive forces
U^{att} – Attractive interaction potential energy
U^{rep} – Repulsive interaction potential energy
U– Total interaction potential energy
ΔG^0_{cryst} - Gibbs free energy of crystallization
h – Planck's constant
HGD – Human γD-crystallin
ΔH^0_{cryst} - Enthalpy of crystallization
IPTG – Isopropyl-β-D-thiogalactopyranoside
K – Retention factor
K_{av} – Partition coefficient
K_D – Net interaction potential
K_B – Boltzmann's constant
LB agar/broth – Luria-Bertani agar/broth
$MgSO_4.7H_2O$ – Magnesium sulphate heptahydrate
MPD – 2-methyl-2,4-pentanediol
$NaH_2PO_4.2H_2O$ – Sodium dihydrogen phosphate dihydrate
$Na_2HPO_4.7H_2O$ – Sodium phosphate dibasic heptahydrate
NaN_3 – Sodium azide
OD – Optical density
PCR – Polymerase chain reaction
PEG – Polyethylene Glycol
pI – Iso-electric point
QELS – Quasi-elastic light scattering
R_h – Hydrodynamic radius
S.O.C – Super optimal broth with catabolic repressor
SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis
ssDNA – Single-stranded DNA
ΔS^0_{cryst} - Entropy of crystallization

$\Delta S^{0}_{protein}$ - Protein entropy changes

$\Delta S^{0}_{solvent}$ - Solvent entropy changes

SE-HPLC – Size-exclusion high performance liquid chromatography

TEMED - Tetramethylethylenediamene

t_R – Retention time

t_0 – Void time

T_c – Critical temperature

T_{cloud} – Clouding temperature

T_{clear} – Clearing temperature

T_{ph} – Phase separation temperature

V_0 – Void volume

V_t – Total volume

V_e – Elution volume

V_R – Retention volume

V_m – Matthew's coefficient

ΔW – Work done

Table of Contents

Chapter 1

Introduction

1.1Self-assembly in soft matter systems

Self-assembly occurs when individual building blocks spontaneously organize into ordered structures, driven mostly by thermodynamic forces (Jones, 2002). The concept of self-assembly originated in chemistry and the motivation was derived from the central importance of self-assembly in biology which includes synthesis of protein molecules from linear polypeptide strands that spontaneously fold into functional three-dimensional structures (Voet & Voet, 2011); spontaneous organization of lipid molecules into bilayer membranes (Jones & Chapman, 1995); formation of viral capsids (Casjens & King, 1975) and the formation of whole organisms from living cells by division, development and organization. Self-assembly has also found wide-spread use in chemistry pertaining to crystallization, formation of membrane vesicles, surface adsorption and supramolecular chemistry (Lehn, 2004).

1.1.1 Colloids as simple models for soft matter systems

Colloids are sub-microscopic particles with dimensions ranging from tens of nanometers to a few microns (figure 1.1). When in the nanometer range, these particles are below the limits of detection of a simple optical microscope, and direct imaging and measurement of the size of nanoscale colloidal particles became possible only with the advent of the electron microscope (Everett, 1988).

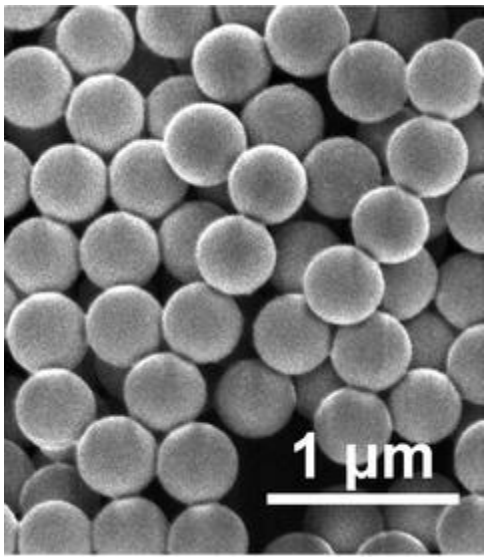

Figure 1.1: *Pictorial representation of silica colloidal particles in a dispersion medium, taken from (Shi et al., 2013).*

The size range is not rigid and has exceptions such as in emulsions and slurries for which particles are larger in size. In systems such as clay and thin films, the colloidal

range is restricted to only one dimension (Everett, 1988). Due to the high surface to volume ratio, interfacial surface chemistry plays an important role in determining the behaviour of colloidal particles and the overall thermodynamic properties of the system, particularly its free energy.

1.1.2 Interaction potential: Intermolecular forces governing colloidal behaviour

To understand the interactions between colloidal particles, it is important to understand the origin and nature of the intermolecular forces that govern colloidal interactions. The existence of attractive forces between non-polar molecules has been traced back to the classical work of van der Waals (1873), but their origin was only understood when London (1930) quantified attractive forces between fluctuating dipoles using quantum mechanics (Everett, 1988).

These attractive forces increase in magnitude as the molecules approach each another and are related to one other as per equation 1.1,

$$F^{att} = \frac{-A}{r^7} \qquad 1.1$$

where the attractive force (F^{att}) is inversely proportional to the seventh power of separation. The work done in separating a pair of molecules from a distance d to infinity is given by equation 1.2,

$$\Delta W = -\int_d^\infty F^{att} dr = A\int_d^\infty \frac{1}{r^7} dr = \frac{A}{6d^6} = \frac{A'}{d^6} \qquad 1.2$$

If the energy at infinite separation is zero, then the attractive potential energy (U^{att}) between the pair of molecules at a separation distance d is given by equation 1.3 (A' is the attractive van der Waals constant),

$$U^{att} = -\Delta W = -\frac{A'}{d^6} \qquad 1.3$$

This attractive interaction potential is schematically shown as (a) in figure 1.2,

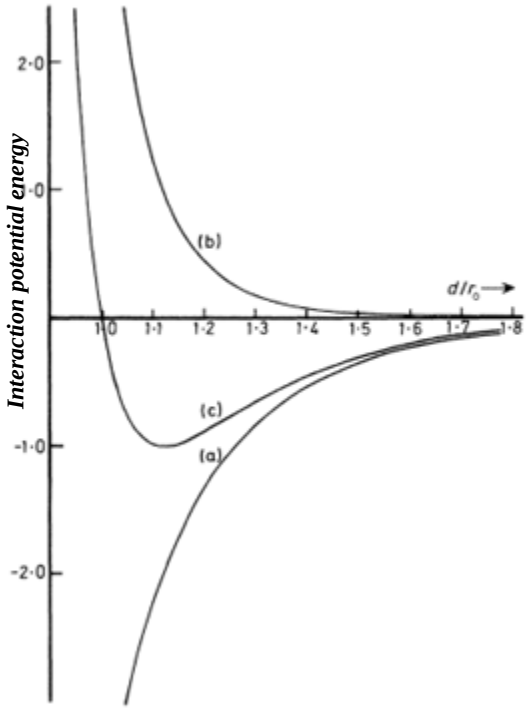

Figure 1.2: *Interaction potential energy between two molecules at a distance d (a) van der Waals attraction, (b) steric repulsion and (c) net interaction potential, taken from (Everett, 1988).*

As a pair of molecules approaches each other, the energies of attraction become increasingly negative. However, as the pair of molecules approach to very close distances, their electron clouds in non-bonding orbitals begin to interact which gives rise to a repulsive force that becomes increasingly infinite as the electron clouds interpenetrate.

Thus, the repulsive potential generated from the intermolecular potential between two interacting pair of molecules is represented by equation 1.4 (B′ is the repulsive Born constant),

$$U^{rep} = \frac{B'}{d^{12}} \qquad 1.4$$

which is schematically shown as (b) in figure 1.2.

The total interaction potential energy between a pair of molecules is the sum of the contributions from the attractive and the repulsive forces represented by equation 1.5 and schematically shown as (c) in figure 1.2.

$$U = U^{rep} + U^{att} = \left(\frac{B'}{d^{12}}\right) - \left(\frac{A'}{d^{6}}\right) \qquad 1.5$$

The above representation is usually referred to as the Lennard-Jones potential. Instead of expressing the above equation in terms of parameters A' and B', it is convenient to express the free energy interaction potential in terms of the depth of the potential well ε and the distance at which the interaction potential is zero, r_0 (equation 1.6);

$$U = 4\varepsilon\left[\left(\frac{r_0}{d}\right)^{12} - \left(\frac{r_0}{d}\right)^{6}\right] \qquad 1.6$$

1.1.3 Interparticle forces in colloidal interactions

1.1.3.1 Electrostatic forces: the electric double layer

One of the most significant interparticle forces that exist between colloidal particles is the electrostatic force that arises due to electrically charged surfaces. Surfaces become electrically charged by a variety of mechanisms which include ionization of surface groups, differential solution of ions from the surface of a sparingly soluble crystal, isomorphous substitution, charged crystal surfaces and specific ion adsorption (Everett, 1988). In order to examine the effect of surface charges on the properties of colloids, the fundamental electrostatic theory (Coulomb's law) is applied.

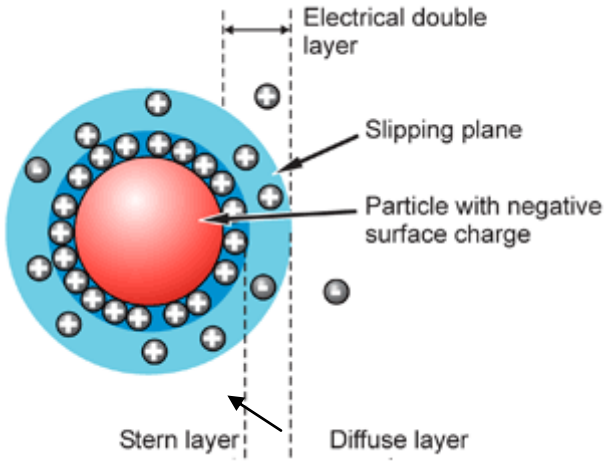

Figure 1.3: *Pictorial depiction of electrical double layer, taken from (nanoComposix Europe 2015).*

A charge cloud or ionic atmosphere is developed around the surface of a charged colloidal particle which is generally referred to as an electric double layer (figure 1.3). The electric double layer consists of charges (positive or negative) distributed on the surface of colloidal particles which is balanced by the total charge in the double layer comprising an excess of oppositely charged ions (counter-ions). The interaction

between the associated double layers surrounding the colloidal particles influences the interparticle interaction between colloidal particles *i.e.* electrostatic repulsion is introduced between two similarly charged colloidal particles when their ionic clouds approach each other. An important consequence of the electric double layer is charge screening or shielding, where long-range electrostatic repulsive forces are screened, resulting in a decrease in the repulsive forces and an increase in the attractive forces between charged molecules (James & McManus, 2012).

1.1.3.2 Steric repulsion: the adsorbed or anchored layers

Adsorbed layers can influence the van der Waals attractive forces and give rise to repulsion between particles. If an adsorbed layer (usually a polymer) is rigid and compact, the adsorbed layer has a tendency to prevent the centres of the colliding particles from coming closer than 2 (a + δ) where *a* is the radius of the particle and *δ* is the thickness of the adsorbed layer. These particles with an effective collision diameter 2 (a + δ) may be regarded as hard spheres. Two situations may arise due to the collision of particles. If the properties of the surface layer are closely similar to the dispersion medium (the medium in which the colloidal particle is dispersed), the attraction between the colloidal particles for a given centre-to-centre distance is unaffected. Since the centre-to-centre distance cannot fall below 2 (a + δ), the attractive potential at contact is lower than for an uncoated particle (δ = 0) (figure 1.4 (a)). If the properties of the surface layer are closely similar to the core material of the colloidal particle, the coated particle will have a greater attractive potential than for the uncoated particle (figure 1.4 (b)). The density of the adsorbed polymer layer also affects the interaction potential of the colloidal particle. The adsorbed polymer layer is not dense enough for the colloidal particle to behave as a hard sphere. These polymer chains of the adsorbed layer extend out into the dispersion medium and depending on its chemical similarity; it begins interacting with the dispersion medium resulting in an open surface structure. On collision with other colloidal particles dispersed in the medium, the polymer chains interpenetrate which will contribute to the entropic repulsion term of the interaction potential. In order to stabilize the colloidal system, this approach is exploited by preparing colloidal particles whose adsorbed surface layer is chemically similar to the dispersion medium and different to the particle itself (Everett, 1988).

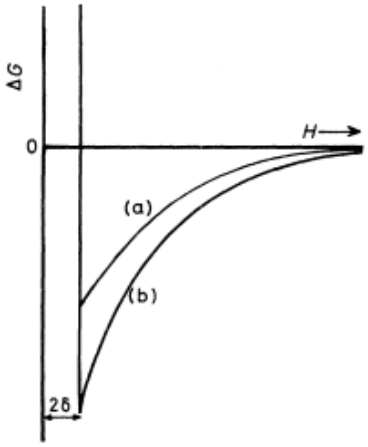

Figure 1.4: *Interaction potential curves (a) when the surface layer has closely similar properties to the dispersion medium and (b) when the surface layer has closely similar properties to the core material, taken from (Everett, 1988).*

1.1.4 Phase diagrams for simple and complex colloidal systems

The simplest colloidal systems are hard spheres which neither attract nor repel at distances greater than the particle diameter. Their characteristic feature is that they cannot interpenetrate, due to which two hard spheres with a radius *a* when held at a centre separation distance *r* produces infinite repulsion if ($r \leq 2a$) *i.e.* when their centre separation distance is less than twice the radius *2a* and zero repulsion when ($r \geq 2a$) (Chaikin & Lubensky, 2000). The only externally controlled parameter affecting the interaction potential is the overall density of spheres which is expressed in terms of its particle volume fraction ϕ.

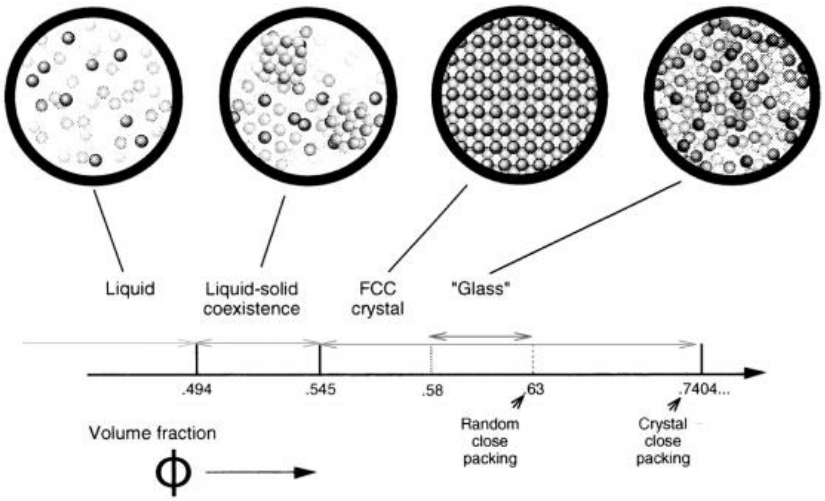

Figure 1.5: *Equilibrium phase diagram for uniformly sized hard spheres, taken from (Cheng et al., 2001).*

When the colloidal suspension consists of large and small particles, the pressure from the small particles produces an attraction between the large particles. This is referred to as depletion force and is the possible cause of phase separation for colloidal particles immersed in a liquid of non-adsorbing polymers (Mao *et al.*, 1995). The equilibrium phase behaviour for hard spheres is one-dimensional, controlled by the volume fraction of the system. For ($\phi \leq 0.49$), the hard spheres exist in a fluid state; for volume fractions in the range ($0.54 \leq \phi \leq 0.74$), the hard spheres enter a face-centred cubic (FCC) crystal state; for volume fractions in the range ($0.49 \leq \phi \leq 0.54$), the hard sphere system partitions into two co-existing phases, one a crystal state and the other a fluid state constituting an ergodic system; and for ($\phi \approx 0.58$), the hard spheres may avoid nucleation to form a kinetically arrested disordered state *i.e.* structurally similar to a liquid but resembles an amorphous glass (figure 1.5) (Sciortino, 2002, Pusey & Van Megen, 1986). The phase diagram obtained for hard spheres modelled using attractive Yukawa potentials and square-well potentials are applied to protein solutions using Mayer and McMillan's treatment of solutions as gaseous suspensions (McQuarrie, 1976). Once this approximation is made, the gas phase is identified with the uniformly dispersed protein liquid phase, the gas-liquid coexistence curve correlates with the liquid-liquid binodal and the sublimation line corresponds to the liquidus or solubility line. When attractive interactions are introduced in hard sphere systems, the phase diagram becomes rich and diverse giving rise to distinct liquid, gas phases (gas phase is effectively a dilute solution, not an actual gas phase) and a liquid-gas critical point (Anderson & Lekkerkerker, 2002). Purely attractive interparticle attractions introduce new parameters such as strength and range of the attraction giving rise to two-dimensional phase diagrams. The first temperature-density phase diagrams for colloidal particles with a short-ranged attraction used hard spheres with attractive Yukawa potentials based on Monte Carlo simulations (Hagen & Frenkel, 1994).

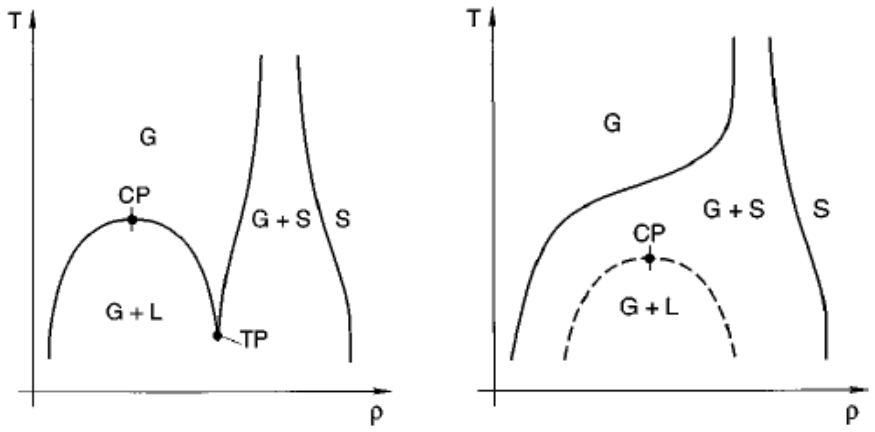

Figure 1.6: *Standard temperature-density phase diagram for hard spheres (diameter σ) with attractive Yukawa potentials (range δ) (left) and for hard spheres (range of interaction $\delta \leq \sigma/7$) (right), taken from (Muschol & Rosenberger, 1997).*

As shown in figure 1.6 (left), gas (G) is the most stable low-density phase at high temperatures. When the temperature is lowered, phase separation into coexisting gas and liquid (G+L) phases occur. As the density of the hard spheres is increased, the gas-liquid (G+L) and the gas-solid (G+S) coexistence curves intersect at the triple point (TP). When the density of the hard spheres reaches a maximum, solidification occurs. This standard phase diagram of hard spheres with attractive interactions is altered when the range of the interaction δ is modified to about one-seventh of the particle hard core diameter σ ($\delta \leq \sigma/7$) and is analogous to the phase diagram for simple fluids (figure 1.6 (right)). On reducing the range of interactions, the triple point vanishes and the gas-liquid coexistence curve shifts to a metastable region beneath the gas-solid coexistence curve as shown in figure 1.6. A similar phase diagram was obtained when hard spheres were modelled using attractive square-well potentials for $\delta < 0.25\sigma$ (Asherie *et al.*, 1996).

1.1.5 Protein phase diagrams

The thermodynamic forces driving the transition of a protein from one phase to another are a combination of processes occurring between protein-protein and protein-solvent molecules (Dumetz, Chockla, *et al.*, 2008). A protein phase diagram is a map which represents the state of a protein as a function of temperature, concentration and various solution conditions such as pH, ionic strength, salt concentration (Asherie, 2004).

The thermodynamics of phase transitions in protein solutions are in many ways closer to the thermodynamics of phase transitions in colloidal solutions (Adams *et al.*, 1998, Hagen & Frenkel, 1994, Asherie *et al.*, 1996, Pusey & Van Megen, 1986, Van Megen & Underwood, 1993) than to those of small-molecule inorganic solutions.

A globular protein is similar to a colloidal particle in many ways. There has been a great deal of experimental work done, measuring phase diagrams for colloidal particles and many of these phase transitions are also observed for globular protein solutions (Asherie, 2004, Muschol & Rosenberger, 1997). A certain number of proteins, including the gamma crystallins and lysozyme have been described as hard spheres with short-range, attractive interactions (Rosenbaum *et al.*, 1996). Proteins such as α-crystallins behave like hard spheres with electrostatic repulsions and weak attractive interactions (Xia *et al.*, 1994). Early work in this area used simple isotropic models to describe proteins as spherical particles with short-range attractive interactions, which could explain certain features of protein phase diagrams such as liquid-liquid phase separation of globular proteins with attractive interactions (Lomakin, Asherie, *et al.*, 1996, Asherie *et al.*, 1996, Rosenbaum *et al.*, 1996, Poon, 1997, Rosenbaum & Zukoski, 1996). These isotropic models work well for proteins in certain contexts, for example, the changes in the phase behaviour of lysozyme with increasing salt concentration (Muschol & Rosenberger, 1997). However they do not quantitatively reproduce many of the features of protein phase diagrams, such as the width of the binodal line (liquid-liquid co-existence line) (Lomakin *et al.*, 1999). Aeolotopic models with nonuniform, short-range interactions, have proved more useful in accurately describing protein phase behaviour (Lomakin *et al.*, 1999) and many newer aeolotopic or "patchy" colloidal models have been developed since (Sear, 1999, Romano *et al.*, 2010, Dorsaz *et al.*, 2012).

Taking into account all the experimental observations and theoretical predictions, a generic phase diagram for globular proteins has been postulated explaining the basic features of protein solutions. As displayed in figure 1.7, three zones have been identified; zone I is identified as the region for protein crystal growth and the shaded region in zone I represents pregelation clustering that can interfere with and interrupt protein crystal growth; zone II represents the metastable region of the phase diagram where liquid-liquid phase separation occurs; and zone III represents the gelation region. The gelation zone corresponds to a dynamically arrested state in which the solution forms a non-equilibrium state such as a gel or glass.

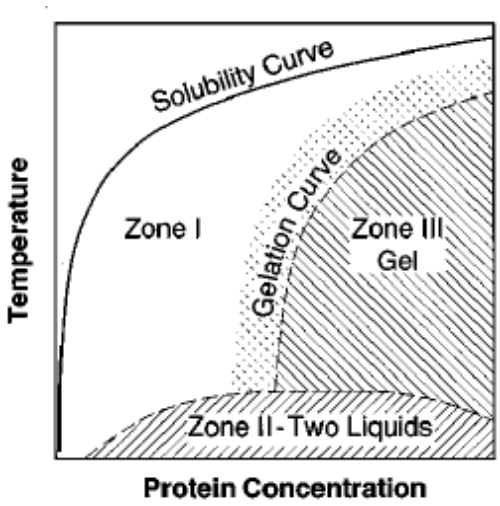

Figure 1.7: *Generic phase diagram for proteins with normal solubility, taken from (Muschol & Rosenberger, 1997).*

The phase diagram shown in figure 1.7 applies to proteins with normal solubility. However, for proteins with inverted solubility the phase diagram is altered and is as shown in figure 1.8.

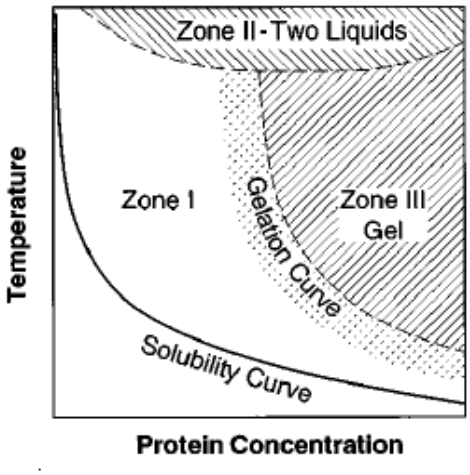

Figure 1.8: *Generic phase diagram for proteins with inverted solubility, taken from (Muschol & Rosenberger, 1997).*

To summarize, the location of phase boundaries of various protein solutions depend on the solution conditions in which the protein is dissolved including pH, ionic strength, physical conditions such as temperatures and the specific chemistry on the surface of proteins. A typical protein phase diagram for proteins interacting via short range attraction consists of i) a liquidus or solubility line denoting the concentration of the monomeric protein solution in equilibrium with the condensed crystal/aggregate phase. ii) a liquid-liquid coexistence line (binodal) metastable to the liquidus or solubility line where the protein solution separates into two co-existing liquid phases of unequal

protein concentrations by temperature quenching the protein solution, one a protein rich phase and the other a protein poor phase. The different phases appear in a protein solution due to concentration fluctuations at temperatures less that the critical temperature for phase separation. iii) a spinodal line separating the biphasic region to a metastable and an unstable region, the spinodal decomposition develops due to decay in concentration fluctuations when the system is quenched into the unstable region and phase separation occurs spontaneously without an activation energy barrier as shown in figure 1.9 (Berland *et al.*, 1992, Broide *et al.*, 1991, Grigsby *et al.*, 2001, Asherie *et al.*, 1996).

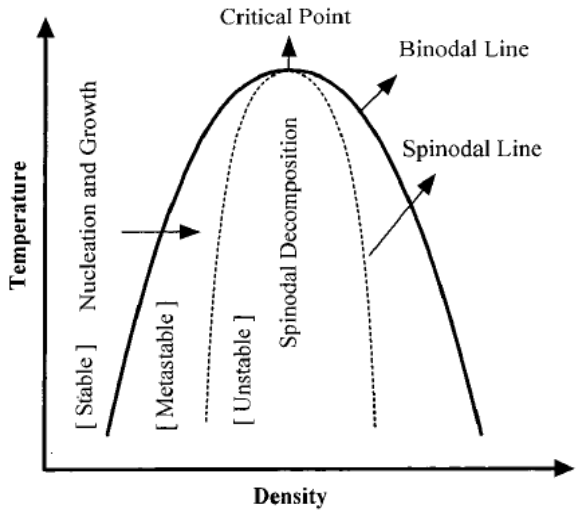

Figure 1.9: *Hypothetical protein phase diagram in the temperature-density plane showing the different phases, taken from (Muschol & Rosenberger, 1995).*

The liquid-liquid co-existence line (binodal) is metastable with respect to the liquidus/solubility line due to the characteristic ranges of interaction between protein molecules in the metastable region of the phase diagram. In this region of the phase diagram, the ranges of all proteins interactions are less than one quarter of the protein diameter in solution (Lekkerkerker *et al.*, 1992, Hagen & Frenkel, 1994, Rascón *et al.*, 1995, Daanoun *et al.*, 1994, Asherie *et al.*, 1996, Rosenbaum *et al.*, 1996).Globular proteins, in general range in diameter from 3-15nm, and as such their inter-particle interactions are determined hydration and hydrophobic interactions which extend to a few water molecular diameters, roughly 10 Å (Leckband & Israelachvili, 2001, Paliwal *et al.*, 2005, Kuehner *et al.*, 1997), ionic or other bridges that have an even shorter range of a few angstroms (McPherson & Cudney, 2006, Dumetz *et al.*, 2007), van der Waals interactions, which extend to about the same distances as ionic bridges (Velev *et*

al., 1998, Song & Zhao, 2004) and the electrostatic range defined by the Debye length depends on the ionic strength of the protein solutions (Muschol & Rosenberger, 1995, Coen *et al.*, 1996). This metastability has been observed in various protein solutions (Berland *et al.*, 1992, Thomson *et al.*, 1987, Broide *et al.*, 1991, Broide *et al.*, 1996, Muschol & Rosenberger, 1997) and in colloidal solutions (Verduin & Dhont, 1995, Ilett *et al.*, 1995, Gast & Russel, 1998), but not in simple fluids where the interaction is long-rage (Zemansky & Dittman, 1997).

During liquid-liquid phase separation in protein solutions, formation of the dense liquid phase is characterised by two phase lines, the co-existence line and the spinodal line. The liquid-liquid phase separation (binodal) is experimentally observed in protein solutions during the formation of dense liquid droplets (clouding) and on changing the temperature conditions these droplets then undergo dissolution into the protein solution (de-clouding) (Muschol & Rosenberger, 1997, Galkin & Vekilov, 2000) whereas the spinodal line is observed when the protein molecules are dynamically arrested and the dense liquid phase forms without the presence of a nucleation barrier (Thomson *et al.*, 1987, Shah *et al.*, 2004). There is a specific point on the phase diagram for protein solutions where the binodal and the spinodal meet, also referred to as the critical point for liquid-liquid phase separation. The critical temperature (T_c) for a protein solution is also defined as the highest temperature at which the two liquid phases co-exist.

If a particular protein solution of a certain composition forms more than one solid phase, each solid phase will exhibit a different solubility line. Lysozyme found in chicken egg white forms two different crystal morphologies; tetragonal and rhombohedral crystals both exhibiting normal solubility with a shift in the liquidus lines depending on the crystal type formed as shown in figure 1.10 (Vekilov, 2012). This indicates that solubility is a property of the solution-crystal equilibrium and not a property of the protein by itself. Protein solid phases can also exhibit two different types of solubility; normal solubility where the concentration of solution in equilibrium with the crystal phase decrease as the temperature is lowered favoring formation and growth of the crystal phase as in lysozyme and inverted solubility (retrograde solubility) where higher temperatures favor liquid-liquid phase separation and protein crystallization as in hemoglobin C (Feeling-Taylor *et al.*, 1999, Vekilov, Feeling-Taylor, Petsev, *et al.*, 2002).

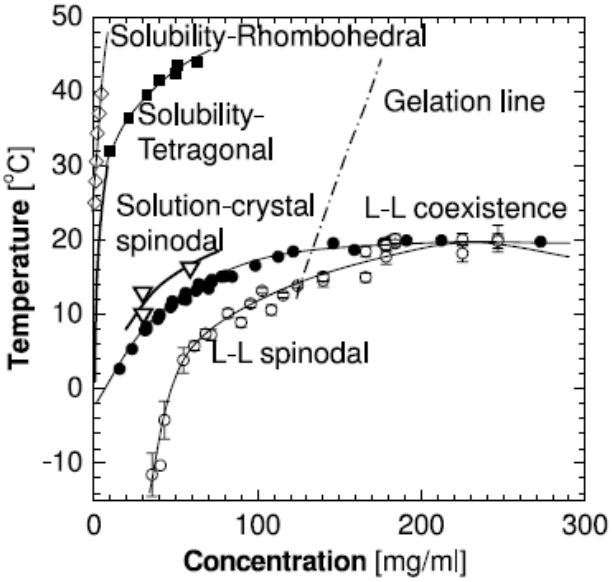

Figure 1.10: *Phase diagram for lysozyme in 0.05M sodium acetate (pH 4.5 and 4% NaCl) showing the liquid-liquid co-existence line (binodal), liquid-liquid spinodal line, liquidus or solubility lines for the different crystal types and the gelation line (taken from (Vekilov, 2012)).*

Besides liquid-liquid phase separation and protein crystallization, protein solutions at higher concentrations form protein gels at temperatures above the critical point for liquid-liquid phase separation (Muschol & Rosenberger, 1997, Galkin & Vekilov, 2000) as indicted by the gelation line in figure 1.10. The molecular mechanism of protein gel formation is not completely understood as yet. Some attribute it to additional long-rage forces (Noro *et al.*, 1999, Sear, 1999, Kulkarni *et al.*, 2003), others to nonuniform anisotropic short-range interactions (Fu *et al.*, 2003). A recent hypothesis for protein gelation has been proposed that relates it to the formation of weak, limited-lifetime networks of protein molecules (Pan *et al.*, 2009).

1.1.5.1 Phase diagram approach for protein crystallization

Protein crystallization is a phase transition phenomena that can be mapped onto a phase diagram (figure 1.11) (Chayen & Saridakis, 2008). In a typical protein crystallization phase diagram, there are four distinct regions: an undersaturated zone where proteins are fully dissolved and do not crystallize; a metastable zone of low supersaturation where protein crystals are stable and they grow, but no further nucleation occurs; a nucleation zone of moderate supersaturation where spontaneous nucleation will occur and a precipitation zone of very high supersaturation where the proteins will precipitate (Chayen *et al.*, 1996, Chayen, 2005, Ducruix & Giegé, 1992).

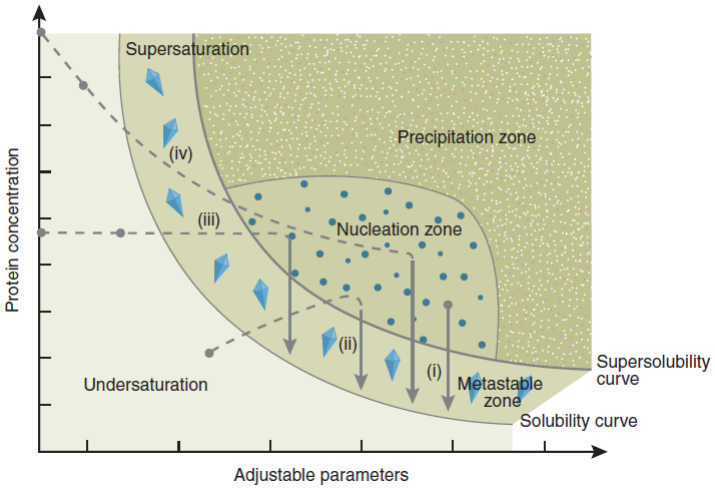

Figure 1.11: *Schematic representation of protein crystallization phase diagram as function of protein concentration and various adjustable parameters such as precipitant concentration, pH and temperature, taken from (Chayen & Saridakis, 2008).*

1.1.6 Self-assembly of anisotropically generated complex colloidal systems

The ability to build materials from anisotropic colloids has generated huge interest in the field of soft matter science. Nanoparticles and colloids have been generated by particle synthesis to yield a broad variety of building blocks with different shapes, compositions, patterns and functionalities (Glotzer & Solomon, 2007). Association of nanoparticles and colloids towards the formation of complex structures can be driven by shape anisotropy (Jones *et al.*, 2010, Jones *et al.*, 2011, Sacanna & Pine, 2011), directional entropic force (Young *et al.*, 2013) and surface functionalization (Kraft *et al.*, 2012, Wang *et al.*, 2013, Wang *et al.*, 2012) of individual building blocks. Another emerging approach that has been adopted for self-assembly of nanoparticles and colloids is anisotropically decorating the surface of particles with synthetic organic or biological molecules (Glotzer, 2004, Manoharan, 2004, Bianchi *et al.*, 2011, Breed, 2007, Zhang & Glotzer, 2004). This biologically inspired self-assembly has been exploited for the self-assembly of nanoparticles and colloids into specific structures and working devices (Long *et al.*, 2007, Bowden *et al.*, 2001).

DNA-functionalized colloids are one example of how biologically inspired self-assembly can be used to create well defined assembled structures. The process of self-assembly is initiated by coating the surface of colloidal nanoparticles with single stranded DNA (ssDNA). Using two particle types, each functionalised with

complementary single stranded DNA oligomers allows highly specific colloidal self-assembly. A 3-strand system has also been used where a linker ssDNA sequence complementary to the ssDNA coating the particles allows particle bridging (figure 1.12) (Geerts & Eiser, 2010).

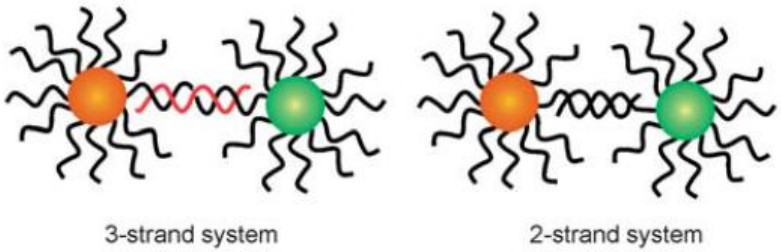

Figure 1.12: *DNA driven self-assembly of colloidal particles, taken from (Geerts & Eiser, 2010).*

In addition, Alivisatos and co-workers have demonstrated that the distance between the DNA-nanoparticle self-assembled structures can be manipulated by the DNA sequences used (Alivisatos *et al.*, 1996). Nature has designed numerous enzymes that specifically aid in cutting and joining DNA, hence, a wide variety of strands may be produced that would recognize its complementary counterparts and may assist the DNA-based self-assembly process. The work by Mirkin and Alivisatos emphasized the potential role of DNA in generating complex three-dimensional structures that can self-assemble in a controlled and reversible manner (Mirkin *et al.*, 1996).

Thus DNA-based self-assembly can generate one-, two- and three-dimensional structures that cannot be made through the assembly of spherical colloidal particles and more importantly, it is directional anisotropic interactions that have inspired the generation of these biologically relevant structures having wide applications in bio-sensing/bio-recognition and DNA/RNA delivery into cells (Geerts & Eiser, 2010, Tan *et al.*, 2013).

1.2 Protein structure and self-assembly

The concept of anisotropy is pivotal to the self-assembly of proteins (Derewenda, 2011, Lomakin *et al.*, 1999). Protein molecules have evolved to adopt a three-dimensional functional structure under physiological conditions. Proteins are polymeric chains of amino acids covalently joined by peptide bonds, which give a protein its primary structure (figure 1.13). The primary structure of a protein is stabilized by peptide bond formation which is a class of covalent bond formed between

a carboxylic acid and an amino group on two neighbouring amino acid molecules by the potential loss of a water molecule as shown in figure 1.13. The peptide bond formed is planar due to the delocalization of electrons giving it a partial double bond character, which restricts conformational freedom and flexibility within the protein molecule. The 20 most commonly known amino acids that participate in peptide bond formation in proteins are sequentially arranged in the form of a linear chain with the amino acid sequence dictated by the nucleotides in the gene encoding it.

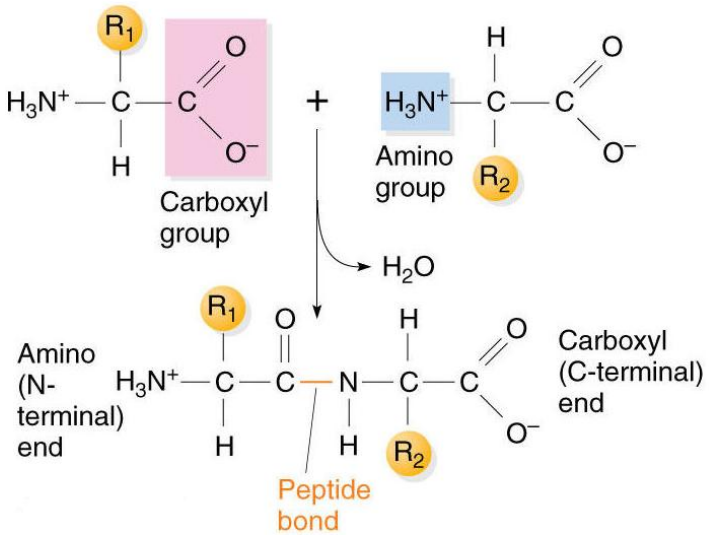

Figure 1.13: *Formation of a peptide bond between two amino acid units by condensation reaction, taken from (Russell, 2011).*

The primary polypeptide chain folds into α-helices or β-sheets formed through hydrogen-bonding interactions between N-H and C=O groups located in the invariant parts of the amino acids in the polypeptide backbone as shown in figure 1.14. The α-helical structure is stabilized by hydrogen bonds formed between the hydrogen atom attached to the electronegative nitrogen atom of a peptide bond and the electronegative carbonyl oxygen atom on the amino-terminal end of another peptide bond that are spaced four residues apart while β-sheets are composed of two or more stretches of at least 5-10 amino acids folded and aligned alongside one another, stabilized by hydrogen bonds as shown in figure 1.14. The elements of the secondary structure of a protein *i.e.* α-helices and β-sheets along with the non-secondary structural elements such as loops and links fold into tertiary structures (figure 1.15). Many proteins also form quaternary structures by the association of folded chains of more than one polypeptide unit as shown in figure 1.15. Therefore, the formation of stable

tertiary/quaternary structures by the process of folding is one major natural self-assembly process found in proteins.

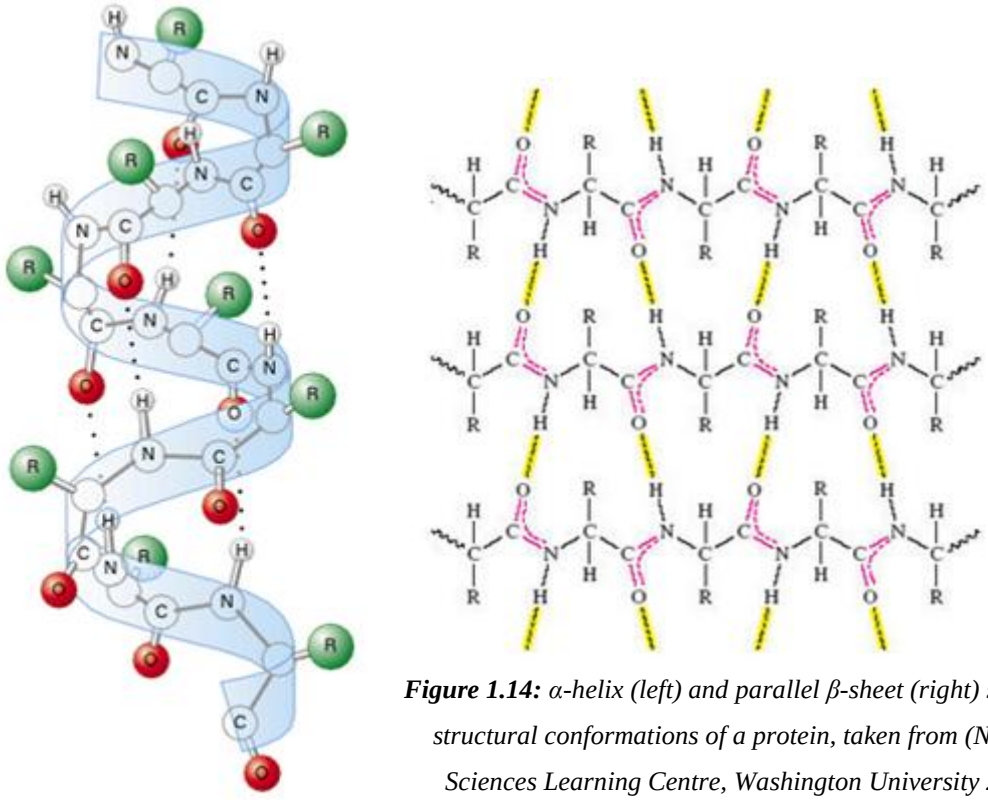

Figure 1.14: *α-helix (left) and parallel β-sheet (right) secondary structural conformations of a protein, taken from (National Sciences Learning Centre, Washington University 2011).*

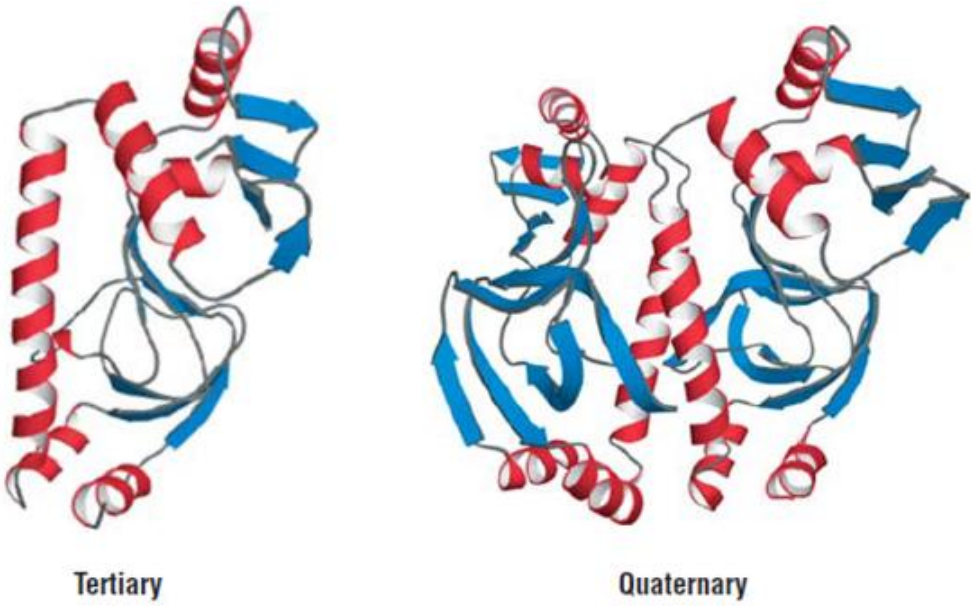

Figure 1.15: *Tertiary (3°) (left) and quaternary (4°) (right) protein structures formed by protein folding, taken from (Petsko & Ringe, 2004).*

1.2.1 Intermolecular self-assembly of proteins

Proteins are divided into two major classes: globular proteins and transmembrane proteins. Globular proteins are water-soluble proteins that occur in the cytosol of living cells and body fluids, while transmembrane proteins are embedded in cellular membranes and serve to transmit signals across the membrane between the extracellular matrix and the cytosol. Transmembrane proteins occur as single molecules in the cellular membrane space, in some cases they self-assemble into complex two-dimensional structures that carry out specific functions (Byrne & Iwata, 2002). In other cases, they self-assemble into two-dimensional crystals; the formation of two-dimensional crystals is mediated by the aggregation of these single molecules which can be compared to a phase transition phenomena, although poorly understood (Oesterhelt & Stoeckenius, 1973). Better understood are the phase transitions of globular proteins that result in the formation of several condensed phases such as dense liquid phase due to liquid-liquid phase separation, protein crystals and protein gels (Vekilov, 2010b).

A change in the physical state of the protein also gives rise to solid phases *i.e.* amorphous or disordered protein aggregates (figure 1.16). A different class of protein solid phase known as amyloid fibrils can also be formed by condensation that is facilitated by the partial unfolding of an amyloidogenic protein or peptide chain into linear, one-dimensional and ordered/structured aggregates (Lomakin, Chung, *et al.*, 1996, Kelly, 1997, Finkelstein & Ptitsyn, 2002). These ordered aggregates are associated with the formation of neurodegenerative disorders such as Alzheimer's and Parkinson's diseases (Carver *et al.*, 2002).

Phase transitions of globular proteins do not affect the structural integrity of protein molecules and the thermodynamic forces or the kinetic pathways that result in the transition phases are not controlled by intermolecular dynamics (Vekilov, 2010b, 2012). A detailed understanding about the physical laws that control these phase transitions and the fundamental mechanisms that drive them are essential to understand the formation of solid and dense-liquid condensed phases of compact, folded globular proteins.

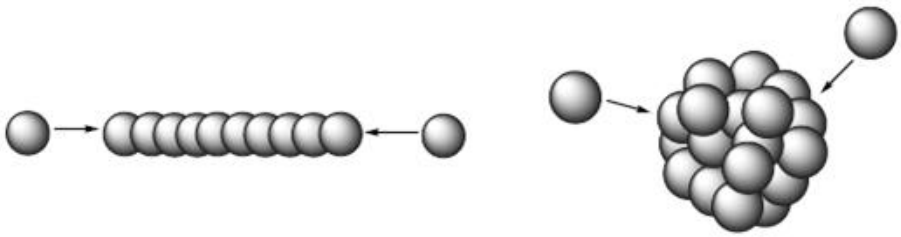

Figure 1.16: *Schematic representation of linear ordered aggregation (left) and disordered three-dimensional amorphous aggregation (right), taken from (Stranks et al., 2009) .*

1.3 Protein condensation

Protein condensation is a widely occurring phenomenon which has relevance in the following areas of study: 1. Protein condensation leading to the formation of diseases such as cataract, sickle cell anemia, Alzheimer's *etc.* 2. Protein condensation leading to the formation of protein crystals have been widely used to obtain the three-dimensional atomic structure of protein molecules by X-ray crystallography. 3. Protein condensation leading to the formation of undesirable protein aggregates in bio-pharmaceutical industries. These are discussed in detail below.

1.3.1 Protein condensation diseases

Protein condensation is a contributing factor during the pathogenesis of several diseases. Some of these include cataract, sickle cell disease and Alzheimer's disease and are as described below.

Cataract is a protein condensation disease, caused by opacity of the eye lens due to changes in its refractive index over distances comparable to the wavelength of transmitted light (Delaye & Tardieu, 1983, Benedek, 1971) (figure 1.17).

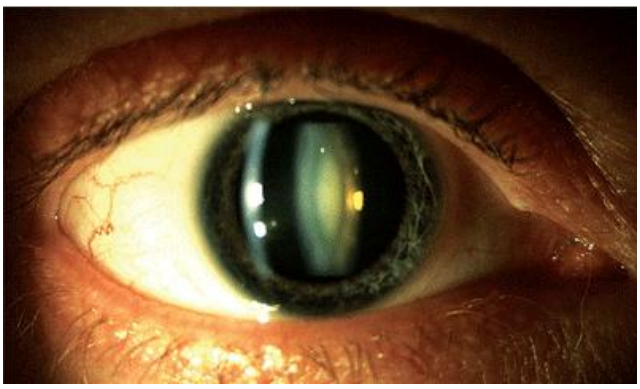

Figure 1.17: *Cataractous lens showing protein deposits taken from (Review of Optometry 2011).*

The vertebrate lens is an avascular cellular structure which is mainly composed of highly elongated, terminally differentiated fiber cells (Bassnett *et al.*, 2011, Kuszak *et al.*, 2004) as shown in figure 1.18. Within the fiber cells, the α-, β- and γ-crystallins are the most abundantly found proteins. The total protein concentration in the nucleus of the eye lens exceeds 400 $mg.ml^{-1}$ (Hejtmancik, 2008). The high concentration of the soluble crystallin proteins contributes to the high refractive index of the eye-lens (Slingsby *et al.*, 2013, Wistow, 2012, Bloemendal *et al.*, 2004), responsible for diffracting and concentrating light on the retinal cells.

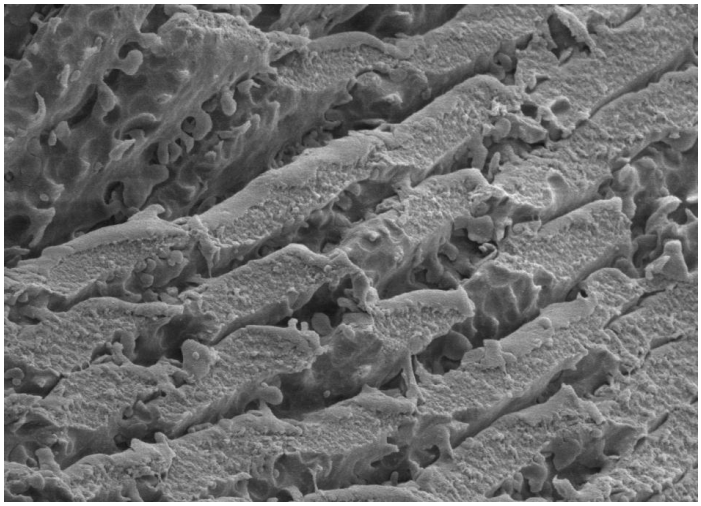

Figure 1.18: *Scanning electron microscopy image of eye lens fibre cells, taken from the (The Institute of Optics, University of Rochester).*

The γ-crystallins belong to a family of structurally homologous, monomeric proteins, expressed in mammalian lenses (Schoenmakers *et al.*, 1984, Summers *et al.*, 1984). The γ-crystallins in the eye lens are small, compact, globular, structural proteins with short-range order and together with α- and β-crystallins are important in maintaining the transparency and refractive index of the eye lens to produce clear focussed images on the light sensitive cells of the retina. However, it is the short-ranged attractive protein-protein interactions that differentiate the γ-crystallins from the α- and β-crystallins (Tardieu *et al.*, 1992). These attractive interactions within the eye lens reduces the osmotic pressure, but make the protein mixture more prone to phase separation and aggregation (Clark, 1994). In spite of this, γ-crystallins remain soluble at very high concentrations exceeding 400 $mg.ml^{-1}$ with very little turn over (Zigler, 1994).

There are two types of cataract that affect human populations worldwide. Genetically induced congenital/childhood cataracts, caused by single-point mutations in the crystallin proteins of the eye lens, where crystallin protein mutations result in the formation of protein condensed phases such as protein crystals (Pande *et al.*, 2001) or protein aggregates (Pande *et al.*, 2005) which significantly scatter light and cause mild to severe forms of opacity. The other is age-related cataract caused by increased susceptibility to environmental changes such as light, hyperglycaemic, osmotic, oxidative or other stresses (Hejtmancik *et al.*, 2001). Certain post-translational chemical modifications such as proteolysis, intermolecular disulfide bridges, deamidation of asparagine and glutamine residues, racemization of aspartic acid residues, phosphorylation, non-enzymatic glycosylation and carbamylation (Hejtmancik *et al.*, 2001) in the eye lens proteins result in the formation of protein aggregates associated with ageing lenses. The progression of a cataractogenic condition is characterized by increased lens turbidity and light scattering due to the formation of higher molecular weight aggregates. These protein aggregates may be held together by covalent forces such as disulfide and glycosidic linkages or non-covalent forces such as hydrophobic, van der Waals, hydrophilic and electrostatic interactions.

Sickle cell disease is a protein condensation disease caused by single-point mutation at the 6^{th} site in each of the two β-chains of the hemoglobin tetramer by substituting the hydrophilic, negatively-charged glutamatic acid residue with the hydrophobic, positively-charged valine residue (Ingram, 1956, Vekilov, 2007). Hemoglobin found in red blood cells, are globular proteins that carry oxygen (O_2) and carbon dioxide (CO_2) molecules throughout the body, each hemoglobin protein molecule comprises of four polypeptide chains, two identical α-chains and two identical β-chains held together by non-covalent interactions.

A single-point mutation of the tetrameric hemoglobin protein results in the formation of hemoglobin S which alters the phase transition of the normal protein under deoxygenated conditions forming polymerized, helically twisted fibres (figure 1.19) that associate to form a gel phase (Bookchin & Nagel, 1981, Hirsch *et al.*, 1985). The resulting fibrillar condensate causes distortion in the shape of red blood cells resulting in a sickle appearance which undergo reversible phase transition from gel phase into aqueous hemoglobin solution upon re-oxygenation (Eaton & Hofrichter, 1990). Homozygous CC disease is another example for a hemoglobin protein condensation

disease in red blood cells caused by a single-point mutation at the 6th amino acid position of the β subunit that replaces negatively charged glutamate with a positively charged lysine (Hunt & Ingram, 1958). The single-point mutation results in the formation of protein crystals in vivo under oxygenated conditions (figure 1.19) (Hirsch *et al.*, 1985, Lin *et al.*, 1989, Feeling-Taylor *et al.*, 2004), with a mild chronic hemolytic anemia (Lessin *et al.*, 1969).

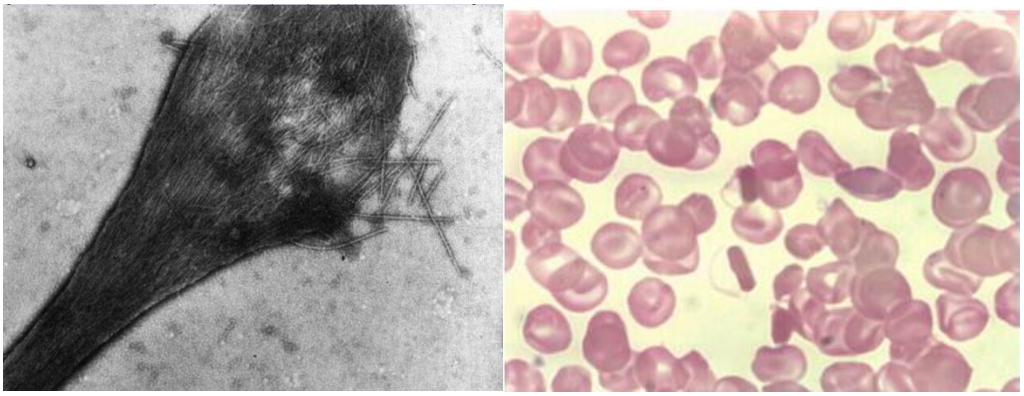

Figure 1.19: *Formation of aggregates or polymerized helically twisted fibres in hemoglobin S disease (left) taken from (Biology Learning Centre, The University of Arizona 2007) and protein crystals in hemoglobin C disease (right) taken from (Interim LSU Hospital).*

Alzheimer's disease is a protein condensation disease formed by the association of the β-amyloid peptide (Aβ) into long fibrils that eventually form amyloid plaques. Both fibrils and plaques are somewhat neurotoxic, but more recent work has indicated that in fact smaller oligomeric forms of the protein are significantly more neurotoxic than the larger aggregates (Selkoe, 1994).

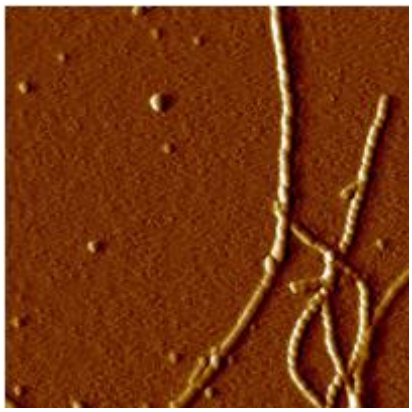

Figure 1.20: *Formation of Aβ fibrils during the pathogenesis of Alzheimer's disease, taken from (The Scripps Research Institute).*

Aβ is a ~4 KDa protein comprising a common core sequence with heterogeneous N- and C- termini. The most common form of Aβ is 40 amino acids long referred to as

$A\beta_{40}$. $A\beta_{42}$, the lesser known component differing by the presence of two additional amino acids is also associated with the disease (Irvine *et al.*, 2008). Aβ peptide is the product of enzymatic breakdown of a large membrane protein called amyloid precursor protein; APP. Alzheimer's disease is characterized by a decrease in the concentration of Aβ monomer resulting in protein condensation due to the formation of Aβ oligomers and long protein fibrils (figure 1.20).

1.3.2 Protein crystallization

1.3.2.1 Understanding self-assembly of proteins leading to protein crystallization

Understanding protein interactions in aqueous solutions is key to understanding the self-assembly of proteins into the formation of solid condensed phases such as protein crystals. The osmotic second virial coefficient, B_{22}, is a thermodynamic parameter that has predominantly been used in the literature to characterise protein interactions arising from weak non-covalent interactions such as electrostatic repulsions, van der Waals attractions, hydrophobic and hydration forces (Zimm, 1946, Neal *et al.*, 1998). George *et al.* have pointed out that negative B_{22} values correspond to attractive interactions that strongly correlated with the crystallization conditions for globular proteins (George *et al.*, 1997). Another thermodynamic parameter that is used to characterise protein interactions is the k_D (net interaction potential) values from QELS (Quasi-elastic light scattering) experiments (James & McManus, 2012), where negative k_D (net interaction potential) values correspond to attractive interactions. While B_{22} and k_D values are useful for some proteins such as lysozyme, it isn't always a good predictor of protein solution behaviour.

Proteins, in general are difficult to crystallize. One possible explanation is that polypeptide chains that make up a protein have significant amount of conformational entropy that affects crystallization properties of proteins (Doye *et al.*, 2004). Thus, proteins have evolved not to crystallize because it compromises the viability of the cell. Secondly, proteins have directional interactions that may not be compatible with simple crystal lattices (Sear, 1999). Thirdly, formation of kinetically trapped phases such as aggregates and gels are enhanced by directional, anisotropic interactions which suppress the nucleation of protein crystals (Muschol & Rosenberger, 1997, Lomakin *et al.*, 1999, Romano *et al.*, 2010).

In order to obtain diffraction quality protein crystals, the protein molecule must be immobilized in a lattice structure with limited dynamic motion. Several physical techniques have been employed to encourage protein crystallization by controlling conditions such as temperature, pH and the addition of chemical excipients to achieve supersaturation, a non-equilibrium condition that favors the spontaneous appearance of a new phase. Protein crystallization is a first-order phase transition that occurs via nucleation and growth of the critical nuclei into large single crystals. The first and most important pre-requisite for protein crystallization is nucleation. If the concentration of the protein solution lies near to the metastable critical point on the phase diagram for the protein, concentration fluctuations that leads to separation of the protein into protein-rich and protein-poor phases enhance protein crystal nucleation (ten Wolde & Frenkel, 1997). The initial critical nucleus possesses high free energy, thus its formation is only possible by overcoming the energy barrier towards nucleation as shown in figure 1.21. The initial critical nucleus is an irreversible, transient, thermodynamically stable phase that is capable of independent existence, acting as crystal centres for further growth into larger crystals by the diffusion and attachment of growth units to its surface and their subsequent incorporation into the protein crystal lattice structure.

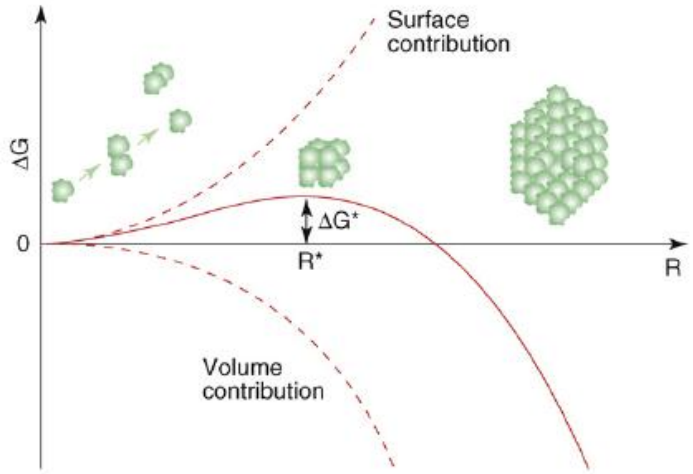

Figure 1.21: *Hypothetical diagram showing energy barrier for nucleation, taken from (Saridakis & Chayen, 2009).*

Nucleation is an entropy driven process. A crystalline state is generally favored by lowering the entropy of the system. In order to counterbalance the loss in entropy, a decrease in enthalpy must occur to trigger the incorporation of a protein molecule into the bulk of a crystal. However, it has been observed that formation of a protein crystal

is accompanied by an increase in enthalpy. This suggests that the loss in entropy due to crystal formation is balanced by lowering the free energy due to the formation of contacts in a protein crystal. Theory, simulations and experiments have predicted that density fluctuations near LLPS can significantly reduce the energy barrier to nucleation and enhance the rate of crystal nucleation (ten Wolde & Frenkel, 1997, Talanquer & Oxtoby, 1998). In addition to the classical nucleation mechanism, a two-step nucleation mechanism (ten Wolde & Frenkel, 1997, Talanquer & Oxtoby, 1998) supported by recent research proposes that protein crystallization proceeds via a two-step nucleation and growth phase (Erdemir *et al.*, 2009, Vekilov, 2010a, 2005).

1.3.2.2 Thermodynamics and driving forces for protein crystallization

The thermodynamics of protein phase transitions are in many ways similar to the thermodynamics of phase transitions in colloidal solutions (Hagen & Frenkel, 1994, Asherie *et al.*, 1996, Adams *et al.*, 1998, Van Megen & Underwood, 1993). Protein molecules in solution prior to the formation of solid condensed phases (crystals/amorphous aggregates), are surrounded by water molecules. As the protein molecules assemble to form a new phase, the water molecules undergo some restructuring or rearrangement. This restructuring is accompanied by changes in the Gibbs free energy (Prausnitz & Foose, 2007). The change in Gibbs free energy of crystallization, ΔG^0_{cryst} at constant temperature T is the sum of the contributions from the enthalpic term ΔH^0_{cryst} and the entropic term ΔS^0_{cryst} defined by equation 1.7,

$$\Delta G^0_{cryst} = \Delta H^0_{cryst} - T\Delta S^0_{cryst} \qquad 1.7$$

When protein molecules are engaged in Brownian motion in solution, there is no order associated with it. However, when a three-dimensional structuring or order is imposed on these protein molecules in the crystal lattice, it is accompanied by a massive decrease in entropy that drives crystal formation. This decrease in entropy is associated with the loss of six translational and rotational degrees of freedom per protein molecule and is only partly compensated for by the newly created vibrational degrees of freedom (Finkelstein & Janin, 1989, Tidor & Karplus, 1994). This contributes to an increase in free energy, ΔG^0_{cryst} which disfavours crystallization. This can only be compensated for by lowering the enthalpic term, ΔH^0_{cryst} or the entropy of accompanying processes. However, thermodynamic measurements for several proteins

such as lysozyme (Schall *et al.*, 1996) and hemoglobin C (Vekilov, Feeling-Taylor, Petsev, *et al.*, 2002), indicate that enthalpic effects are unlikely to favor crystallization and that entropic effects are more important. The incorporation of a protein molecule into a crystal lattice is associated with the release or additional trapping of structured water/solvent molecules (Vekilov, 2012). Hence, the solvent and protein entropy changes are integrated into equation 1.7 which is modified into equation 1.8 as shown below,

$$\Delta G^0_{cryst} = \Delta H^0_{cryst} - T(\Delta S^0_{protein} + \Delta S^0_{solvent}) \qquad 1.8$$

A more negative ΔG^0_{cryst} that aids crystallization is favored by positive sum of $\Delta S^0_{protein} + \Delta S^0_{solvent}$. The estimated values for $\Delta S^0_{protein}$ from experimental studies are in the negative region and the negative contribution is only overcome by positive contribution from the $\Delta S^0_{solvent}$ values (Vekilov & Chernov, 2002). This increase in entropy ($\Delta S^0_{solvent}$) is accompanied by the release of water/solvent molecules upon the incorporation of a protein molecule into a crystal lattice. (Vekilov, Feeling-Taylor, Yau, *et al.*, 2002).

1.3.2.3 Methodology for protein crystallization

Creating the correct conditions for crystallization to occur, often involves screening of large numbers of potential solution conditions, which may favour nucleation. Currently brute force approaches are employed using high-throughput systems to prepare large number of conditions that favour protein crystallization. The general approaches are:

i. Altering the pH of the buffer in which the protein is dissolved. Changing the pH of the protein solution alters the charge on the protein surface.
ii. Altering the chemical activity of the hydration layer around the protein by the addition of salts (*e.g.* ammonium sulphate).
iii. Altering the degree of attraction between protein molecules. This is achieved by changing the pH or the addition of bridging ions (*e.g.* ligands).
iv. Altering the nature of interactions between protein molecules and the solvent by the addition of organic solvents (Cohn *et al.*, 1947), long-chain polymers (Arakawa & Timasheff, 1985) and non-volatile organic compounds (Stauber *et al.*, 2015).

v. Altering the temperature and concentration of the protein solution.
vi. Inducing crystallization in proteins due to the influence of precipitant stereochemistry (Stauber *et al.*, 2015, Asherie *et al.*, 2008).

Chemical precipitants for protein crystallization

Chemical precipitants are widely used to obtain supersaturation for protein crystal growth. The most commonly used precipitants fall into four broad categories: (i) salts, (ii) organic solvents, (iii) long-chain polymers, (iv) low molecular weight polymers and non-volatile organic compounds.

The effect of salts on protein crystallization is highly influenced by the ionic species that constitute it. Salting-out salts such as CH_3COONa, NaCl and Na_2SO_4 are preferentially excluded from the protein surface thereby hydrating the protein molecule and the increase in chemical potential associated with the presence of these salts can lead to phase separation and formation of the solid phase (Arakawa & Timasheff, 1982). In contrast, salting-in salts such as KSCN, $MgCl_2$ and $CaCl_2$ extensively bind to protein surfaces with their anionic or cationic portion binding to specific protein surfaces and the distribution of the counter-ions in the electric double layer around the protein, this screening of electric charges on the surface of the protein can lead to phase separation and formation of the solid phase (Arakawa & Timasheff, 1982). Organic solvents (*e.g.* ethyl alcohol) induce crystallization in protein solutions by reducing dielectric constant of the medium. It is thought that screening effects of the solvent promote attraction between protein molecules favoring formation of the solid phase (Englard & Seifter, 1990, Cohn *et al.*, 1947). The third class of precipitants include long-chain polymers (*e.g.* polyethylene glycol), which are preferentially excluded from the protein surface. (Arakawa & Timasheff, 1985, Ingham, 1990, McPherson, 1976). Solid phase formation in the presence of non-volatile organic compounds (*e.g.* 2-methyl-2,4-pentanediol) is also thought to be via preferential exclusion from the protein surfaces due to repulsion from their charged residues (Bolen, 2004).

Physical techniques for protein crystallization

The process of protein crystallization begins from screening for favorable crystallization conditions to optimizing crystal growth to obtain diffraction quality crystals. Protein crystallization screening trials are usually used to find ‘hits’ or ‘leads’

for the conditions that favor crystallization (Chayen & Saridakis, 2008). Observation of crystals, crystalline precipitate and phase separation in solution are leads to identifying the suitable conditions that favour the process. Protein phase diagrams have also been very effective tools in determining conditions that favour protein crystallization (Asherie, 2004). When the conditions are identified, it is followed by optimization of the crystallization process. This may be achieved by varying the concentration of protein, type and concentration of precipitant/additive, pH or temperature.

Protein crystal screening can take place in the following formats:

i. Microbatch experiments

In microbatch experiments, small volumes of protein and precipitant are mixed and incubated under low-density paraffin oil (0.87 $g.ml^{-1}$) as shown in figure 1.22. The aqueous crystallization drops sink beneath the denser paraffin oil, where they are protected from evaporation and contamination (Chayen & Saridakis, 2008). The conditions for protein crystallization in a microbatch experimental set-up are constant within the time frame of a crystallization experiment and there are no changes in the drop volume or pH and the crystals generally do not dissolve (Chayen *et al.*, 1990, Chayen, 1998).

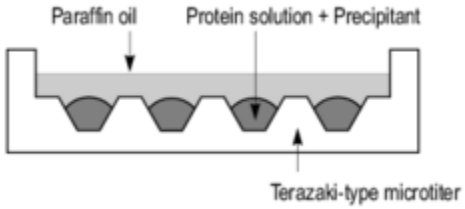

Figure 1.22: *Schematic representation of microbatch protein crystallization technique widely used in crystallizing proteins.*

ii. Vapour diffusion

The most commonly used crystallization technique is vapour diffusion as shown in figure 1.23, where the protein solution is either a sitting or a hanging drop. This is then allowed to equilibrate against a reservoir containing a crystallizing agent at either higher or lower concentration than in the drop (Chayen & Saridakis, 2008). Diffusion based methods are highly dynamic systems where the conditions change throughout the crystallization process.

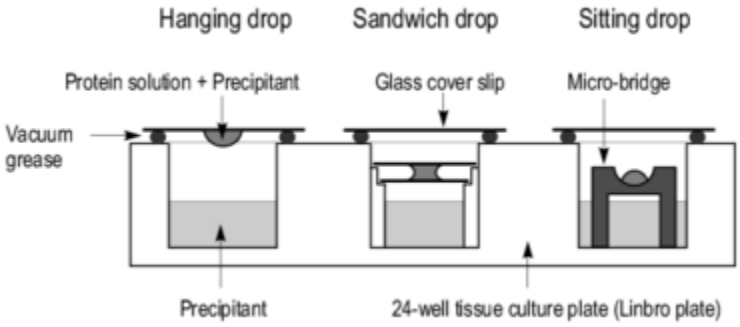

Figure 1.23: *Schematic representation of vapour-diffusion protein crystallization technique widely used in crystallizing proteins.*

iii. Dialysis

In this method, the protein solution is initially contained within a semi-permeable membrane, which is then allowed to equilibrate against a precipitant solution as shown in figure 1.24. Equilibration and diffusion against the precipitant slowly raises the supersaturation of the protein solution within the dialysis membrane eventually resulting in crystallization (Chayen & Saridakis, 2008).

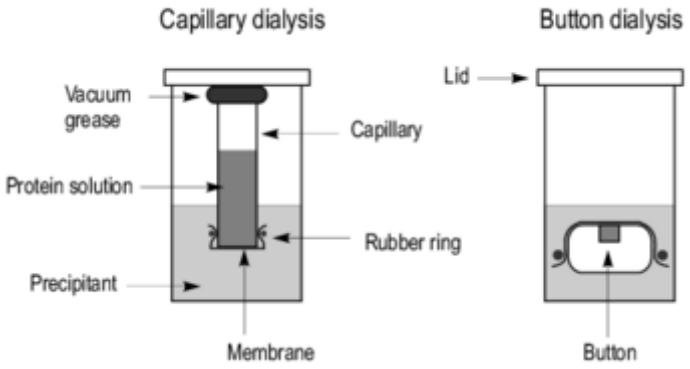

Figure 1.24: *Schematic representation of dialysis protein crystallization technique widely used in crystallizing proteins.*

iv. Free interface diffusion

This method involves placing the protein and precipitant solutions in close proximity whereby a concentration gradient is established with time and this directs the system towards self-nucleation and growth (Chayen & Saridakis, 2008, McPherson, 1999, Garcia-Ruiz *et al.*, 2002).

v. Microfluidics

High-throughput crystallization screening is carried out using microfluidic chips (Hansen & Quake, 2003). This method not only requires minimum quantities of protein and works with maximum pace and efficiency, but also follows fluid flow and mass transport principles which may increase the likelihood of protein crystallization in solutions (Chayen & Saridakis, 2008, Zheng *et al.*, 2003, Li & Ismagilov, 2010).

1.3.2.4 Use of molecular engineering and synthetic symmetrisation to promote protein crystallization

Obtaining single well-diffracting single crystals is the major limiting step towards protein crystal structure prediction from X-ray crystallographic analysis (DeLucas & Bugg, 1987, Chayen *et al.*, 1996). Protein structural data forms the basis of all structural genomic studies involving protein structure-function correlations (Darby & Creighton, 1993) and also plays a significant role in rational drug design. One of the major constraints towards the formation of single crystals is the presence of large hydrophilic residues on the surface of the protein that prevents the formation of intermolecular contacts required for protein crystallization. Thus, protein surface properties play a determining role in the formation of tight intermolecular contacts, required for crystal lattice stability. The concept of 'surface-entropy reduction' was recently proposed to overcome the crystallization barrier by substituting large, bulky surface-exposed residues with small amino acids such as alanine that has minimum conformational entropy using site-directed mutagenesis (Derewenda, 2004, Longenecker *et al.*, 2001, Derewenda & Vekilov, 2006), which was further supported by strong theoretical work (Cieślik & Derewenda, 2009). Among the polar amino acids that are surface-exposed, lysine and glutamate are the most prominent (Baud & Karlin, 1999), that have evolved a 'surface-entropy shield' which prevents their non-specific aggregation and precipitation (Doye *et al.*, 2004) and their side chains are characterized by high conformational energy of ~2 kcal/mol (Avbelj & Fele, 1998). Thus, replacing lysine and glutamate residues constitute a good target for surface modification and crystal engineering of proteins (Derewenda, 2010).

Another concept that has been employed to crystallize proteins is by synthetically symmetrizing proteins (Banatao *et al.*, 2006). This has been achieved by generating symmetric proteins such as homodimers that have increased tendency to crystallize

from naturally monomeric proteins by forming disulfide bridges between individual cysteine residues introduced by mutagenesis. On similar lines, Yeates and Kent have described a concept called 'racemic protein crystallography' to solve protein structures (Yeates & Kent, 2012). All natural proteins are chiral and the chirality of protein molecules has important implications for how they interact with other molecules in the cell and how they crystallize. They have utilized the chirality of the crystal space groups to generate the mirror image (enantiomer) of a protein molecule, and the mirror images that form racemic mixtures had greater potential to crystallize in a wider range of possible space groups (Yeates & Kent, 2012). Thus, racemic protein crystallography was another novel approach towards obtaining well-ordered crystals for protein structure determination (Berg & Goffeney, 1997, Matthews, 2009).

1.3.3 Understanding protein condensation leading to protein aggregation in bio-pharmaceutical industries

Protein aggregation in bio-pharmaceutical industries is an undesirable form of protein self-assembly as it leads to reduced or no biological activity of the potential drug molecule and enhanced potential for immunogenicity or side effects. It has emerged as one of the most important areas of extensive research and development since the primary goal of bio-pharmaceutical industries has been to identify the cause of protein aggregation and control protein aggregation at its earliest stage. These protein aggregates have primarily influenced the commercialization of drug products by reducing their efficacy. Significant amount of research has been carried out to understand several aspects of protein self-assembly leading to protein aggregation. The broad field of protein aggregation encompasses various aspects such as mechanism and driving forces leading to physical aggregation (Chi *et al.*, 2003), aggregation mechanisms (Philo & Arakawa, 2009), aggregation kinetics (Morris *et al.*, 2009), aggregation predictions (Weiss *et al.*, 2009), theoretical considerations of aggregation (Gsponer & Vendruscolo, 2006), inhibition of aggregation (Wang, 2005, Hamada *et al.*, 2009) and aggregation related to diseases (Wang *et al.*, 2008).

1.3.4 Aqueous two-phase separation systems for industrial protein purification

Polyethylene glycol has been widely used as a crystallizing agent, creating an effective attraction between protein molecules due to depletion forces (Gilliland, 1988, Anderson & Lekkerkerker, 2002). However, polyethylene glycol has most often been used in combination with a salt, resulting in liquid-liquid phase separation of proteins. At low salt concentrations, liquid-liquid phase separation of protein occurs due to repulsive electrostatic interaction between protein molecules while the nature of the salt dictates phase separation at relatively high salt concentrations (Dumetz, Lewus, *et al.*, 2008).

The partitioning of protein between the two phases also depends on the hydrophobicity of the protein and the polymer concentration. Protein concentration also influences partitioning of protein between the two phases. This method of aqueous two-phase separation have found wide applications in the purification of a large number of industrial enzymes and separation of virus and virus-like particles (Asenjo & Andrews, 2011).

Chapter 2

Experimental methods and materials

2.1 Preparation of materials

2.1.1 Preparation of buffers

Analytical grade buffer salts were purchased from Fisher Scientific (Dublin, Ireland) and used without further purification. All buffers were prepared in Milli-Q water, adjusted to the appropriate pH using either sodium hydroxide (NaOH), hydrochloric acid (HCl) or acetic acid (CH_3COOH) solutions and filtered through a 0.45 µm nylon membrane filter using a Millipore glass assembly filter under vacuum (Merck Millipore) before use. Buffers used for SE-HPLC were filtered using the above method and degassed using the online degasser. Buffers used for gel electrophoresis were degassed using an external vacuum pump.

2.1.1.1 Preparation of sodium phosphate buffer

0.1 M sodium phosphate buffer at pH 7.0 was prepared using 15.6 g of sodium dihydrogen phosphate dihydrate (Mw of $NaH_2PO_4.2H_2O$ = 156.007 $g.mol^{-1}$) and 26.8 g of sodium phosphate dibasic heptahydrate (Mw of $Na_2HPO_4.7H_2O$ = 268.07 $g.mol^{-1}$) in Milli-Q water and made upto 1 L. The pH was adjusted using a NaOH solution. 0.02% (w/v) sodium azide (NaN_3) was added to prevent bacterial growth in buffer solutions.

2.1.1.2 Preparation of sodium acetate buffer

0.275 M sodium acetate buffer at pH 4.5 was prepared using 16.5 g of glacial acetic acid (CH_3COOH) and 4.5 g of sodium hydroxide (NaOH) in 1 L milli-Q water. While preparing sodium acetate buffers, direct mixing of glacial acetic acid with sodium hydroxide was avoided to prevent an explosive exothermic reaction.

Sodium acetate buffer with salt. 0.275 M sodium acetate buffer at pH 4.8 with 0.325 M sodium chloride (NaCl) contained 16.5 g of glacial acetic acid (CH_3COOH), 6.7 g of sodium hydroxide (NaOH) and 76.1 g of sodium chloride (NaCl) in 1L Milli-Q water. The pH of sodium acetate buffers were adjusted using CH_3COOH or NaOH solution as appropriate. 0.02% sodium azide (NaN_3) was added to prevent bacterial growth in buffer solutions.

2.1.1.3 Preparation of Tris-HCl buffers

1 M Tris-HCl at pH 8.8 and 0.5 M Tris-HCl at pH 6.8 were prepared. 1 M Tris-HCl was prepared by dissolving 12.11 g of Tris Base (Mw of Tris Base = 121.1 $g.mol^{-1}$) in Milli-Q water and made up to 100 ml. 0.5 M Tris-HCl was prepared by dissolving 6.06 g of Tris Base in Milli-Q water and made up to 100 ml. The pH was adjusted using HCl solution.

2.1.2 Preparation of reagents

2.1.2.1 Preparation of cell lysis solution

The cell lysis solution containing 50 mM Tris HCl, 25 mM NaCl and 2 mM EDTA (Ethylene diamine tetraacetic acid disodium salt hydrate) in Milli-Q water, adjusted to pH 8.0 by the addition of NaOH solution was prepared in a volume of 10 ml, followed by the addition of one complete protease inhibitor cocktail tablet (25 MU, Roche) and 155 mg of DTT (Dithiothreitol).

2.1.2.2 Preparation of destaining solution for gel electrophoresis

The destaining solution for gel electrophoresis was prepared by mixing together Milli-Q water, methanol and acetic acid in the volume ratio of 60/30/10 (v/v/v).

2.2 Recombinant protein production

2.2.1 Bacterial cells for DNA transformation reactions

Two types of bacterial cells were utilized for our experiments involving DNA transformation reactions. Library Efficiency® DH5α™ Competent Cells purchased from invitrogen (Life Technologies, Dublin, Ireland) were used for the transformation of native and mutated DNA material. The transformed cells were grown, harvested and subjected to plasmid extraction and purification procedures to obtain recombinant DNA. The recombinant DNA was then transformed into BL21-Gold (DE3) Competent Cells purchased from Stratagene (U.S.A) for recombinant protein over-expression.

2.2.2 Bacterial cell culture

2.2.2.1 Preparing LB agar plates and LB (Luria-Bertani) broth for bacterial growth

LB agar (Fisher Scientific, Dublin, Ireland) was prepared according to the manufacturer's instructions in Milli-Q water and sterilized by autoclaving at 121°C and 0.212 MPa pressure for 20 minutes on a liquid heating cycle in a TOMY SX-500E high pressure autoclave. The sterile LB agar was allowed to cool to below 50°C before the addition of ampicillin to a final concentration of 100 $\mu g.ml^{-1}$. The warm LB agar was poured into sterile petri dishes under aseptic conditions and left at room temperature for a few hours to solidify. The solidified LB agar plates were stored at 4°C until further use.

2.2.2.2 Preparation of LB Broth

LB broth (Fisher Scientific, Dublin, Ireland) was prepared as per the manufacturer's instructions in Milli-Q water and sterilized by autoclaving at 121°C and 0.212 MPa pressure for 20 minutes on a liquid heating cycle in a TOMY SX-500E high pressure autoclave. Ampicillin was added to the cooled broth to a final concentration of $100\mu g.ml^{-1}$.

2.2.2.3 Transformation of Library Efficiency® DH5α™ Competent Cells

Library Efficiency® DH5α™ Competent Cells purchased from invitrogen (Life Technologies, Dublin, Ireland) were transformed using a standard protocol provided with the transformation kit. Briefly, the competent cells were thawed, gently mixed with the DNA and incubated on ice, followed by a heat-shock procedure. The heat-shock method involves heat-treatment that triggers the opening of bacterial cell membrane pores that then facilitate the entry of DNA material into the cell. The heat-shocked cells were added to 0.9 ml of S.O.C. medium (Super optimal broth with catabolic repressor) purchased from invitrogen (Life Technologies, Dublin, Ireland) at room temperature and thoroughly mixed at 37°C with 225 rpm shaking for 1 hour in an Innova 42 incubator shaker (New Brunswick Scientific Co., INC). The experimental reactions were diluted 1:5 with S.O.C. medium. 200 µl of the diluted mixture was spread on LB agar plates containing 100 $\mu g.ml^{-1}$ ampicillin. The plates were incubated overnight at 37°C for the appearance of white colonies. After 12-18 hours, a single colony was picked and subsequently grown in 5 ml of LB (Luria-

Bertani) broth containing 100 µg.ml^{-1} ampicillin to be used for plasmid DNA extraction and purification.

2.2.2.4 Growth and bacterial harvesting for plasmid extraction

A single colony from a freshly streaked LB agar plate was picked and inoculated in a 5 ml LB medium containing 100 µg.ml^{-1} ampicillin. The inoculated broth was incubated for 8 hours at 37°C with vigorous shaking (300 rpm). The starter culture was diluted 1/1000 into 100 ml of LB medium and further grown for 12-16 hours at 37°C with vigorous shaking (300 rpm). The bacterial cells were harvested by centrifugation at 6000 g for 15 minutes at 4°C.

2.2.2.5 Plasmid extraction and purification

Plasmid DNA extraction was carried out using QIAGEN plasmid purification midi kit (Qiagen GmbH, Hilden, Germany) according to a standard manufacturer protocol. The supernatant containing the plasmid DNA following plasmid extraction was loaded onto the QIAGEN-tip, washed and eluted out. The eluate was subjected to isopropanol precipitation; DNA pellets from isopropanol precipitation have a glassy appearance and may sometimes be loosely attached to the sides of the tube. Hence, the DNA pellets were carefully separated by gently decanting out the supernatant. The glassy DNA pellet was washed with 70% ethanol and centrifuged further. The supernatant was removed without disturbing the DNA pellet; it was air-dried, redissolved in a suitable volume of water and stored at -20°C until further use.

2.2.2.6 Determination of DNA yield

The purity of DNA was determined by measuring the absorbance of the diluted DNA sample at 260nm and 280nm. The ratio of (Absorbance $_{260}$/Absorbance $_{280}$) is an indication of the nucleic acid purity. A value greater than 1.8 indicates nucleic acid content greater than 90%. The concentration of DNA solutions (equation 2.1) were estimated by measuring UV absorbance at 260nm, multiplying by the dilution factor, and using the relationship that Absorbance$_{260}$ of 1 = 50 µg.ml^{-1} of dsDNA.

$$\text{DNA concentration} = \text{Absorbance}_{260} \times 50 \times (\text{dilution factor}) \ \mu g.ml^{-1} \qquad 2.1$$

2.2.2.7 Transformation of BL21-Gold (DE3) competent cells

BL21-Gold (DE3) competent cells purchased from Stratagene (U.S.A) were transformed using a standard protocol; hence a detailed step by step description is not required. The competent cells were thawed, gently mixed with the DNA material and incubated on ice, followed by a heat-pulse reaction. The heat-shock method involves heat-treatment that triggers the opening of bacterial cell membrane pores that then facilitate the entry of DNA material into the cell. The heat-pulsed cells were added to 0.9 ml of S.O.C. medium (Super optimal broth with catabolic repressor) purchased from invitrogen (Life Technologies, Dublin, Ireland) at room temperature and then incubated at 37°C with shaking at 225 rpm for 1 hour in an Innova 42 incubator shaker (New Brunswick Scientific Co., INC). The transformed cells were centrifuged at 200 × g for 3 minutes and plated on LB agar plates containing 100 $\mu g.ml^{-1}$ ampicillin. The plates were incubated overnight at 37°C for the appearance of white bacterial colonies.

2.2.2.8 Preparation of bacterial cell stock

The transformed E.coli BL21-Gold (DE3) competent bacterial cells with native and mutated DNA material were grown on LB agar plates and incubated overnight at 37°C. A single colony was picked and subsequently grown in 100 ml of LB broth containing 100 $\mu g.ml^{-1}$ ampicillin for 12-16 hours. 2 ml of the inoculated overnight culture was pipetted into Nalgene cryogenic vials under aseptic conditions and stored at -80°C in an ultra low temperature freezer (New Brunswick Scientific) for further use.

2.2.2.9 Growing bacterial cultures for recombinant protein production

The BL21-Gold (DE3) competent cells purchased from Stratagene (U.S.A) and transformed with either native or mutated DNA material from frozen stocks were streaked using 1 µl clear sterile disposable loops onto sterile LB agar plates. A single colony from the streaked plate was picked and allowed to grow in a sterile small culture flask containing 100 ml of LB medium and 100 $\mu g.ml^{-1}$ ampicillin for 12-16 hours at 37°C in an Innova 42 incubator shaker (New Brunswick Scientific Co., INC) with shaking at 225 rpm until the optical density (OD_{600}) was between 0.8 and 1. The inoculated small culture medium was divided equally among three large culture flasks containing 1.2 L of sterile LB broth with 100 $\mu g.ml^{-1}$ ampicillin and grown for 4 hours at 37°C with shaking at 225 rpm until the optical density (OD_{600}) was 1. 1 ml of 1 M IPTG (Isopropyl-β-D-thiogalactopyranoside) purchased from Fisher Scientific

(Dublin, Ireland) was added to each flask and the culture was incubated with shaking for a further time period of 4-5 hours. Finally, the cell culture was centrifuged at 6000 rpm for 8 minutes in 250 ml fluorinated Nalgene bottles to pellet the cells. The supernatant was discarded and the pellets were stored at -80°C for subsequent protein extraction and purification.

2.2.2.10 Protein extraction from cell lysate

The bacterial cell pellets were thawed in a water bath at 30°C. To the thawed cell pellets, 10 ml of cell lysis solution was added. The bacterial cell pellets were re-suspended by gentle vortexing to obtain a homogeneous mixture and left to incubate at room temperature for 2 hours. 160 µl of 50 $mg.ml^{-1}$ lysozyme was then added to the homogeneous cell suspension. The solution was vortexed and incubated for a further 30 minutes. The cell solution was freeze-thawed 4 × times using liquid nitrogen for the freeze and 30°C water bath for the thaw. For the final step 2 ml of 1 $mg.ml^{-1}$ DNase I (Bovine Pancreas) and 1 ml of 1 M $MgSO_4$ (Magnesium sulphate heptahydrate) mixed just prior to use, was added to the cell lysate. This was incubated at room temperature for 30 minutes. The pH of the final cell solution was adjusted to pH 4.5 by the addition of glacial acetic acid. The lysate was centrifuged at 10000 rpm for several hours in a fiberlite rotor attached to a Thermo Scientific HERAEUS MULTIFUGE 3SR+ centrifuge until the cell debris had completely pelleted. The soluble protein solution was decanted from the top of the pellet and filtered through 0.22 µm Millex-Gv Millipore low protein binding durapore (PVDF) membrane filter units (Millipore, Co.Cork, Ireland) before purification by chromatography.

2.2.3 Site-directed mutagenesis, molecular cloning and DNA sequencing

Recombinant HGD prepared by amplification of the coding sequence from a human fetal lens cDNA library were utilized for site-directed mutagenesis and molecular cloning experiments. The forward and reverse oligonucleotide primers used to make the double mutant proteins were synthesized by Life Technologies (Dublin, Ireland). The following forward and reverse primers for the single mutants were obtained:

P23V Mutant

Forward: 5′-GCA GCA GCG ACC ACG TCA ACC TGC AGC CCT AC-3′.
Reverse: 5′-GTA GG CTG CAG GTT GAC GTG GTC GCT GCT GC-3′.

P23T Mutant

Forward: 5′-GCA GCA GCG ACC ACA CCA ACC TGC AGC CC-3′.
Reverse: 5′-GGG CTG CAG GTT GGT GTG GTC GCT GCT GC-3′.

R36S Mutant

Forward: 5′-GCA ACT CGG CGA GCG TGG ACA GCG GC-3′.
Reverse: 5′-GCC GCT GTC CAC GCT CGC CGA GTT GC-3′.

R58H Mutant

Forward: 5′-CCA GTA CTT CCT GCA CCG CGG CGA CTA TGC-3′.
Reverse: 5′-GCA TAG TCG CCG CGG TGC AGG AAG TAC TGG-3′.

Site-directed mutagenesis was performed with QuikChange II site-directed mutagenesis kit from Stratagene (USA) using oligonucleotide primers for the single mutants and HGD wild type plasmid DNA isolated and purified using QIAGEN plasmid purification midi kit (Qiagen GmbH, Hilden, Germany). Based on the concentration of plasmid DNA and oligonucleotide primers that have been utilized to introduce the mutations, table 2.1 illustrates the amounts of different components in the PCR mixture.

Mutants of HGD	**P23V**	**R36S**	**R58H**	**P23VR36S**	**P23TR36S**	**P23VR58H**
Reaction buffer	5µl	5µl	5µl	5µl	5µl	5µl
dS DNA	1.01µl (HGD)	1.01µl (HGD)	1.01µl (HGD)	1.01µl (P23V)	0.94µl (R36S)	1.08µl (R58H)
Oligo #1	0.89µl	1.07µl	0.375µl	1.07µl	0.39µl	0.889µl
Oligo #2	0.92µl	1µl	0.331µl	1µl	0.37µl	0.919µl
dNTP	1µl	1µl	1µl	1µl	1µl	1µl
ddH_2O	41.18µl	40.92µl	42.284µl	39.82µl	42.3µl	41.112µl
Pfu Ultra	1µl	1µl	1µl	1µl	1µl	1µl

Table 2.1: *Composition of PCR mixture for creating double mutants; described in table 2.1 are various components (dS DNA: double stranded DNA, Oligo #1, 2: Oligonucleotides 1, 2, dNTP: deoxynucleotide mix, Pfu Ultra: Pfu Ultra DNA polymerase) that make up the PCR mixture whose individual component composition (volume) is expressed in microlitre (µl).*

Polymerase chain reaction (PCR) generates multiple copies of the DNA material. Following polymerase chain reaction (PCR), the mutated DNA material is transformed

into Library Efficiency® DH5α™ Competent Cells (invitrogen /Life Technologies, Dublin, Ireland) for plasmid DNA extraction and purification. Each of the three double mutant plasmids obtained after mutagenesis were sequenced with the T7 promoter primer by using an automated capillary DNA sequencer (MRCPPU, College of Life Sciences, University of Dundee, Scotland). In order to confirm the desired mutations, the plasmid DNA sequences for the double mutants were aligned and compared against the wild type plasmid DNA using BLAST (Basic Local Alignment Search Tool) available on the NCBI (National Centre for Biotechnology Information, U.S. National Library of Medicine) website.

2.3 Analytical techniques for protein purification and protein purity determination

2.3.1 Size-exclusion chromatography

2.3.1.1 Working Principle

Size-exclusion (also known as gel filtration) chromatography is an analytical technique which separates proteins and other biological macromolecules based on their differences in molecular size (figure 2.1).

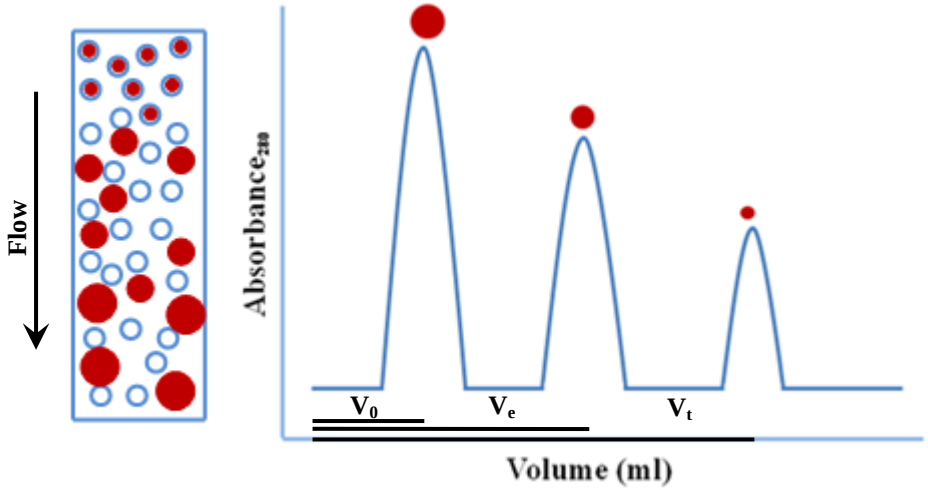

Figure 2.1: *Separation of a sample mixture based on molecular size by size-exclusion chromatography (left). Chromatogram (right) shows the elution profile of molecules depending on their size.*

The stationary solid phase is packed with a porous matrix of spherical particles; the mobile phase fills the pores of the matrix and the space between the particles. The larger molecules that remain exterior to the pores move quickly through the matrix, eluting out at earlier times and the smaller ones that remain trapped in the pore takes

longer to elute through the column (Cutler, 2008). This result in selective partitioning of the various components of a crude mixture based on their size.

Those molecules whose actual size is greater than the pore size of the gel used for separation, elute out together as the first peak in the chromatogram, and this is referred to as the 'total exclusion volume' or void volume (V_0) which defines the exclusion limit for a particular column. Those molecules that can enter the pores of the gel matrix diffuse into the internal pore structure and reside within the pores for a certain amount of time called the 'average residence time' depending on the size of the molecules and the pore size distribution of the gel. This volume region on the chromatogram over which effective separations can occur is referred to as the 'selective permeation' region. Molecules that are smaller than the pore size of the solid matrix are retained for the longest residence time and will elute out of the column as the last peak on the chromatogram in the total volume fraction (V_t), which determines the 'total permeation' limit for a particular column. The largest elution or retention volume for any size-exclusion column is the total mobile-phase in the column also known as void volume (V_0) (Kazakevich & Lobrutto, 2007). The exclusion range comprises of those molecules with molecular weights greater than the exclusion limit as shown in figure 2.2.

A given protein can elute between V_0 and V_t, designated by the elution volume V_e which is related to the partition coefficient K_{av} as given by equation 2.2.

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \qquad 2.2$$

The partition coefficient, K_{av} shares a log-linear relationship with the hydrodynamic radius or radius of gyration of a molecule, which is essentially the effective size of the molecule.

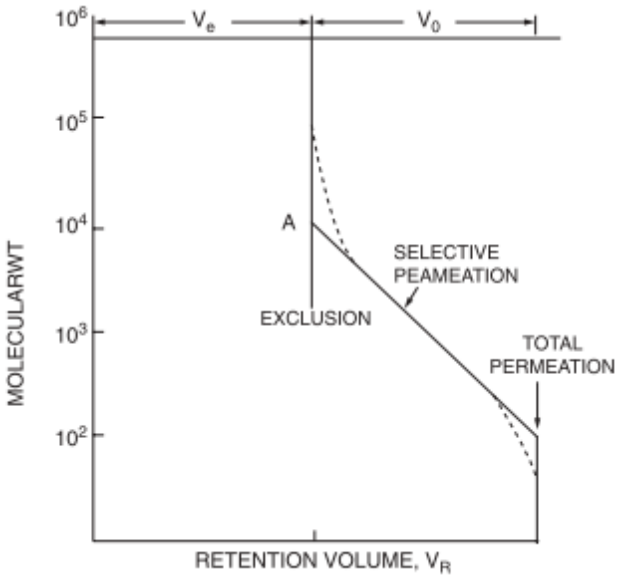

Figure 2.2: *Elution of molecules in size-exclusion chromatography, taken from (Kazakevich & Lobrutto, 2007) .*

2.3.1.2 Instrumentation

Size-exclusion chromatography was performed using an XK50 column connected to an external AKTAprime plus system (GE Healthcare BioSciences AB, Sweden) run using the PrimeView 5.0 software. The XK50 column was packed with Sephacryl S-200 HR (GE Healthcare BioSciences AB, Sweden) which is a composite medium prepared by covalently cross-linking allyl dextran with N,N'-methylene bisacrylamide to form a hydrophilic matrix of high mechanical strength. This packing medium is ideal for separating globular proteins in the range between 5 kDa and 250 kDa. The column matrix which forms the stationary phase was equilibrated with two bed volumes of 0.275 M sodium acetate buffer (pH 4.5) pumped at a flow rate of 2 $ml.min^{-1}$, controlled by the AKTA prime plus system (GE Healthcare) set to a maximum pressure limit of 1 MPa. The proteins were detected by monitoring their absorbance profile at 280nm, an extinction coefficient value of 2.09 mg^{-1} ml cm^{-1} was used here for protein concentration measurements.

2.3.1.3 Protein sample preparation and elution

The clear supernatant containing the crude protein obtained after centrifuging the cell lysate was filtered sterilized through 0.22 μm Millex-Gv Millipore low protein binding durapore (PVDF) membrane filter units (Millipore, Co.Cork, Ireland) before loading on to the column. The sample was loaded by manual run method where the flow rate was set to 2 $ml.min^{-1}$ on the AKTA prime plus system.

The protein sample fractions were isocratically eluted by running 0.275 M sodium acetate buffer (pH 4.5) at a flow rate of 2 $ml.min^{-1}$ using a programmed method (table 2.2) and collected in long borosilicate glass test tubes on a fraction collector.

Breakpoint	Volume (ml)	Flow rate $ml.min^{-1}$	Fraction (ml)
1	0	2	0
2	400	2	0
3	2300	2	20
4	2300	2	0

Table 2.2: *Programmed method for separation of native HGD and the double mutant proteins from crude mixtures by Size-exclusion chromatography.*

2.3.2 Ion exchange chromatography

2.3.2.1 Working Principle

Ion-exchange chromatography is a process for separating proteins and biological molecules in a sample solution based on their charge characteristics. The charged groups on the surface of a protein interact with the oppositely charged groups immobilized on the ion-exchange medium (Williams & Frasca, 2001). Negatively charged groups bind to positively charged medium and positively charged groups bind to negatively charged medium.

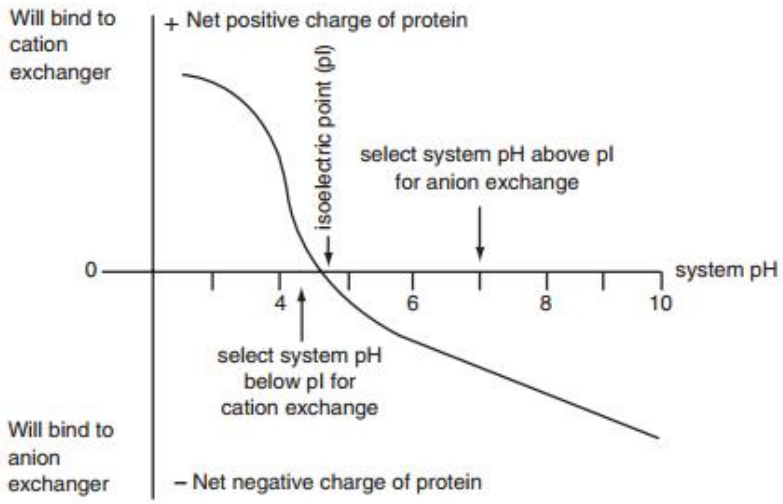

Figure 2.3: *Diagram showing the binding affinity of the protein to anion or cation exchangers based on net charge of the protein as a function of operating buffer pH, taken from (Williams & Frasca, 2001).*

The net charge of a protein depends on the operating buffer pH (figure 2.3). The pH at which the net charge of a protein is zero (*i.e.* the number of positive charges balances the number of negative charges) is referred to as iso-electric point (pI). When the operating buffer pH is less than the pI of the protein, the protein will have a net positive charge and will bind to the cation-exchange medium, whereas when the operating buffer pH is greater than the pI of the protein, the protein will have a net negative charge and will bind to the anion-exchange medium (figure 2.3). This result in selective partitioning of the molecules in a protein mixture based on their surface charge differences. Ion-exchange chromatography is performed by adsorption of protein to an ion-exchange medium followed by elution using increasing salt gradient or using a buffer with a different pH.

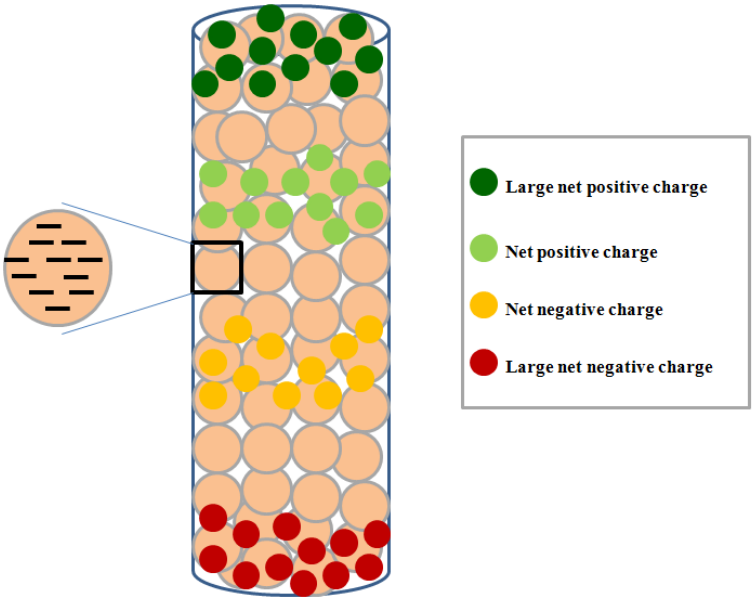

Figure 2.4: *Separation of sample mixtures based on their surface charge by cation-exchange chromatography.*

2.3.2.2 Instrumentation and technique

Cation-exchange chromatography was performed using an XK50 column connected to an external AKTA prime plus system (GE Healthcare BioSciences AB, Sweden) run using the PrimeView 5.0 software. The XK50 column was packed with SP sepharose fast flow (GE Healthcare BioSciences AB, Sweden) which is a strong cation exchanger composed of crosslinked 6% agarose beads, sulphopropyl (Q) with strong cation exchange groups. The column matrix comprises of spherical particles substituted with ionic groups that are either negatively or positively charged. The highly porous matrix which is packed into the column was then equilibrated with at

least five bed volumes of 0.275 M sodium acetate buffer (pH 4.5) pumped at a flow rate of 25 ml.min^{-1} controlled by the AKTAprime plus system set to a maximum pressure limit of 1 MPa. The equilibration buffer fills the pores of the matrix and the space between the spherical particles. The pH and ionic strength of the equilibration buffer are selected to ensure binding of the proteins of interest to the medium and thorough elution of other impurities. The proteins were detected by monitoring their absorbance profile at 280nm.

2.3.2.3 Protein sample preparation and elution

The protein fractions from the size-exclusion chromatographic separation were pooled and applied to the pre-equilibrated cation-exchange column at a flow rate of 15 ml.min^{-1}. The proteins that bind are concentrated onto the column and the ones that possess the same surface charge as of the ionic groups on the column are eluted out along with the buffer during sample loading or washing. The wash step was carried out by applying five bed volumes of 0.275 M sodium acetate buffer (pH 4.5), at a flow rate of 25 ml.min^{-1} to remove unbound protein. When all the unbound protein has passed through the column during the wash, a programmed method (table 2.3 for native HGD) and (table 2.4 for the double mutant proteins) with a flow rate of 10 ml.min^{-1} and an increasing salt gradient of 0 to 0.325 M sodium chloride (NaCl) was set to collect the bound proteins eluting out in 0.275 M sodium acetate elution buffer (pH 4.8) and into long borosilicate glass test tubes on a fraction collector. The salt ions (Na^+ and Cl^-) compete with the bound proteins for the cross-linked resin, the proteins with the lowest net charge at a selected pH will elute out first. Proteins with the highest net charge will be most strongly bound to the medium and elute out at the end.

Breakpoint	Volume(ml)	Salt (%)	Flow rate (ml.min^{-1})	Fraction (ml)
1	0	15	10	0
2	1379	15	10	0
3	1380	30	10	22
4	2904	43	10	22
5	3844	85	10	22

Table 2.3: *Programmed method for separation of native HGD from protein mixtures by cation-exchange chromatography. A linear gradient elution against increasing salt concentration is used to elute the proteins of interest.*

Breakpoint	Volume(ml)	Salt (%)	Flow rate (ml.min-1)	Fraction (ml)
1	0	15	10	0
2	614	15	10	0
3	615	30	10	23
4	2800	43	10	23
5	3844	85	10	23

Table 2.4: *Programmed method for separation of HGD double mutant proteins from protein mixtures by cation-exchange chromatography. A linear gradient elution against increasing salt concentration is used to elute the proteins of interest.*

After each chromatographic separation, any irreversibly bound material was washed off the column with 1 M NaOH in Milli-Q water at a flow rate of 20 ml min^{-1} and the medium was regenerated for the next chromatographic separation by washing with at least five bed volumes of 0.275 M sodium acetate buffer (pH 4.5).

2.3.3 Polyacrylamide gel electrophoresis

2.3.3.1 Working principle

Electrophoresis is an analytical technique used for qualitative characterisation of biological molecules including proteins, peptides, sugars and nucleic acids. The charged molecules migrate under the influence of an electric field towards the electrode carrying the opposite charge. The mobility of different molecules in a mixture depends on their mass allowing them to migrate at different velocities, thus separating them into well defined fractions (Westermeier, 2005). The migration velocity of a charged molecule, such as protein depends on its pI (iso-electric point) and on the pH and ionic strength of the buffer. The pI for a protein is constant and is influenced by the composition of amino acids and carbohydrates. The migration of protein molecules in an electric field is measured and expressed in terms of its electrophoretic mobility (μ),

$$\mu = \frac{d}{Et} \qquad 2.3$$

where *d* is the distance travelled from the origin (cm), *E* is the strength of the electric field (V/cm) and *t* is the duration of electrophoresis (s) (Keren, 2003).

2.3.3.2 SDS-PAGE

SDS-PAGE is an electrophoresis method for separating proteins based on their molecular weights. SDS is an anionic detergent that binds to the protein at a constant weight ratio of 1.4 g SDS/1g protein forming an SDS-protein complex, masking the intrinsic charge of the protein and giving it an overall negative charge density. Besides SDS, the addition of 2-mercaptoethanol or dithiothreitol (DTT) to the sample buffer breaks down the disulphide bridges between cysteine residues on the protein. This type of gel is commonly referred to as reducing SDS-PAGE.

Protocol for SDS-PAGE

Preparation of running gel (12.6%):

Composition	Quantity
Milli-Q water	3.15 ml
1M Tris-HCl pH 8.8	2.5 ml
10% (w/v) SDS	100 µl
30% Acrylamide/Bis solution (37:5:1)	4.2 ml

The following components were added to the above mixture after degassing for about 15 minutes.

10% ammonium persulfate (prepared fresh)	50 µl
TEMED	5 µl

Preparation of stacking gel (4%):

Composition	Quantity
Milli-Q water	3.05 ml
0.5M Tris-HCl pH 6.8	1.25 ml
10% (w/v) SDS	50 µl
30% Acrylamide/Bis solution (37:5:1)	0.65 ml

The following components were added to the above mixture after degassing for about 15 minutes.

10% ammonium persulfate (prepared fresh)	50 µl
TEMED	5 µl

Preparation of sample buffer:

Composition	Quantity
Milli-Q water	4 ml
0.5M Tris-HCl pH 6.8	1 ml
Glycerol	0.8 ml
10% (w/v) SDS	1.6 ml
Bromophenol blue	0.5 mg

Preparation of protein molecular weight markers:

Combined low range molecular weight protein standards (Bio-Rad, Dublin) with SDS reducing sample buffer in the ratio 1:20. The mixture was heated for 5 minutes at 95°C, cooled and loaded 10 µl/well for full length wells.

Stock sample buffer (stored at room temperature)

Composition	Quantity
Milli-Q water	4.8 ml
0.5M Tris-HCl pH 6.8	1.2 ml
Glycerol	1.0 ml
10% (w/v) SDS	2.0 ml
0.1% (w/v) Bromophenol blue	0.5 ml

SDS reducing sample buffer (prepared immediately before use)

2-mercaptoethanol	25 µl
Stock sample buffer	475 µl

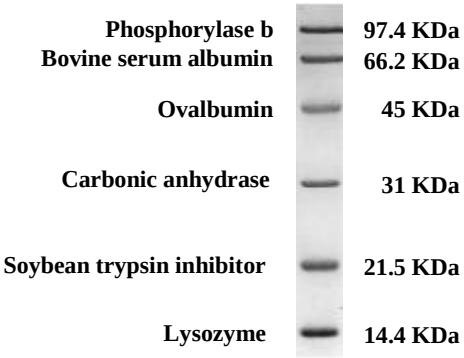

Figure 2.5: *SDS-PAGE low molecular weight markers run on 12% gel, stained with Coomassie R-250.*

Preparation of protein samples

Protein samples were combined with the sample buffer in a 1:1 ratio and boiled for 5 minutes to linearize the proteins. In the case of reducing SDS-PAGE, protein samples were combined with pre-mixed sample buffer and 2-mercaptoethanol. Linearized protein at a concentration of 0.1 µg .µl^{-1} was loaded into the wells at 30 µl/well.

Casting the gels

The gels were run using a mini-protean tetra cell (Bio-Rad, Ireland) assembly. A clean pair of spacer and short plates were assembled on a casting stand and secured firmly using a casting frame. 10% ammonium persulfate solution and TEMED were added to the thoroughly degassed running gel (12.6%) to initiate polymerisation. The combined running gel mixture was then dispensed into the space between the clamped plates using pasteur pipettes up to 2/3 height. When the gel started to set, tert-amyl alcohol (2-methyl-2-butanol) was dispensed on the surface of the gel mixture to prevent oxidation of polyacrylamide. The running gel was left to set for about 35-40 minutes. When the gel was completely set, the tert-amyl alcohol was washed away with milli-Q water and the remaining water on the surface of the gel blotted with a Kimwipe. The stacking gel mixture was then poured and allowed to set with the combs inserted to form wells. The prepared protein samples and molecular weight markers were loaded into the wells by gentle pipetting.

Running the gels

The loaded gels were run at 200 V (400 mA) until the band was observed visually at a height of 1 cm above the end of the glass plates.

Staining and de-staining the gels

The gels were stained in Coomassie Brilliant Blue R-250 staining solution purchased from Bio-Rad (Ireland) for 2-5 hours and then destained using the destaining solution. The molecular weight of the unknown protein was determined by comparing the migration of the protein band against the molecular weight standards.

2.3.4 High Performance Liquid Chromatography

2.3.4.1 Working principle/technique

High performance liquid chromatography (HPLC) is an analytical technique which separates various components in a mixture by forcing the mobile phase under high pressure through the stationary phase of the column. Size-exclusion-high performance liquid chromatography (SE-HPLC) separates molecules based on their molecular size. Larger molecules will swiftly pass through the stationary phase of the column eluting out first and smaller molecules will take longer to pass through the stationary phase

thus eluting out later. Each specific molecule is represented as a peak on the chromatogram; the peaks are usually symmetrical and resemble a normal Gaussian distribution curve. The distance of the peak maxima from the injection point expressed in time units is referred to as the retention time (t_R) (Kazakevich & Lobrutto, 2007) and this time describes the characteristic behaviour for individual molecules depending on their molecular size. Analyte retention volume (V_R) is related to the analyte retention time (t_R) as shown by equation 2.4, provided the HPLC system provides a stable mobile phase flow (F).

$$t_R = \frac{V_R}{F} \qquad 2.4$$

Retention factor is another chromatographic descriptor which relates retention time (t_R) to the retention volume (V_R) by taking into account the void time (t_0) and the void volume (V_0). The difference between the total retention time (t_R) and the void time (t_0) is called the reduced retention time (t_R'), similarly the difference between total retention volume (V_R) and the void volume (V_0) is referred to as reduced retention volume (V_R'). The ratio of the reduced retention time (t_R') to the void time (t_0) or the ratio of the reduced retention volume (V_R') to the void volume (V_0) is referred to as retention factor (k) as shown by equation 2.5 (Kazakevich & Lobrutto, 2007).

$$k = \frac{V_R - V_0}{V_0} = \frac{V_R'}{V_0} = \frac{t_R - t_0}{t_0} = \frac{t_R'}{t_0} \qquad 2.5$$

Thus, retention factor allows valid comparison of the analyte results obtained using different column dimensions and mobile phase flow rate.

2.3.4.2 Instrumentation

SE-HPLC was performed either using a Superdex 200 10/300 GL or Superdex 75 10/300 GL SE-HPLC column (GE Healthcare) connected to a Shimadzu SPD HPLC system run using Shimadzu LC solution software. The protein samples were introduced to the column using the instrument autosampler and detected using a diode array detector at 190-800nm. The packing medium for Superdex 200 10/300 GL or Superdex 75 10/300 GL SE-HPLC column is a composite mixture of cross-linked agarose and dextran. Superdex 200 10/300 GL is ideal for separating proteins in the

range between 1000 and 600000 Da and Superdex 75 10/300 GL is ideal for the range between 3000 and 70000 Da.

2.3.4.3 Sample preparation and loading

SE-HPLC samples for HGD and the double mutant proteins were prepared in 0.1 M sodium phosphate (pH 7.0). Prior to loading, 20mM DTT was added to the HGD and double mutant proteins to reduce crosslinks formed due to disulfide bridges between cysteine residues on the protein surface (Cys-110) (Pande *et al.*, 2000).

The freshly prepared protein samples with 20mM DTT was forced through either a Superdex 200 10/300 GL or a Superdex 75 10/300 GL SE-HPLC column (GE Healthcare) and the output which is a size-exclusion chromatogram correlates the measured retention time with the molecular weight for the protein. The protein samples were initially filter sterilized using 0.22 μm Millex-Gv Millipore low protein binding durapore (PVDF) membrane filter units (Millipore, Co.Cork, Ireland). The filtered samples were placed in Chromacol borosilicate glass vials (Fisher Scientific, Ireland) with a crimp cap and pre-fitted seal and introduced to the column from the instrument autosampler. The following specifications were chosen to run a SE-HPLC method on Superdex 200 10/300 GL column:

Parameters	**Specifications**
Mobile phase	0.1 M sodium phosphate (pH 7.0)
Flow rate	0.75 $ml.min^{-1}$
Run time	60 minutes
Sample injection volume	40 μl
Concentration of sample	0.1 $mg.ml^{-1}$-1$mg.ml^{-1}$
Pressure	5-10 bar

The specific conditions chosen to run a SE-HPLC method on Superdex 75 10/300 GL are detailed as follows:

Parameters	**Specifications**
Mobile phase	0.1 M sodium phosphate (pH 7.0)
Flow rate	0.15 $ml.min^{-1}$
Run time	30 minutes
Sample injection volume	40 μl
Concentration of sample	0.2 $mg.ml^{-1}$-1$mg.ml^{-1}$
Pressure	5-10 bar

2.3.4.4 Calibration of the column

Proteins samples for the calibration plot were prepared by dissolving 1mg.ml^{-1} of each protein in 0.1M sodium phosphate (pH 7.0), filter sterilized using 0.22 μm Millex-Gv Millipore low protein binding durapore (PVDF) membrane filter units (Millipore, Co.Cork, Ireland) and loaded on both Superdex 200 10/300 GL (flow rate of 0.75 ml.min^{-1}) and Superdex 75 10/300 GL (flow rate of 0.15 ml.min^{-1}) SE-HPLC columns (GE Healthcare). All proteins were run separately on both the columns and from the retention times obtained for each protein, a characteristic calibration plot was drawn for Superdex 200 10/300 GL (figure 2.6) and Superdex 75 10/300 GL (figure 2.7).

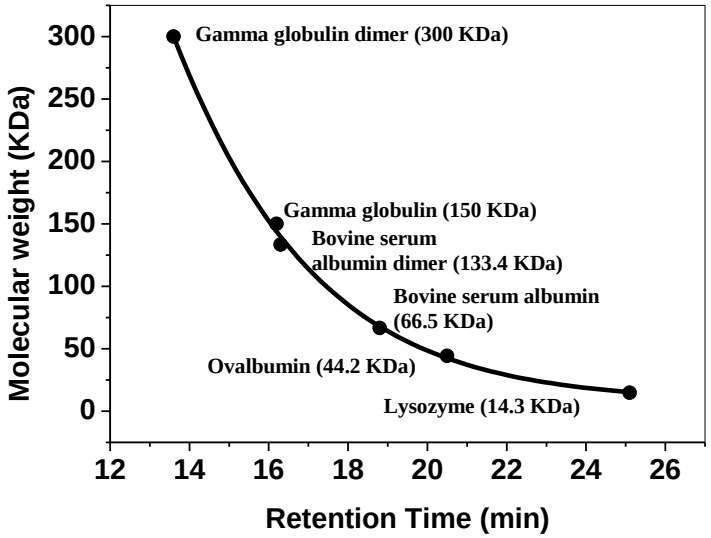

Figure 2.6: *Calibration plot for Superdex 200 10/300 GL SE-HPLC column using a range of different molecular weight proteins.*

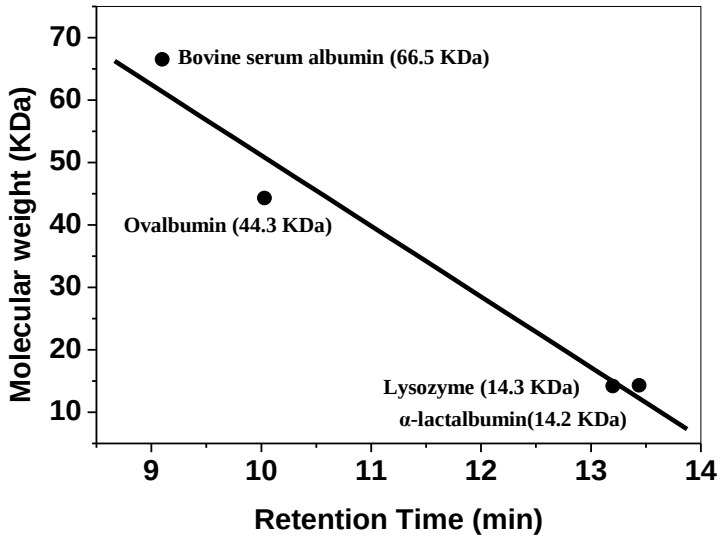

Figure 2.7: *Calibration plot for Superdex 75 10/300 GL SE-HPLC column using a range of different molecular weight proteins.*

2.4 Spectroscopic techniques for protein characterization

2.4.1 Introduction to spectroscopy

Spectroscopy is a powerful technique that measures and interprets the spectra arising from the interaction of matter with electromagnetic radiation. This has led to the advent of using spectroscopy for analysing the individual components of complex biological molecules such as proteins and nucleic acids. Electromagnetic radiation in the form of light is characterised by its energy, *E*, which is linked to frequency, *v*, or wavelength, λ, of the radiation through Planck relationship:

$$E = h\nu = h\frac{c}{\lambda} \qquad 2.6$$

where *c* is the speed of light in vacuum (c = 2.998x10^{8} m/s) and *h* is the Planck's constant (h = 6.625x10^{-34} J-s).

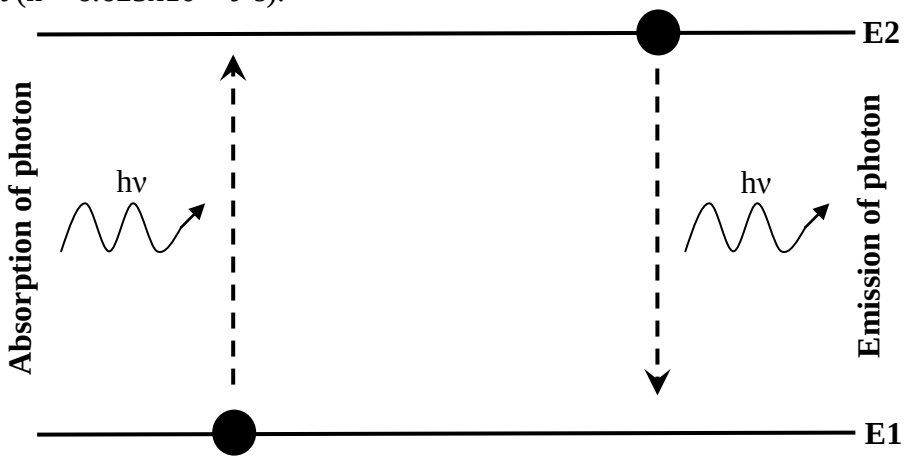

Figure 2.8: *Energy transition due to the absorption of electromagnetic radiation.*

Radiation is an electromagnetic wave with perpendicularly oscillating electric and magnetic components. The fundamental process involved in molecular spectroscopy is the absorption of a discrete amount of energy in the form of photons to initiate the transition of a charged particle from a lower energy state (E1) to a higher energy state (E2) (figure 2.8). The excited molecule reverts back to the lower energy state by the emission of energy equal to or less than the absorbed radiation referred to as fluorescence which will be discussed in section 2.4.1.3. The electrons in the molecule are displaced to higher energy levels if the electro-magnetic field intensity is sufficiently large and the time taken for the absorption process is about 10^{-15} s

(Hammes, 2005). An absorption spectrum is obtained when the absorption of light is measured as a function of its frequency or wavelength.

2.4.1.1 UV-Visible absorption spectroscopy: Working principle

UV-Visible absorption spectroscopy uses the general principle of molecular spectroscopy, *i.e.* when uv-visible light passes through a sample, energy from the radiation promotes the electrons from a bonding or non-bonding orbital to an anti-bonding orbital as illustrated in figure 2.9. UV-Visible spectroscopy uses electromagnetic radiation that falls in the wavelength range between 190-380nm (ultraviolet region) and 380-780nm (visible region) (Schmid, 2001). The energy gained by a molecule due to absorption of radiation also referred to as its potential energy is the sum of its electronic, vibrational and rotational energies and can be represented as a series of discrete levels or states.

$$E_{\text{total}} = E_{\text{electronic}} + E_{\text{vibrational}} + E_{\text{rotational}} + E_{\text{translational}} \qquad 2.7$$

The differences in energy among the different states follow the order:

$$E_{\text{electronic}} > E_{\text{vibrational}} > E_{\text{rotational}} \qquad 2.8$$

Energy absorbed by a molecule in the UV-Visible region produces changes in the electronic energy levels of the molecule. Electromagnetic radiations of longer wavelengths in the infrared region of the spectrum are required to bring about vibrational and rotational energy changes to the molecules. However in practice, the absorption spectrum in the UV-Visible region would consist of a larger electronic energy level transition which is usually accompanied by simultaneous smaller changes between the vibrational and rotational levels associated with the molecule (Primer, 1996).

Electronic transition is an event triggered by the absorption of a photon due to which an electron is promoted from an occupied orbital (usually a bonding σ (sigma), π (pi) or non-bonding n orbital) of the molecule in its ground state to an unoccupied orbital (an antibonding orbital σ^* or π^*) in an excited state. For most molecules, the lowest energy occupied molecular orbitals are σ orbitals, formed either from two s atomic orbitals, or from one s and one p atomic orbital, or from two p atomic orbitals having a collinear axis of symmetry. The π orbitals, formed from two p atomic orbitals overlapping laterally lie at relatively higher energy levels than σ orbitals and the non-

bonding orbitals that hold unshared electron pairs lie at even higher energy levels. The antibonding orbitals (σ^* and π^*) are orbitals of highest energy (Valeur & Berberan-Santos, 2012).

The following electronic transitions are possible as shown in the energy diagram below:

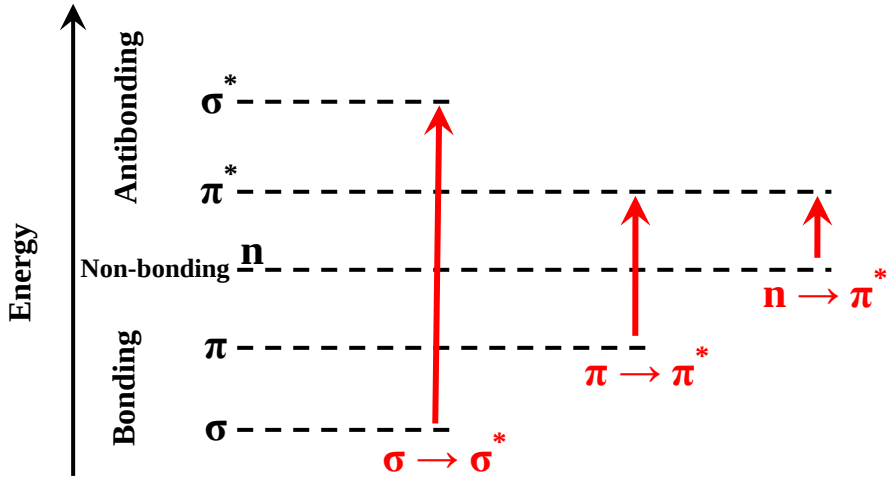

Figure 2.9: *Hypothetical energy diagram showing electronic transitions among various orbitals.*

Absorption of photon of appropriate energy can promote one of the π electrons to an anti-bonding π^* orbital and this type of transition is called $\pi \rightarrow \pi^*$ transition, the promotion of an anti-bonding electron to an anti-bonding π^* orbital results in $n \rightarrow \pi^*$ transition. The σ to σ^* transition requires absorption of photon with a wavelength which does not fall in the UV-Visible range, thus, only π to π^* and n to π^* occur in the UV-Visible region.

The energy of electronic transitions is generally in the following order:

$$n \rightarrow \pi^* < \pi \rightarrow \pi^* < n \rightarrow \sigma^* < \sigma \rightarrow \pi^* < \sigma \rightarrow \sigma^* \qquad 2.9$$

Electronic transitions arise in proteins due to the presence of the chromophores of the aromatic amino acids (phenylalanine, tyrosine and tryptophan), peptide bonds and disulphide bonds. Electronic transitions in tyrosine and tryptophan are primarily responsible for the absorption spectrum in the region of 280nm. The contribution from the aromatic amino acid phenylalanine is more pronounced at 260nm (figure 2.10). A strong absorbance profile centred at ~190nm indicates electronic transitions associated with the peptide bond. Thus the absorbance maxima in the 190-210nm range is attributed to the presence of peptide bonds. The absorption spectra of proteins are not only dependant on the aromatic amino acids and peptide bonds but also strongly

influenced by their local molecular environments. The structural characteristics of proteins that determine the local charge distribution and the local dielectric constant also influence the absorption spectrum of proteins. The presence of hydrogen bonding and ion-pair interactions also influence the spectrum. This environmental sensitivity of protein spectra can be used to obtain detailed information about the amount of α-helix, β-sheet and random coil configurations of proteins (Hammes, 2007).

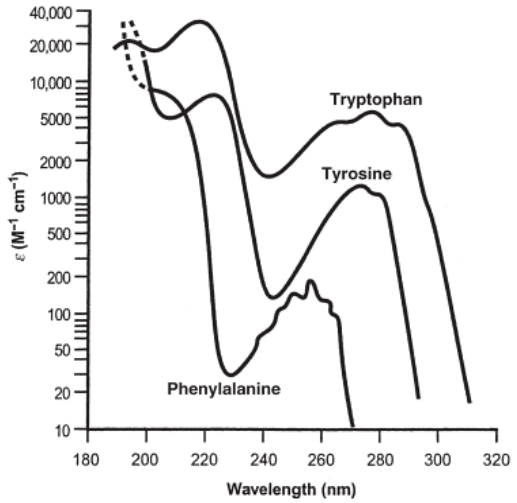

Figure 2.10: *Absorption spectra of phenylalanine, tyrosine and tryptophan at pH 6.0, taken from (Hammes, 2007).*

2.4.1.2 Protein quantification using UV absorption spectroscopy

Protein concentrations for native HGD and the double mutants were determined spectrophotometrically by UV absorbance using the extinction coefficient value of 2.09 mg^{-1} ml cm^{-1} (Pande *et al.*, 2001). According to Beer-Lambert law,

$$A = \varepsilon c l \qquad 2.10$$

where A is the absorbance at 280nm, ε is the extinction co-efficient of the protein in $mg^{-1}.ml.cm^{-1}$, c is the concentration of the protein in $mg.ml^{-1}$ and l is the path length of light in cm.

2.4.1.3 Fluorescence spectroscopy: Working principle

The term fluorescence refers to the emission of light associated with electrons, which in the process of re-arrangement is displaced from an excited state to the ground state. Fluorescence is a very useful tool in studying biological processes as it is very sensitive to the chemical nature of the molecule and the local environment surrounding the molecule. Fluorescence measurements are more complex than absorbance

measurements because the molecules are excited at a specific wavelength and fluorescence is observed at a higher wavelength. Fluorescence spectral data is generally presented as an emission spectrum. A fluorescence emission spectrum is a plot of fluorescence intensity vs wavelength. Emission spectra are sensitive to the chemical structure of the fluorophore and the solvent conditions. Fluorophores belong to two general classes – intrinsic and extrinsic. Intrinsic fluorophores occur naturally, whereas extrinsic fluorophores are added to impart spectral properties to a non-fluorescent molecule. Proteins belong to the class of intrinsic fluorophores; the dominant fluorophore in proteins is the indole group of tryptophan. Indole absorbs near 280nm and emits near 340nm. The emission spectrum of indole is also highly sensitive to solvent polarity. The emission of indole is blue shifted if the group is buried within the native protein and its emission may shift to higher wavelengths (red shift) when the group is exposed as in an unfolded protein (Lakowicz, 2007).

2.4.1.4 Protein intrinsic fluorescence measurements by fluorescence spectrometer

Intrinsic fluorescence measurements of proteins were carried out using Molecular Devices SpectraMax M2e (Molecular Devices, USA). Protein samples were placed in Hellma Analytics micro fluorescent quartz cells, and excited at 280nm (excitation of tyrosine and tryptophan residues) and 295nm (excitation of tryptophan residues only). Fluorescence emission spectra were recorded over the wavelength range from 250 to 450nm with a data interval of 1nm. Buffer samples were used as reference and were subtracted from the intrinsic fluorescence measurements of the protein samples.

2.4.1.5 Circular dichroism spectroscopy: Working principle

Molecular asymmetry is a general feature of biological macromolecules. Such molecules whose mirror images are not identical are referred to as chiral molecules. Chiral molecules are distinguished from achiral molecules by their interactions with polarized light. Plane polarized light exhibits unidirectional wave propagations. It can be decomposed into two circularly polarized components of equal magnitude, one rotating in the clock-wise direction (right-handed, R) and the other in the counter-clockwise direction (left-handed, L) as shown in figure 2.11.

The two types of measurements that are commonly used to determine the effect of molecules on polarized light are optical rotation and circular dichroism (CD). Optical rotation is a measure of the rotation of linearly polarized light by a molecule, whereas

circular dichroism is a measure of the differential absorption of the left-handed and right-handed circularly polarized light by molecules (Hammes, 2007). When a chiral molecule interacts with plane polarized light, the plane is rotated with the amount of rotation and direction depending on the molecular asymmetry. If the L and R components of the plane polarized light are either unabsorbed or absorbed to equal extents, the recombination regenerates the original plane polarized light. However, if they are absorbed to different extents, the resulting radiation would possess elliptical polarization as shown in figure 2.11 (Kelly *et al.*, 2005).

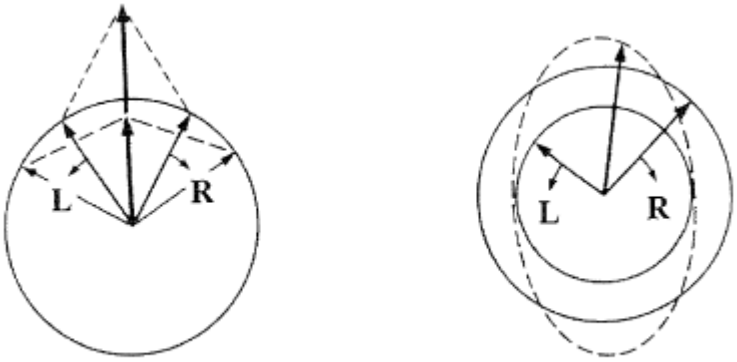

Figure 2.11: *The left (L) and right (R) circularly polarized components of plane polarized light. When the two components have the same amplitude, they combine and generate plane polarized light (left), while different magnitudes combine and result in elliptically polarized light (right), taken from (Kelly et al., 2005).*

Circular dichroism is an excellent tool for rapid determination of secondary structures and folding properties of recombinant proteins. It has been widely used to study protein interactions, determine whether a recombinant protein has been folded correctly or if mutation has affected its structural conformation (Kelly *et al.*, 2005). CD measurements are recorded by spectropolarimeters and are generally reported in terms of ellipticity (θ) in degrees which is a measure of the differential absorption of the L and R circularly polarized components. CD spectral bands only arise where absorption of radiation occurs and they are assigned to explain distinct structural features of a molecule. In proteins, the presence of chromophores such as aromatic amino acids (absorption in the range from 260 to 320nm), peptide bonds (absorption below 240nm) and disulphide bonds (weak absorption centred at 260nm) generally give rise to characteristic spectral features. These spectral features contain detailed information on the secondary and tertiary structural composition of proteins (Kelly *et al.*, 2005, Greenfield, 2006).

The peptide bonds in proteins provide information on the secondary structural components such as % helix, sheet, and turns. This is principally determined from absorption of radiation due to peptide bonds in the region below 240nm. Absorption of radiation below 240nm results in a weak but broad $n \rightarrow \pi^*$ transition centred at 220nm and a more intense $\pi \rightarrow \pi^*$ transition centred at 190nm.

Aromatic amino acids may also contribute significantly to the spectral data in this region. The different secondary structural components of proteins that give rise to characteristic CD spectral data in the far-UV region are illustrated in figure 2.12 (Kelly *et al.*, 2005).

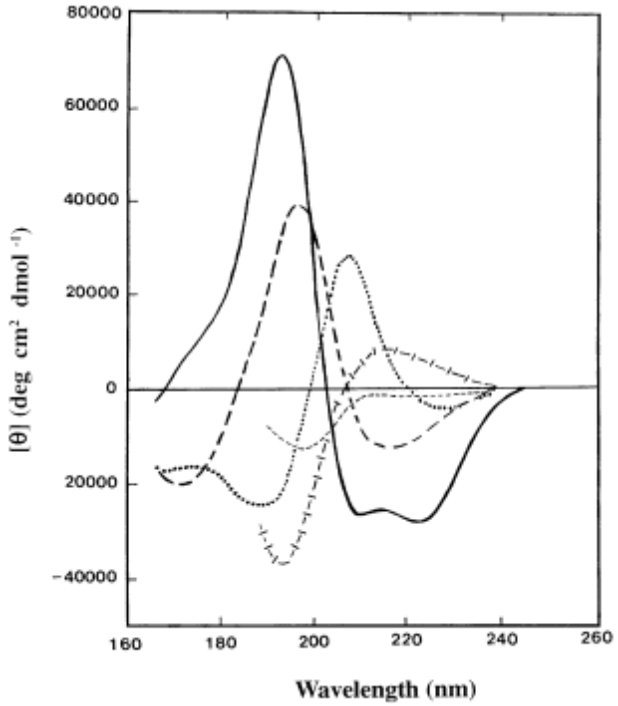

Figure 2.12: *Far-UV CD spectra associated with various secondary structural components. Solid line represents α-helix; long dashed, anti-parallel β-sheet; dotted line, type I β-turn; cross dashed line, 3_1-helix or poly (Pro) II helix; short dashed line, irregular structure, taken from (Kelly et al., 2005).*

The near-UV CD spectrum that features the tertiary structure of proteins principally absorbs in the 260-320nm region due to the aromatic amino acid side chains. Each of the amino acids has a characteristic wavelength profile. Phenylalanine shows weaker but sharper bands with fine structure between 255 and 270nm, tyrosine shows a peak between 275 and 282nm and tryptophan shows a peak close to 290nm with fine structure between 290 and 305nm. It has also been observed that the spectral bands that arise from tyrosine may overlap with tryptophan (Kelly *et al.*, 2005).

The factors that contribute to the shape and magnitude of the near-UV CD spectrum are: the number of each type of aromatic amino acid present, their mobility, the nature of their environment (H-bonding, polar groups and polarisability) and their spatial arrangement in the protein. Although, near-UV CD spectra is not sufficiently advanced to yield significant structural insights but it is a reliable way of analysing protein structural changes due to amino acid substitutions by site-directed mutagenesis. Thus, the near-UV CD spectrum of a protein provides a valuable fingerprint of the tertiary structure of a protein, that can be used to compare the native and mutant forms of the same protein (Kelly *et al.*, 2005).

2.4.1.6 Protein sample preparation for CD spectroscopic measurements

Proteins for CD measurements were prepared at a concentration of 0.2 mg ml^{-1} in 0.01 M sodium phosphate (pH 7.0) for both far-UV and near-UV CD spectroscopic measurements. Far-UV and near-UV CD spectra were recorded on a JASCO J-810 spectropolarimeter (Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland, UK). 0.01 M sodium phosphate buffer (pH 7.0) was used as reference. CD spectral data processing required baseline correction for the CD spectra of the protein samples. CD spectral data for native HGD and the double mutant proteins were drawn and analysed using Origin v6.1 software.

2.5 Protein sample preparation and liquid-liquid phase separation (co-existence curve) measurements

Protein solutions for liquid-liquid phase separation (co-existence curve) measurements were prepared by diafiltration against 0.1M sodium phosphate buffer containing 20mM DTT (Dithiothreitol), pH 7.0 using Ultracel 10 kDa ultrafiltration discs (Merck Millipore, Co.Cork, Ireland). Prior to phase separation measurements, protein solutions were concentrated by ultrafiltration using Amicon Ultra-4 centrifugal filter units (Merck Millipore, Co.Cork, Ireland). Alternatively, to achieve high protein concentrations, phase separation of concentrated protein solutions were carried out. Protein samples concentrated to 100 $mg.ml^{-1}$ by ultrafiltration were placed in an eppendorf tube and subjected to centrifugation at speeds of 1500-2000 rpm for more than an hour in a Thermo Scientific HERAEUS MULTIFUGE 3SR+ centrifuge cooled to temperatures below the critical temperature for the protein (T_c). This resulted in

protein phase separation and the subsequently formed protein-rich phase was collected for phase separation measurements.

The co-existence curve for native HGD was measured using the same method as described previously (Broide *et al.*, 1991, Asherie, 2004). Protein sample of known concentration was pipetted into a cuvette and placed in Perkin Elmer Lambda 35 UV/Vis spectrophotometer operated using UV Win Lab-Timedrive Lambda 35 software. The percent transmission of the sample was recorded at 600nm. The temperature of the system was controlled by an external thermostatized water bath unit (Thermo Scientific DC10-K10) and monitored using an Omega HH509R thermocouple. As the protein solution is cooled, the temperature at which the intensity of transmitted light falls to half of its initial value is defined as the clouding temperature (T_{cloud}). Once the transmitted intensity reaches zero the protein solution is warmed and the temperature at which the intensity of transmitted light returns to half of its initial value is defined as the clearing temperature (T_{clear}). The phase separation (T_{ph}) temperatures are calculated as the average of T_{cloud} and T_{clear}. The phase separation temperatures (T_{ph}) plotted as a function of protein concentrations defines the co-existence curve. Co-existence curves were drawn using Origin v6.1 software.

2.6 Protein crystal/aggregate formation and solubility measurements

2.6.1 Protein crystal growth

Protein solutions were initially prepared by diafiltration against 0.1M sodium phosphate buffer containing 20mM DTT (Dithiothreitol), pH 7.0 using Ultracel 10 KDa ultrafiltration discs (Merck Millipore, Co.Cork, Ireland). Protein solutions for crystal growth were concentrated by ultrafiltration using Amicon Ultra-4 centrifugal filter units (Merck Millipore, Co.Cork, Ireland). Native HGD protein crystals were grown in 5mm NMR glass tubes (Hilgenberg GmbH) from protein solutions at a concentration of 50mg.ml^{-1} in the presence of 20µl of 2-methyl-2,4 pentanediol (MPD) (Asherie, 2012). Double mutant protein crystals/aggregates were also grown in 5mm NMR glass tubes (Hilgenberg GmbH) from protein solutions in 0.1M sodium phosphate buffer containing 20mM DTT (Dithiothreitol), pH 7.0 at concentrations less than 1mg.ml^{-1} for P23VR36S double mutant incubated at 4°C, concentrations less than 1mg.ml^{-1} for normal and inverted solubility crystals of P23TR36S double mutant

incubated at 4°C and 37°C respectively, and concentrations ≈ 10mg.ml^{-1} for aggregates/crystals of P23VR58H double mutant incubated at 25°C over a time period of 12 hours, they were then allowed to settle and gently separated from the supernatant. The solid condensed phase *i.e.* crystals/aggregates separated from the proteins monomers were dispersed in fresh buffer and stirred gently at constant temperature to ensure uniform mixing.

2.6.2 Solubility measurements

The experimental apparatus for solubility measurements consisted of an insulated polystyrene box with internal temperature control via a water coil controlled by an external thermostatized water bath unit (Thermo Scientific DC30-K20) and monitored using a Kane May KM340 differential temperature thermometer with a K type thermocouple. Uniform mixing of the sample was regulated by fixing the sealed glass vial onto a motor controlled rotor. The solubility lines were measured as described previously (Pande *et al.*, 2001, Berland *et al.*, 1992). As the crystals/aggregates melted at a fixed temperature, the concentration of the supernatant protein solution in equilibrium with the condensed phase is monitored by measuring the absorbance (280nm) spectrophotometrically at regular time intervals over a time period of 24-48 hours. To measure protein concentration in the supernatant, the motor was stopped and given sufficient time to allow the protein crystals/aggregates to settle at the bottom of the glass vial. The drawn volume of supernatant from the glass vial was replenished with fresh buffer and left to equilibrate. After a time period of 24-48 hours, the concentration of the supernatant protein solution in equilibrium with the solid condensed phase reaches a constant value, this monomer concentration is defined as the solubility of the protein for a given temperature. The solubility curves were obtained by plotting the protein monomer concentration corresponding to a given temperature for a series of temperature values ranging from 0 to 40°C. Solubility curves were drawn using Origin v6.1 software.

2.7 Protein crystal/aggregate formation for imaging

In order to obtain good quality crystals for phase contrast and polarized light microscopy, it was essential to grow protein crystals under controlled solution and temperature conditions. Concentrated protein solutions in and around the phase boundary for the double mutant proteins were introduced into CoverWell perfusion

chambers (Grace Bio-labs) attached to standard microscopic slides, covered and sealed with a cover slip. The protein solution within the sealed perfusion chamber is then incubated at the specific temperature for crystal/aggregate formation.

Protein crystals/aggregates formed by the above method were imaged using phase-contrast and polarization microscopy. An Olympus BX61 microscope with either 60x or 100x magnification oil-immersion lens were used to visually observe protein crystals/aggregates. The images were recorded using Cell^{F} software and image analysis performed using Image^{J} software (James *et al.*, 2015).

2.8 Protein aggregate formation for kinetic studies

P23VR58H double mutant proteins prepared in 0.1M sodium phosphate buffer containing 20mM DTT (Dithiothreitol), pH 7.0 at concentrations of 5, 7.5 and 10 mg.ml^{-1} were introduced into CoverWell perfusion chambers (Grace Bio-labs) attached to standard microscopic slides, covered and sealed with a cover slip. The monomeric protein solution within the sealed perfusion chamber is then incubated at 30°C on a DC95 stage controller (Linkam Scientific Instruments, UK) to allow aggregate formation to occur.

The growth of the aggregates was monitored and imaged using phase-contrast and polarization microscopy. An Olympus BX61 microscope with either 60x or 100x magnification oil-immersion lens were used to visually observe protein crystals/aggregates. The images were recorded using Cell^{F} software and the diameter of protein aggregates measured using a built-in feature in the Cell^{F} software. The P23VR58H double mutant growth curves were drawn and analysed using Origin v6.1 software.

2.9 Protein crystallography

2.9.1 Crystallization

Protein crystals for X-ray crystallography were grown in 5mm NMR glass tubes (Hilgenberg GmbH) from concentrated protein solutions in and around the phase boundary (solubility line) for the protein. These solutions were incubated at specific temperature conditions for the growth of protein crystals suitable for X-ray crystallographic experiments.

2.9.2 X-ray data collection

X-ray data collection was kindly performed by Prof. Amir Khan, Trinity College Dublin. X-ray diffraction data for P23TR36S double mutant were collected from both rhombic-shaped and rod-shaped crystals that were flash-frozen using glycerol as cryoprotectant. Both data sets were collected at PROXIMA 2 beamline at the French National Synchrotron Source SOLEIL on an ADSC Q315 area detector. All diffraction data were processed, integrated and scaled using iMosflm (Leslie & Powell, 2007) and eventually converted to MTZ format using the CCP4i program package (Collaborative, 1994).

2.9.3 Molecular replacement and refinement

The structures of rhombic and rod-shaped crystal types of P23TR36S double mutant were solved by molecular replacement using P23T single mutant structure (PDB ID 4JGF) and R36S single mutant structure (PDB ID 2G98) as search models in Phaser-MR (McCoy, 2007). After generation of the initial model, iterative cycles of manual model rebuilding were carried out using Coot (Emsley & Cowtan, 2004).

Chapter 3

Production, characterisation and phase diagrams for HGD and the HGD P23VR36S double mutant

3.1 Introduction

3.1.1 Background

Human γD-crystallin (HGD) is one of the most important structural proteins found in the eye-lens. Its phase behaviour is defined by net attractive, short-range, anisotropic interactions responsible for liquid-liquid phase separation and protein crystallization (Pande *et al.*, 2001, Pande *et al.*, 2000). HGD does not crystallize spontaneously, however, there are known single mutants of HGD formed by single amino acid substitutions that spontaneously crystallize with significantly lowered solubility when compared to the native protein (Pande *et al.*, 2001). One such point mutation is the R36S single mutant of HGD which is responsible for the formation of congenital cataracts. Arginine at position 36 is not a crystal contact point in the native protein but it becomes one in the mutant protein resulting in the crystallization behaviour (Kmoch *et al.*, 2000, Basak *et al.*, 2003). The replacement of arginine at position 36 by serine results in loss of a polar, charged residue and its replacement by a less polar, uncharged residue. Loss of a charged residue makes the protein interactions less repulsive, resulting in lowered solubility of the mutant protein (Kmoch *et al.*, 2000, Pande *et al.*, 2001). A second type of single amino acid substitution in HGD at position 23 (P23V, P23T, P23S) inverts the temperature dependence on solubility resulting in an unusual phase behaviour (Pande *et al.*, 2005, McManus *et al.*, 2007, Pande *et al.*, 2009). P23V single mutant is a variant of the naturally occurring P23T single mutant which has been found in diverse population groups worldwide (Santhiya *et al.*, 2002, Mackay *et al.*, 2004, Nandrot *et al.*, 2003, Shentu *et al.*, 2004, Burdon *et al.*, 2004) and is associated with the formation of lamellar (Santhiya *et al.*, 2002), coralliform (Mackay *et al.*, 2004), cerulean (Nandrot *et al.*, 2003), fasciculiform (Shentu *et al.*, 2004), bilateral aculeiform (Vanita & Singh, 2012) and an unclassified silica-like nuclear (Burdon *et al.*, 2004) cataract phenotypes. The P23V single mutant does not appear naturally in human populations and is formed by replacing proline with valine at position 23 on the amino acid chain of native HGD protein. The P23V mutation results in the same characteristic phase behaviour as that of the P23T single mutant, exhibiting inverted temperature dependence of the solubility line, *i.e.* the protein forms condensed phases (crystals/aggregates) at higher temperatures and melts as the temperature is lowered (Pande *et al.*, 2005, McManus *et al.*, 2007). In the native protein, position 23 is not a crystal contact point and replacing proline at position 23

neither alters the crystal structure of the protein nor does it lead to the formation of new crystal contacts in the single mutant protein (Pande *et al.*, 2005, Ji *et al.*, 2013). The lowered solubility is attributed to changes in binding energy of the mutant which is highly temperature dependent (McManus *et al.*, 2007).

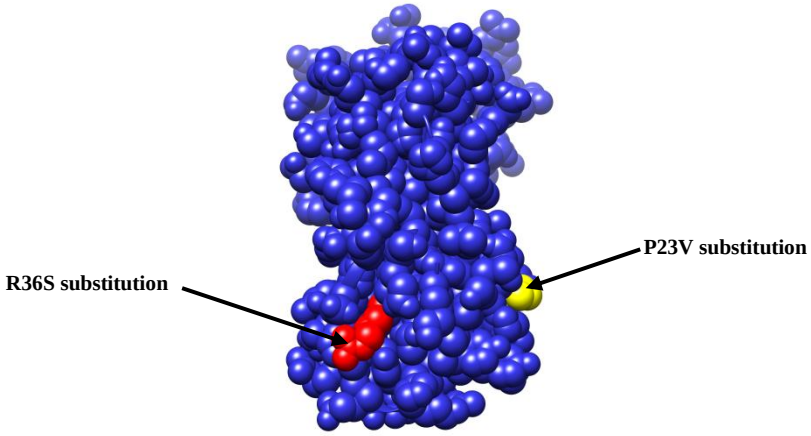

Figure 3.1: *3D representation of the two amino acid substitutions (P23V and R36S) incorporated in the native HGD protein to create the P23VR36S double mutant.*

The first section of this chapter is about the phase diagram for native human γD-crystallin protein and how well it compares with previously published data and secondly, creation of a double mutant protein, incorporating the R36S and P23V single mutants (figure 3.1) and exploring the phase behaviour for the newly formed double mutant protein.

3.2 Aim of the study

Single mutants of HGD have indicated that changes to specific amino acid residues on the protein surface contribute to changes in the protein phase behaviour. To increase the complexity of the system, we have introduced two single amino acid substitutions to the same protein. The first double mutant, P23VR36S, is created by inclusion of both the P23V and R36S single mutants of HGD (James *et al.*, 2015). Both single mutants are well characterized and demonstrate explicitly different phase behaviours 1. Inverted solubility due to the P23V single mutant and 2. Propensity to crystallize due to the R36S single mutant. This phase diagram approach was introduced to describe the competitive nature of anisotropic interactions in proteins, to qualitatively determine if the characteristic behaviour of the individual single mutants were maintained in the double mutant and to identify if the position of the phase boundaries of the single mutants has been altered when incorporated within the double mutant. Among the

double mutant proteins created to study phase behaviour prediction, P23VR36S was first chosen as the phase boundaries for the individual single mutants do not overlap when the phase boundaries were re-drawn on the same diagram. The key question is which behaviour will be exhibited by the double mutant protein?

3.3 Results and Discussions

3.3.1 Production and characterisation of human γD-crystallin

Human γD-crystallin (HGD) proteins were over-expressed in *E.coli* as described in chapter 2. Chromatographic separation of the crude lysate by size-exclusion chromatography (figure 3.2) and cation-exchange chromatography (figure 3.4), following extraction indicated that typically 200 mg of purified protein was obtained from 7.2 L of bacterial cell culture.

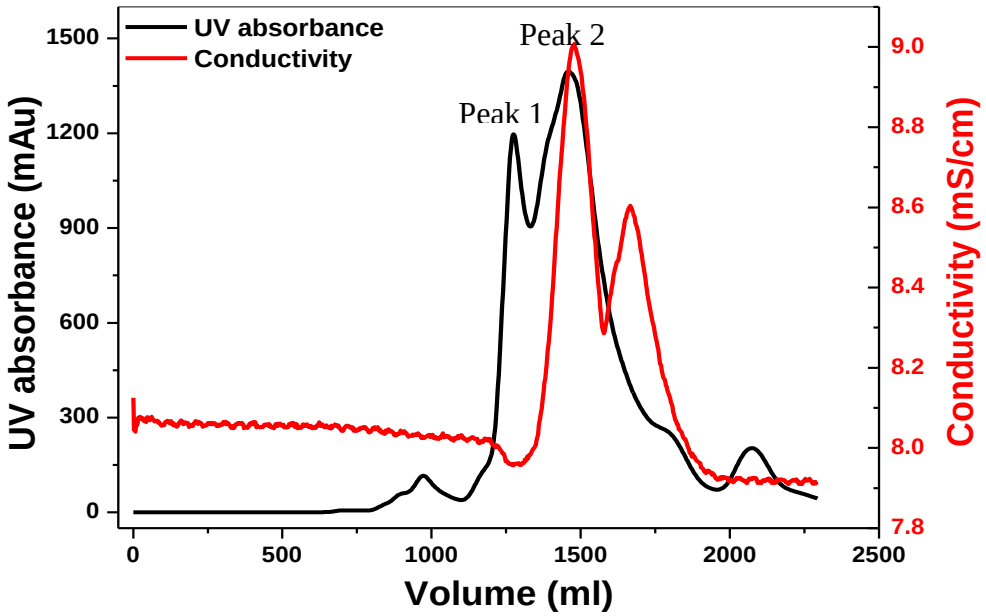

Figure 3.2: *Size-exclusion chromatogram depicting the primary purification step for native HGD protein.*

The protein fractions associated with peak 1 and peak 2 were analysed using SDS-PAGE. The HGD native protein has a molecular mass of 20,607 ± 1 Da as confirmed by intact molecular weight analysis by electrospray ionization mass spectrometry performed at the Finger prints proteomics facility, College of life sciences, University of Dundee. This molecular mass for HGD is consistent with previously published results (Hanson *et al.*, 1998). The presence of a strong band at 21.5 kDa in the lane for peak 1 (figure 3.3) confirms that protein extraction has been successful (figure 3.2).

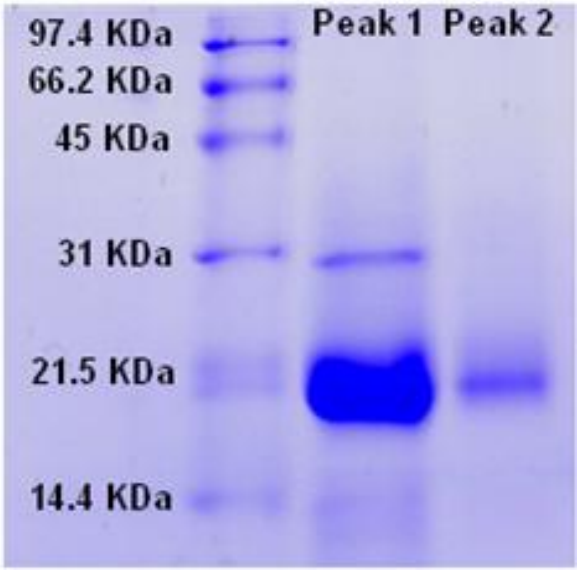

Figure 3.3: *SDS-PAGE gels confirming elution of native HGD protein into fractions associated with peak 1.*

Ion-exchange chromatography of the band identified as HGD completed the purification process (figure 3.4).

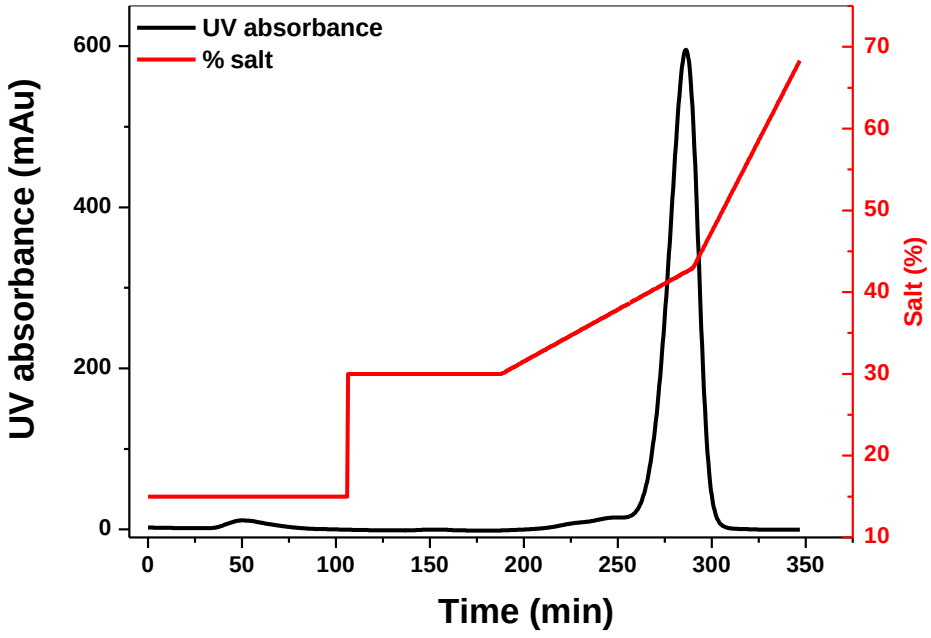

Figure 3.4: *Cation-exchange chromatogram depicting clean elution profile for native HGD protein.*

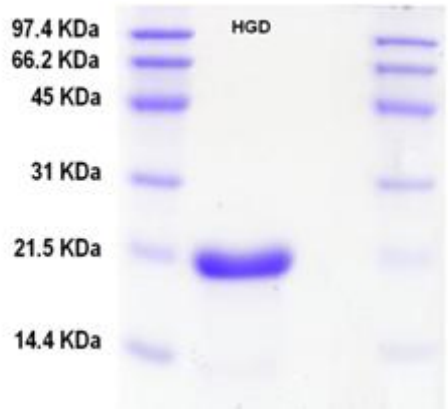

Figure 3.5: *12.6% reducing SDS-PAGE gel confirming the purity of native HGD protein.*

The purified protein was characterised by SDS-PAGE (figure 3.5) showing a clean monomer band.

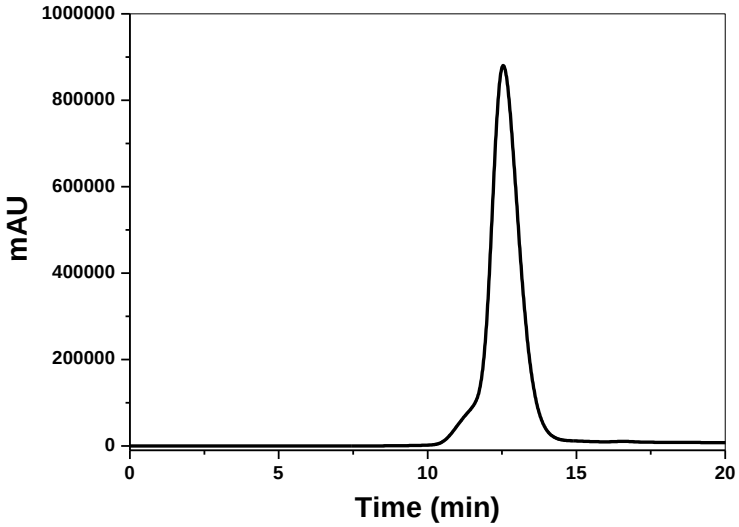

Figure 3.6: *SE-HPLC profile for native HGD, when freshly prepared in 0.1M sodium phosphate buffer pH 7.0 in the presence of 20mM DTT.*

Chromatographic analysis of SE-HPLC data confirmed that native HGD protein is mostly monomeric (≥ 99% purity), which was confirmed by a reducing SDS-PAGE gel (figure 3.5) and shows a major peak at 12.52 min (figure 3.6). From the calibration of the Superdex 75 10/300 GL SE-HPLC column, a peak at 12.52 min indicates a molecular weight in the region of 20.6 KDa, which is consistent with published results (Hanson *et al.*, 1998).

3.3.2 Phase diagram for human γD-crystallin: Solubility and phase separation measurements

The coexistence curve (liquid-liquid phase separation) for native human γD-crystallin (figure 3.7) was measured. This has been measured previously (Asherie *et al.*, 1998, Broide *et al.*, 1991), but to ensure consistency with previous work, this was repeated here.

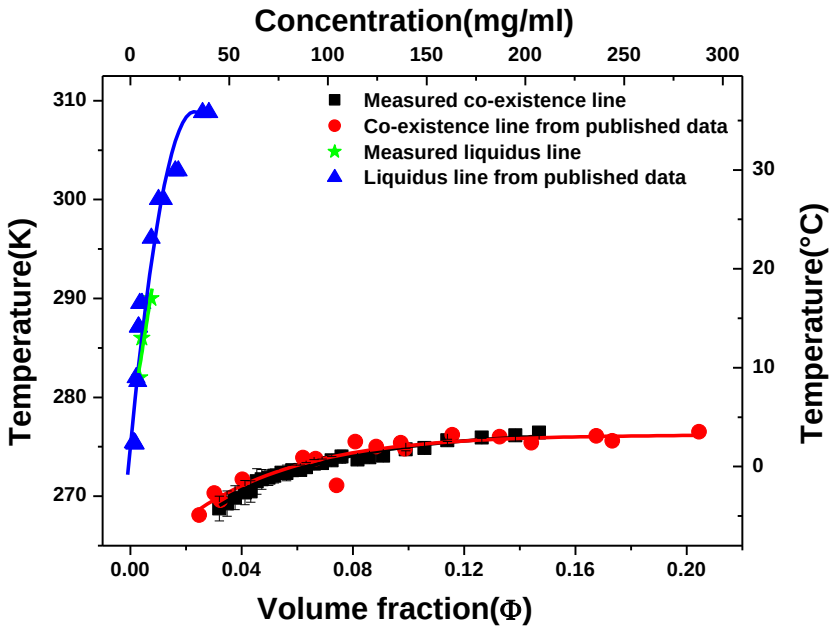

Figure 3.7: *The phase diagram of human γD-crystallin showing the liquid-liquid coexistence phase boundary and liquidus line. To compare the accuracy of our experimentally measured coexistence curve (squares) and solubility curve (star-shaped), we redrew previously measured data points for both coexistence (circles) (Broide et al., 1991, Asherie et al., 1998) and solubility (upright triangles) (Pande et al., 2001) lines.*

The coexistence curve marks the boundary between a homogenous single phase and a heterogeneous two phase region at a particular temperature (Asherie *et al.*, 1998, Thomson *et al.*, 1987). Figure 3.7 clearly shows that our phase separation measurements for native human γ-D crystallin and previously published phase separation measurements (Asherie *et al.*, 1998) are in good agreement. The phenomena of liquid-liquid phase separation resulting in the formation of protein-rich and protein-poor phases formed by temperature quenching protein solutions is metastable with respect to protein crystallization. However for HGD the formation of crystals does not occur spontaneously. Various precipitants such as PEG 8000 (Polyethylene Glycol) at

20% (w/v) and (*RS*)-2-methyl-2,4-pentanediol (MPD) were tested to induce protein crystallization. Crystallization was only observed in native HGD solutions in the presence of (*RS*)-2-methyl-2,4-pentanediol (MPD). Protein crystals grew from protein solutions at a concentration of 50 $mg.ml^{-1}$ in the presence of 20µl MPD solvent. HGD crystals formed were examined by polarization microscopy (figure 3.8).

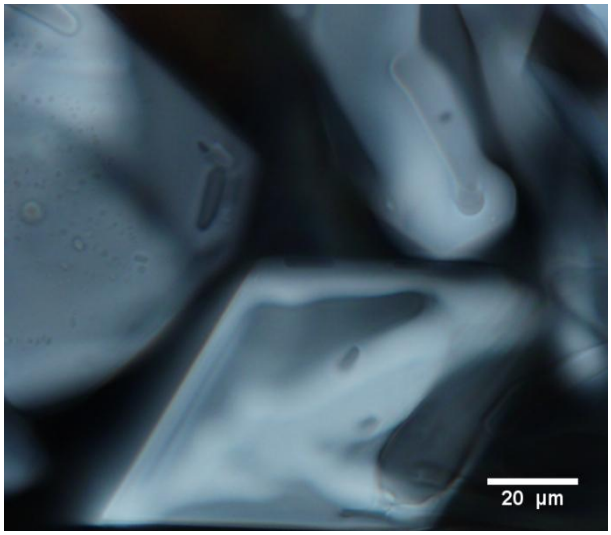

Figure 3.8: *Polarization microscopy image showing crystals formed by native HGD protein in the presence of (RS)-2-methyl-2,4-pentanediol (MPD).*

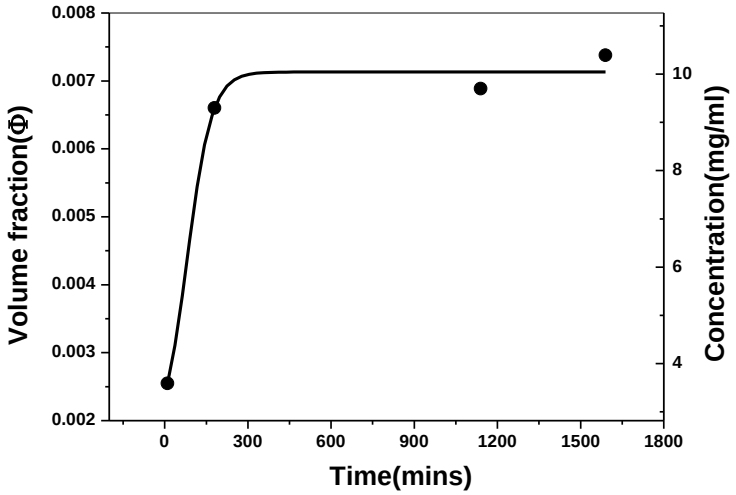

Figure 3.9: *Concentration-time profile for native HGD protein monomers in equilibrium with protein crystals at 22°C over a time period of 27 hours.*

Protein crystals thus formed were then used for solubility analysis to determine the liquidus line for HGD. The solubility curve (liquidus line) for native human γ-D crystallin measured is shown in figure 3.7. Every point on the solubility line is a measure of the concentration of protein monomers in solution that has attained

equilibrium with protein crystals at a specific temperature, over a time period of 24 - 48 hours. Figure 3.9 indicates the change in monomer concentration for native HGD as it reaches equilibrium at 17°C over 27 hours.

3.3.3 Production and characterisation of P23VR36S double mutant protein

The P23VR36S double mutant plasmid DNA prepared by site-directed mutagenesis was sequenced with the T7 promoter and confirmed to incorporate the two single amino acid substitutions, proline (CCC) to valine (GTC) at position 23 and arginine (CGC) to serine (AGC) at position 36 within the double mutant (figure 3.10).

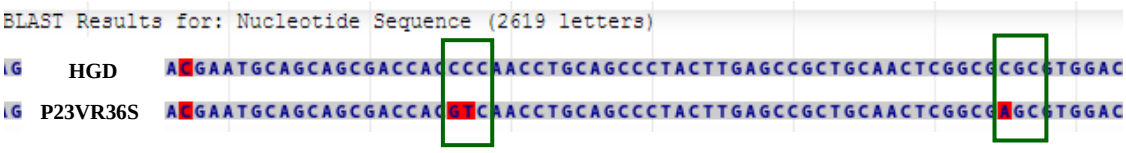

Figure 3.10: *Plasmid DNA sequence for P23VR36S double mutant of HGD aligned and compared against native HGD using BLAST.*

3.3.3.1 Determination of plasmid DNA yield

The purity and concentration of plasmid DNA obtained following mutagenesis and plasmid extraction for native, single and double mutants of HGD are as listed below (table 3.1):

	HGD	P23V	R36S	P23VR36S
Abs_{260}/Abs_{280}	1.8	1.75	1.8	1.77
Conc of DNA (ng/µl)	49.3	44.9	52.9	51.4

Table 3.1: *Purity and concentration of DNA obtained from plasmid purification.*

The P23VR36S double mutant protein was isolated and purified as described in chapter 2. Chromatographic separation of the crude lysate by size-exclusion chromatography (figure 3.11) and cation-exchange chromatography (figure 3.13), following extraction indicated that typically 29 mg of purified protein was obtained from 7.2 L of bacterial cell culture.

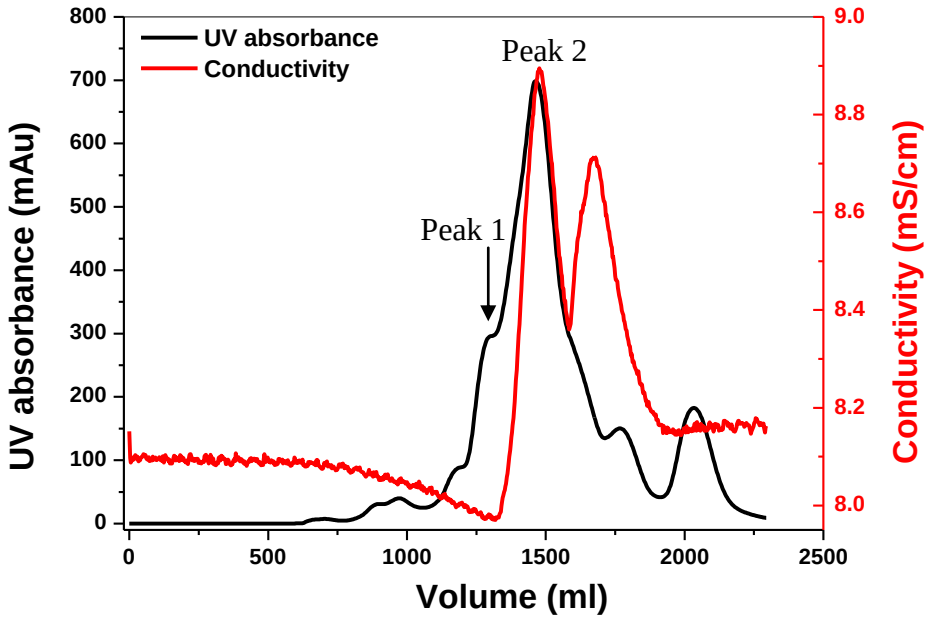

Figure 3.11: *Size-exclusion chromatogram for P23VR36S double mutant of HGD.*

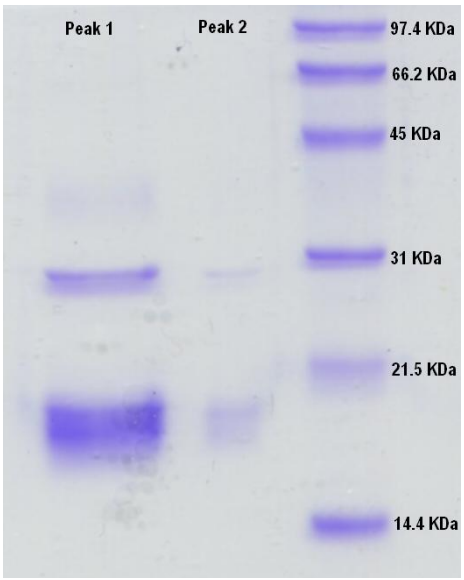

Figure 3.12: *SDS-PAGE gels confirming elution of P23VR36S double mutant protein into fractions constituting peak 1.*

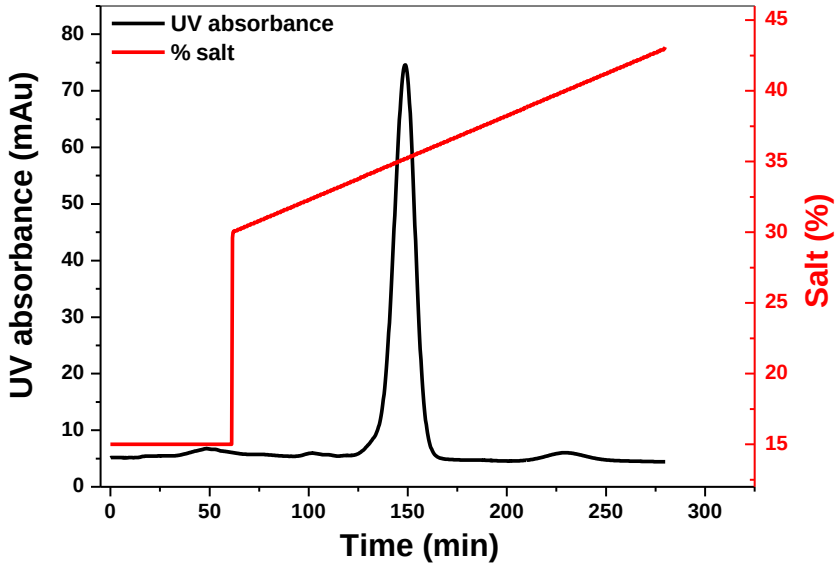

Figure 3.13: *Cation-exchange chromatogram depicting clean elution profile for the P23VR36S double mutant.*

The size-exclusion protein fractions associated with peak 1 and peak 2 (figure 3.11) for the P23VR36S double mutant were analysed using SDS-PAGE. The P23VR36S double mutant protein has a molecular mass of 20,539 ± 1 Da confirmed by intact molecular weight analysis by electrospray ionization mass spectrometry. This molecular mass for the P23VR36S double mutant is consistent with the expected mass value for the specified amino acid substitutions. Peak one corresponded to a molecular weight band at less than 21.5 kDa on SDSPAGE (figure 3.12). Cation-exchange chromatographic separation completed the purification process (figure 3.13).

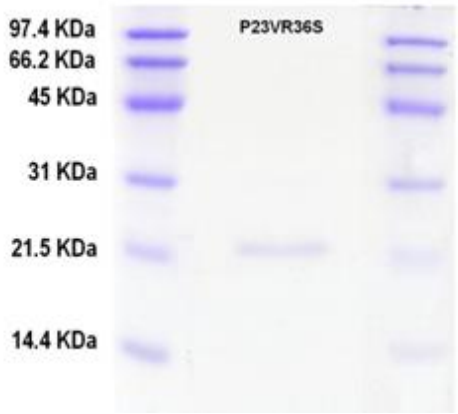

Figure 3.14: *12.6% reducing SDS-PAGE gel showing a single monomer band confirming the purity of P23VR36S double mutant protein.*

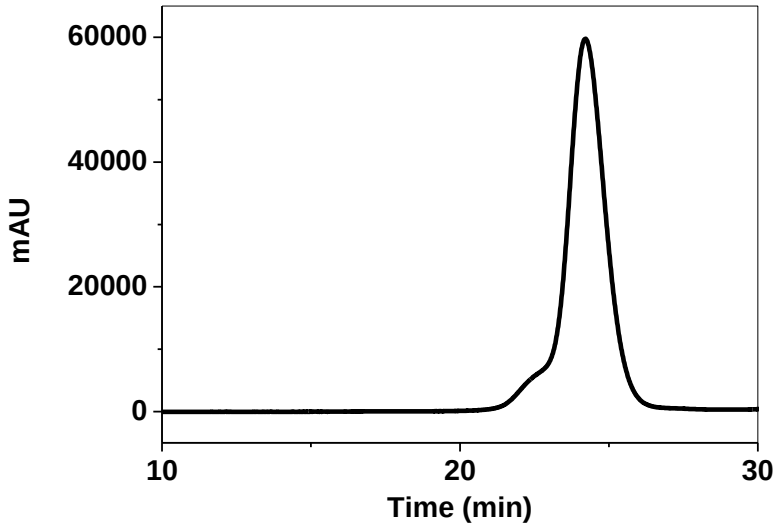

Figure 3.15: *SE-HPLC profile for P23VR36S double mutant when freshly prepared in 0.1M sodium phosphate buffer pH 7.0 in the presence of 20mM DTT.*

Chromatographic analysis using SE-HPLC confirmed that P23VR36S double mutant protein is mostly monomeric (≥ 98% purity), also confirmed by reducing SDS-PAGE gel (figure 3.14) and shows a major peak at 24 min (figure 3.15). From the calibration of the Superdex 200 10/300 GL SE-HPLC column used to isolate the protein, a peak at 24 min indicates a molecular weight is in the region of 20.5 KDa, which is within the range expected for P23VR36S double mutant.

3.3.3.2 Structural studies of P23VR36S double mutant protein using circular dichroism spectroscopy and fluorescence spectroscopy

Circular dichroism spectra. To examine if structural and conformational changes have occurred in the protein due to this double mutation, we have compared the CD spectral data of P23VR36S double mutant and native HGD. CD spectra in the far-UV region monitors secondary structural or conformational changes in proteins and those in the near-UV region monitors tertiary structural changes in proteins. Examination of the far-UV and near-UV CD spectra of P23VR36S double mutant shows that it has nearly identical secondary and tertiary structures to native HGD. This suggests that the P23VR36S double mutation occurs without any significant secondary or tertiary structural changes to the native protein. Both P23VR36S double mutant and native HGD show a negative ellipticity at around 218 nm on the far-UV CD spectra and also display similar spectral features on the near-UV CD spectra for the proteins (figure 3.16).

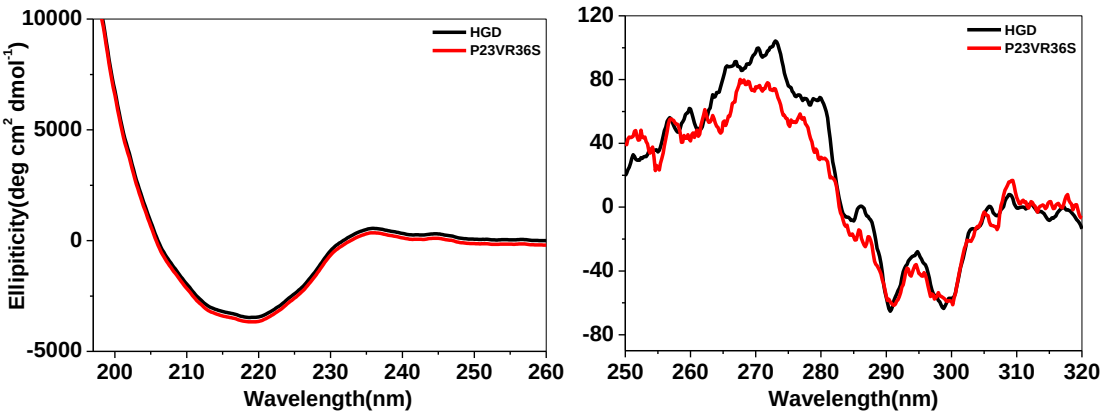

Figure 3.16: *Comparison between secondary and tertiary structures of P23VR36S double mutant and native HGD using far-UV CD spectra (left) and near-UV CD spectra (right).*

There is a slight broadening and blueshift below 220 nm associated with the far-UV CD spectra for the P23VR36S double mutant protein (figure 3.16). This blueshift has been observed before (Pande *et al.*, 2001). This was explained as the serine (36) residue disrupting the normal β-sheet structure resulting in the formation of a 'β-buldge' which is responsible for the observed changes in the far-UV CD spectra (Pande *et al.*, 2001). We assume that the same explanation applies to our data.

Fluorescence spectra. Fluorescence emission spectra of γ-crystallins arise mainly due to four tryptophan residues buried within the protein core (Mandal *et al.*, 1985). Fluorescence emission spectra for the P23VR36S double mutant and native HGD were measured using excitation at 280 nm and 295 nm. The fluorescence emission spectra were normalized with respect to protein concentration. The λ_{max} (emission maximum) value for P23VR36S double mutant protein was well within the range of those observed for all γ-crystallins (λ_{max} from 324 to 335 nm (Kono *et al.*, 1990)) which indicates that tryptophan residues in the mutant protein remain buried in the hydrophobic core (Mandal *et al.*, 1985). However, a red shift in λ_{max} for tryptophan fluorescence is observed in fluorescence spectra of P23VR36S double mutant protein (figure 3.17). This red-shift was also observed in the R36S single mutant and is due to the greater solvent-exposure of the closest tryptophan (trp-42) residue to the serine (ser-36) residue, which resulted in the red shift of the fluorescence emission spectrum (Pande *et al.*, 2001).

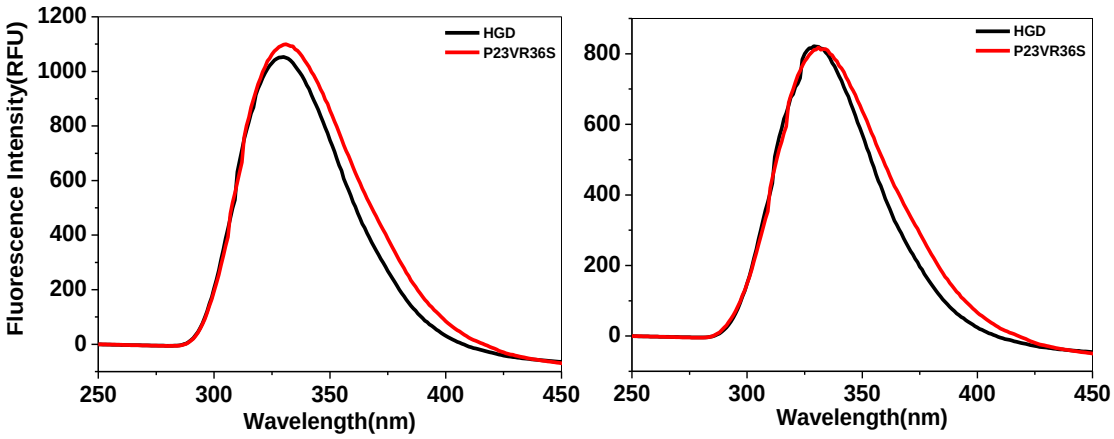

Figure 3.17: *Fluorescence emission spectra (excitation wavelength=280 nm (left); excitation wavelength=295 nm (right)) for P23VR36S double mutant protein in comparison with native HGD.*

3.3.4 Phase diagram (solubility line) for P23VR36S double mutant

The double mutant protein crystallized spontaneously at low protein concentrations (less than 1mg/ml). The liquidus line for the double mutant was measured. The phase diagram (figure 3.18) indicates that the liquidus line for the P23VR36S double mutant overlaps with that measured for the R36S single mutant (James *et al.*, 2015).

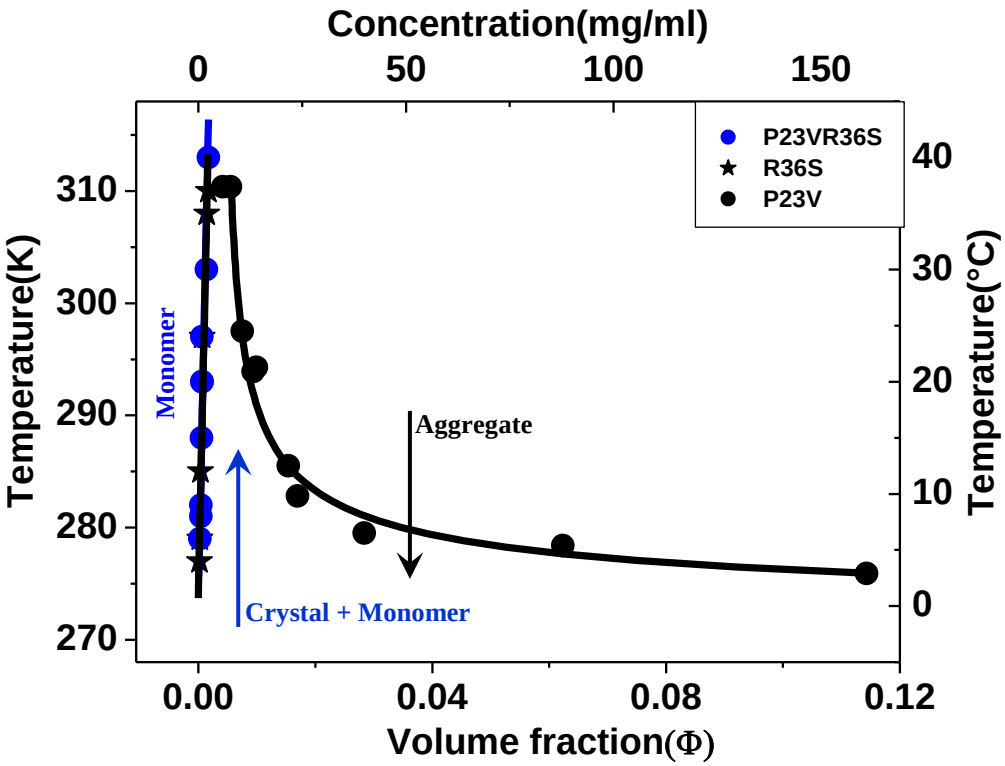

Figure 3.18: *Phase diagram for the single mutants of HGD, P23V and R36S (black lines) and the double mutant P23VR36S (blue line) taken from (James et al., 2015). Data for the single mutants of HGD are taken from (Pande et al., 2000) and (Pande et al., 2005).*

The liquidus line for the P23VR36S double mutant indicates that the crystal has normal solubility. The inverted temperature dependence of the solubility line associated with the P23V mutation is absent (James *et al.*, 2015). It is most likely that it is not possible to access the solubility line associated with the P23V mutant as it occurs at significantly higher concentrations than it is possible to attain for this double mutant, since crystallization occurs at lower protein concentrations and on shorter timescales (James *et al.*, 2015).

3.3.4.1 Controlled crystallization of P23VR36S double mutant

In order to obtain good quality crystals for phase contrast and polarized light microscopy, protein crystals were grown from dilute solutions (less than 1 $mg.ml^{-1}$). The double mutant protein formed crystals that are consistent in morphology (figure. 3.19) and behaviour with crystals of R36S single mutant (James *et al.*, 2015).

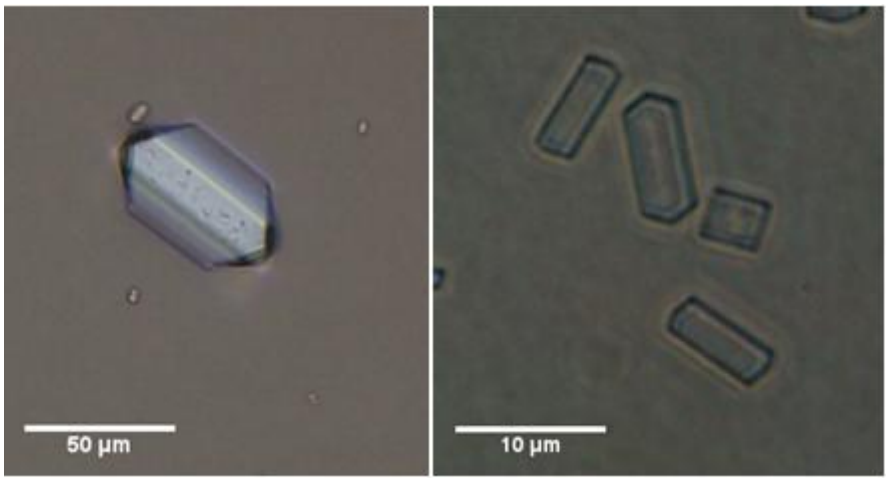

Figure 3.19: *Phase contrast microscopy images showing crystals formed by R36S single mutant (left) and P23VR36S double mutant of HGD (right), taken from (James et al., 2015).*

3.3.4.2 Analysis of P23VR36S double mutant solubility data

Analysis of the solubility data for P23VR36S double mutant reveals interesting information about the phase diagram for the protein. The transfer chemical potentials, $\Delta\mu_{trans}$, for native HGD and a few of its single mutants have been estimated from their liquidus/solubility lines as discussed previously (McManus *et al.*, 2007).

The same method described by (McManus *et al.*, 2007) is employed here to calculate the transfer chemical potentials and to examine if the phase boundaries for the double mutants can be quantitatively compared against the phase boundaries for the single

mutants. This helps us probe if the value of binding energy for the crystals in the double mutants are the same as for the single mutants (McManus *et al.*, 2007).

The transfer chemical potential, $\Delta\mu_{trans}$, relates to the volume fraction, ϕ, using the Van't Hoff law by equation 4.1,

$$\ln \phi = -\frac{\Delta\mu_{trans}}{\kappa T} \qquad 4.1$$

Defining $X=T_0/T$ and assigning T_0= 303K (the temperature at which the solubility lines for the P23T and R36S single mutants intersect), the above expression 4.1 for the Van't Hoff law becomes,

$$\phi = \exp -\left(\frac{\Delta\mu_{trans}}{\kappa T_0}\right)X \qquad 4.2$$

It is observed that at volume fractions greater than ≈ 0.02, $\Delta\mu_{trans}/kT_0$ does not follow a linear relationship with temperature (T). To account for this non-linearity, the second virial coefficient, B_2, which is defined by (Lomakin *et al.*, 1999) as,

$$B_2 = -4\left[\left(e^{\varepsilon} - 1\right)\left(\lambda^3 - 1\right) - 1\right] \qquad 4.3$$

where $\varepsilon = \varepsilon_{eff}/kT$, and from the LLPS, λ=1.25 and ε_{eff}/kT_C = 1.27 is included in the calculation for the transfer chemical potentials.

If we employ the linear dependence of $\Delta\mu_{trans}/kT_0$ in the Van't Hoff law, we find that in the low volume fraction region (ϕ < ≈0.02) the solubility line becomes,

$$\phi \cong \phi_0 \exp\left[\left(V - S\right)\Delta\tau\right] \qquad 4.4$$

where

$$\phi_0 = \exp -\left(\frac{\Delta\mu_{trans}(T_0)}{\kappa T_0}\right) \qquad 4.5$$

and $V = \dfrac{\Delta\mu_{trans}(T_0)}{\kappa T_0}$; $S = \left[\partial\left(\dfrac{\Delta\mu_{trans}}{\kappa T_0}\right)\partial\tau\right]_{\tau=1}$; and $\Delta\tau = \dfrac{(T - T_0)}{T_0}$.

The transfer chemical potential, $\Delta\mu_{trans}$ for each of the solid phases of the double mutant proteins were determined utilizing their solubility data. While plotting $\Delta\mu_{trans}/kT_0$ vs. T/T_0, we find that proteins with normal solubility have negative slope

values and those with inverted solubility have positive slope values. The quantity K determines solubility dependence, where $K = \left(\left[\frac{\Delta \mu_{trans}(T_0)}{\kappa T_0} \right] - S \right)$.

Examining the solubility data for the P23VR36S double mutant and the R36S single mutant suggests that their liquidus lines almost overlap qualitatively as shown in figure 3.18. Quantitative estimation of solubility data is allowed by plotting $\Delta\mu_{trans}/kT_0$ vs. T/T_0 and by determining the value of V for the proteins.

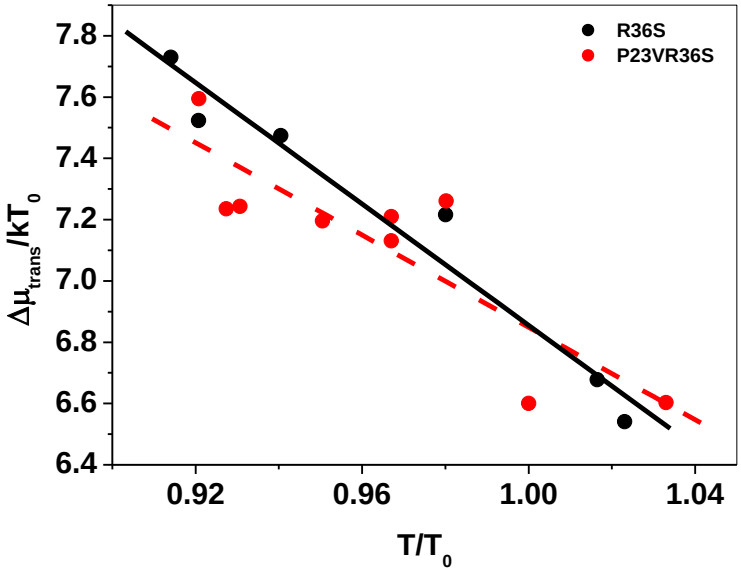

Figure 3.20: *Temperature dependence on solubility for R36S single mutant and P23VR36S double mutant of HGD.*

As illustrated in figure 3.20, the negative slope for $\Delta\mu_{trans}/kT_0$ vs. T/T_0, indicates normal temperature dependence on solubility for the P23VR36S double mutant. There are two competing interactions for the solubility behaviour of the double mutant, P23VR36S and it clearly demonstrates that the R36S mutation dominates the behaviour (James *et al.*, 2015). The values for the parameters S, the slope for $\Delta\mu_{trans}/kT_0$ vs. T/T_0 and V, the intercept at $T=T_0= 303K$ for the single mutants (P23V, R36S) and the double mutant, P23VR36S are detailed in table 3.2. In each case, the value of K indicates the solubility nature of the protein. The V values for the single mutant, R36S and the double mutant, P23VR36S are almost the same. This clearly demonstrates that the behaviour for the P23VR36S double mutant is exclusively influenced by the R36S substitution for the double mutant.

Protein	S	V	K = V – S
HGD	-19	4.4	23.4
P23V	20	5.1	-14.9
R36S	-10	6.9	16.9
P23VR36S	-7.5	6.8	14.3

Table 3.2: *Values of the parameters S, slope for* $\Delta\mu_{trans}/kT_0$ *vs.* T/T_0 *and V, the intercept at* $T=T_0= 303K$ *for each protein.*

3.3.4.3 Conclusion

The phase diagram (co-existence curve and liquidus line) for native human γD-crystallin was measured and is consistent with previously published results. The first double mutant, P23VR36S is a good starting point to validate the concept of protein phase diagram prediction based on the phase behaviour of the individual single mutants. P23VR36S double mutant created by incorporating the two individual single mutants P23V and R36S are soluble following extraction and purification procedures, they maintain their native structure upon mutagenesis confirmed by spectroscopic techniques and form a solid condensed phase (crystals) in a similar way to one of the parent single mutants (James *et al.*, 2015). In addition, the equilibrium solubility line for the P23VR36S double mutant overlaps with the solubility line for the R36S single mutant. Thus, the morphological and behavioural similarity suggests that the phase behaviour due to the second single mutant P23V displaying inverted solubility is suppressed in the double mutant. This double mutant provides significant insights into the phase behaviour of double mutant proteins that were created by mutagenesis of native HGD. However, predicting the phase behaviour of one double mutant from their individual single mutants is not sufficiently convincing and hence a second double mutant, P23TR36S was created.

The creation of the first double mutant, P23VR36S also raised a number of questions. P23V single mutant has a very unusual phase diagram with inverted solubility, however the coexistence curves of the P23V single mutant and native HGD overlapped indicating that there is very little difference in the net interaction potential between the two proteins (McManus *et al.*, 2007). Therefore, anisotropic/aeolotopic interactions accounted for the inverted solubility behaviour which was only engaged in the solid phase. Although the amino acid substitution associated with inverted solubility is present in the double mutant, the solid phase only displays normal solubility while the inverted behaviour due to the P23V substitution is absent (James *et al.*, 2015).

Chapter 4

Phase diagram and crystal structures for the HGD P23TR36S double mutant

4.1 Introduction

4.1.1 Background

The P23T single mutant of HGD is a naturally occurring mutation, formed by replacing proline with threonine at position 23 on the amino acid chain of the native HGD protein (Pande *et al.*, 2005). It exhibits an inverse temperature dependence of the solubility line, *i.e.* the protein forms condensed phases (crystals/aggregates) at higher temperatures that melt as the temperature is lowered (Pande *et al.*, 2005, McManus *et al.*, 2007), and this characteristic behaviour is observed at significantly lower concentrations than for the native protein. A number of different structural studies, using circular dichroism spectroscopy, NMR and X-ray crystallography have indicated that replacement of proline at position 23 with a threonine does not alter the structure of the protein (Pande *et al.*, 2009, Pande *et al.*, 2005, Ji *et al.*, 2013). Furthermore, there is no indication from these studies that any new crystal contact points are formed in crystals of the mutated protein (Pande *et al.*, 2005, Ji *et al.*, 2013). Thus, the lowered solubility is attributed to changes in binding energy of the mutant which is highly temperature dependent (McManus *et al.*, 2007). The second mutation used here is a single amino acid substitution in HGD at position 36 (R36S). As described previously in section 3.1.1, this substitution results in spontaneous crystallization of the protein and significantly lower solubility than the native protein (Pande *et al.*, 2001).

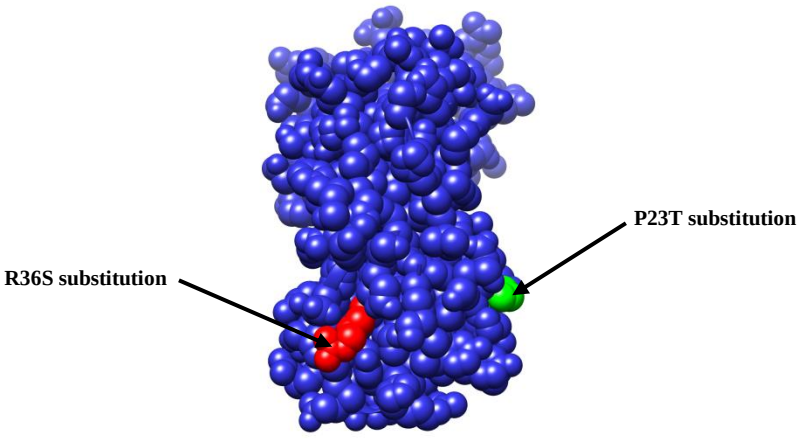

Figure 4.1: *3D representation of the two amino acid substitutions (P23T and R36S) incorporated in the native HGD protein to create the P23TR36S double mutant.*

This chapter provides a detailed description on the creation of a second double mutant P23TR36S (figure 4.1) and exploring the phase diagram for the newly formed double

mutant. This mutant formed two different crystal types and x-ray crystallography has been used to investigate the structures of the crystals formed. The results from a preliminary analysis of this data are presented.

4.2 Aim of the study

The first double mutant created *i.e.* P23VR36S was shown to exhibit the phase behaviour associated with only one of the amino acid substitution *i.e.* the R36S single mutant. The behaviour associated with the other single mutant P23V which demonstrated inverse temperature dependence on solubility was not observed. This was not surprising given the low solubility of the double mutant protein. It was not possible to concentrate the protein to sufficiently high concentrations to the region of the phase diagram, where condensation associated with the P23V single mutant was observed. (James *et al.*, 2015). To address this point, a second double mutant *i.e.* P23TR36S which incorporates the two individual single mutants P23T and R36S was prepared. While the propensity to crystallize due to the R36S mutation remains, the P23T mutation has much lower solubility than P23V (Pande *et al.*, 2005). This allowed us to explore if the phase behaviour for the newly created double mutant P23TR36S displays both normal and inverted phase behaviours associated with the individual single mutants as the liquidus/solubility lines for the individual single mutants (P23T and R36S) overlap if they are superimposed. In addition to the analysis of solubility data to predict the double mutant behaviour, initial structural analysis for the two crystal types formed by the double mutant was performed by Prof. Amir Khan, Trinity College Dublin.

4.3 Results and Discussions

4.3.1 Production and characterisation of P23TR36S double mutant protein

The P23TR36S double mutant plasmid DNA prepared by site-directed mutagenesis using primers for the two individual single mutants (P23T and R36S) was sequenced with the T7 promoter. DNA sequencing confirmed the two single amino acid substitutions, proline (CCC) to threonine (ACC) at position 23 and arginine (CGC) to serine (AGC) at position 36 within the double mutant (figure 4.2).

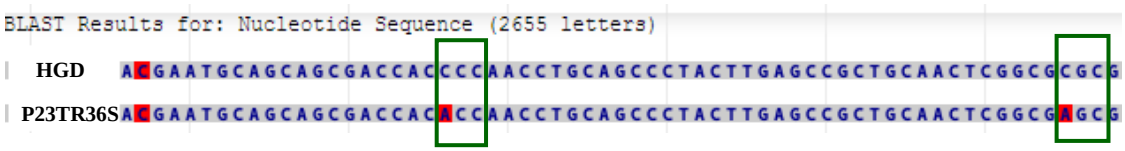

Figure 4.2: *Plasmid DNA sequence for P23TR36S double mutant of HGD aligned and compared against native HGD using BLAST.*

The P23TR36S recombinant plasmid DNA obtained following mutagenesis was analysed for its purity and concentration by measuring the Abs_{260}/Abs_{280} ratio by UV absorbance and equal to 1.86 (a value greater than 1.8 indicates greater than 90% nucleic acid content) and 37.6 ng.µl^{-1} respectively. The protein was expressed and purified as described in chapter 2.

The mutant protein was identified from the various peaks observed in the size exclusion chromatogram (figure 4.3) by SDS-PAGE. These fractions were pooled and further purified by ion exchange chromatography, producing a well purified product (figure 4.4). P23TR36S double mutant protein has a molecular mass of 20,541 ± 1 Da as determined by intact molecular weight analysis by electrospray ionization mass spectrometry performed at the Finger prints proteomics facility, College of life sciences, University of Dundee. This molecular mass for P23TR36S double mutant is consistent with the expected mass value for the specified amino acid substitutions.

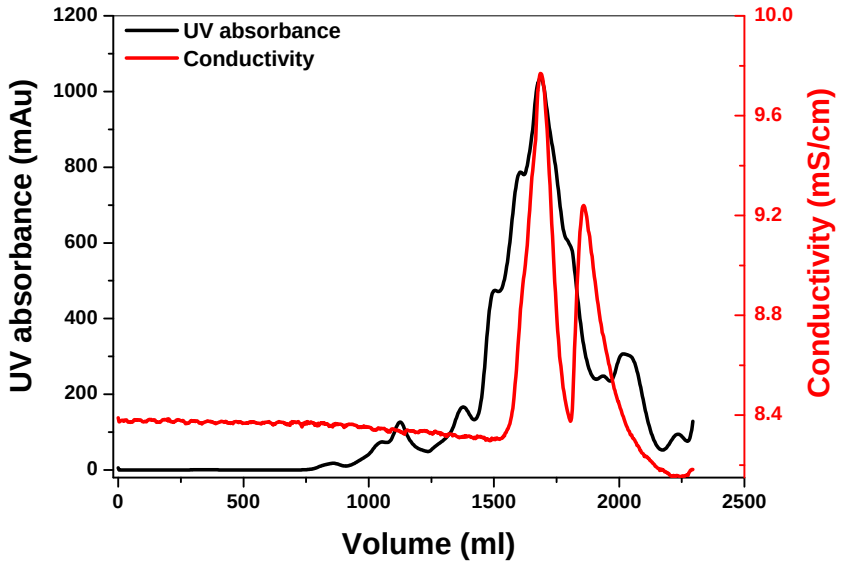

Figure 4.3: *Size-exclusion chromatogram depicting the primary purification step for P23TR36S double mutant of HGD.*

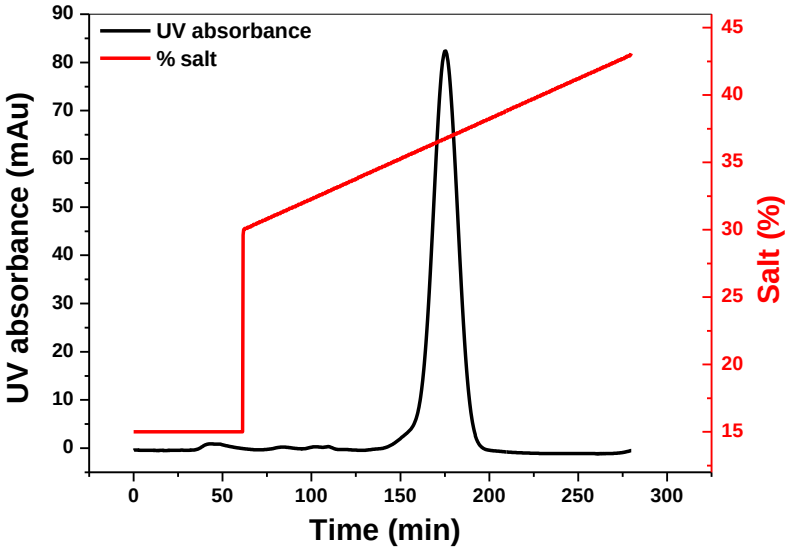

Figure 4.4: *Cation-exchange chromatogram depicting clean elution profile for P23TR36S double mutant.*

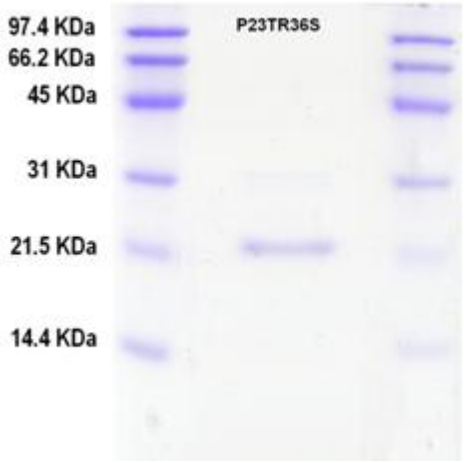

Figure 4.5: *12.6% reducing SDS-PAGE gel showing a single monomer band confirming the purity of P23TR36S double mutant protein.*

Chromatographic analysis using SE-HPLC data confirmed that P23TR36S double mutant protein is mostly monomeric (≥ 99% purity), also confirmed by reducing SDS-PAGE gel (figure 4.5) and shows a major peak at 24 min (figure 4.6). From the calibration of the Superdex 200 10/300 GL SE-HPLC column used to isolate the protein, a peak at 24 min indicates a molecular weight in the region of 20.5 KDa, which is within the range expected for P23TR36S double mutant.

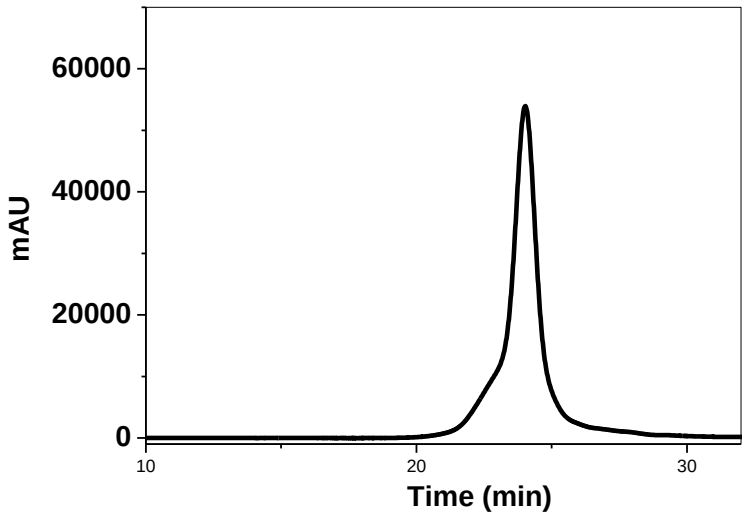

Figure 4.6: *SE-HPLC profile for P23TR36S double mutant when freshly prepared in 0.1M sodium phosphate buffer pH 7.0 in the presence of 20mM DTT.*

4.3.1.1 Structural studies of P23TR36S double mutant protein using circular dichroism spectroscopy and fluorescence spectroscopy

Circular dichroism spectra. Comparison of the CD spectral data for the P23TR36S double mutant and native HGD was performed to determine if there were structural or conformational changes that occurred in the protein due to this double mutation. CD spectra in the far-UV region and near-UV region monitor structural changes in proteins. The far-UV and near-UV CD spectral analysis of the P23TR36S double mutant shows that it has very similar secondary and tertiary structures to native HGD. This suggests that the P23TR36S double mutation occurs without any significant secondary or tertiary structural changes to the native protein. Both the P23TR36S double mutant and native HGD show a negative ellipticity at around 218 nm on the far-UV CD spectra and also display similar spectral features on the near-UV CD spectra for the proteins (figure 4.7). Additional broadening and blue-shift below 220 nm are observed in the far-UV CD spectrum for the P23TR36S double mutant protein (figure 4.7). This observation is similar to that observed in the R36S single mutant where the serine residue disrupts the normal β-sheet structure resulting in the formation of a 'β-buldge' which is responsible for the observed changes in the far-UV CD spectra for the double mutant protein.

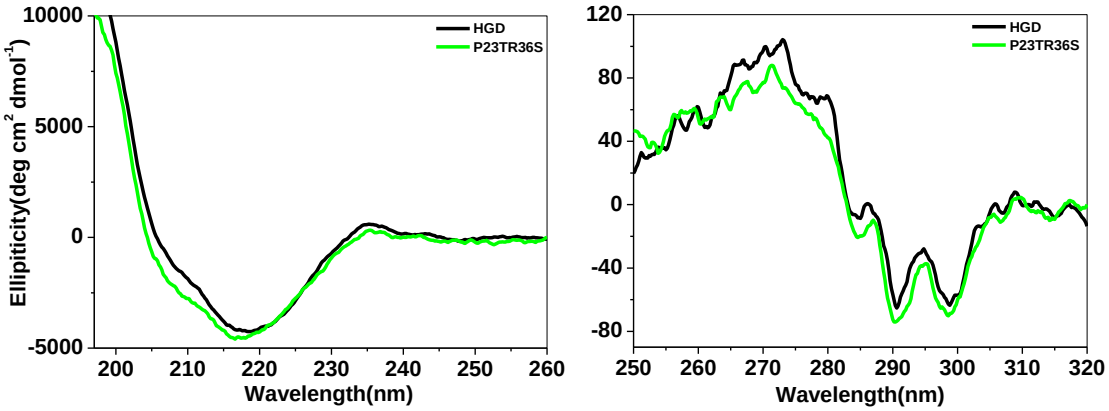

Figure 4.7: *Comparison between secondary and tertiary structures of P23TR36S double mutant and native HGD using far-UV CD spectra (left) and near-UV CD spectra (right).*

Fluorescence spectra. Fluorescence emission spectra for γ-crystallins arise mainly due to four tryptophan residues buried within the protein core (Mandal *et al.*, 1985). These tryptophan residues are regarded as excellent reporter groups that monitor structural integrity of the protein (Pande *et al.*, 2000). Comparison of the fluorescence emission spectra of the P23TR36S double mutant and native HGD at excitation wavelengths of 280 nm and 295 nm were performed as a confirmation of the CD data. The fluorescence emission spectra were normalized with respect to protein concentration. The λ_{max} (emission maximum) value for P23TR36S double mutant protein was well within the range of those observed for all γ-crystallins (λ_{max} from 324 to 335 nm (Kono *et al.*, 1990)) which indicates that tryptophan residues in the mutant protein remain buried in the hydrophobic core (Mandal *et al.*, 1985). However, a red shift in λ_{max} for tryptophan fluorescence is observed in the fluorescence spectra of P23TR36S double mutant protein (figure 4.8). A similar red-shift was also reported for the R36S single mutant of HGD due to the greater solvent-exposure of the closest tryptophan (trp-42) residue to the serine (ser-36) residue (Pande *et al.*, 2001).

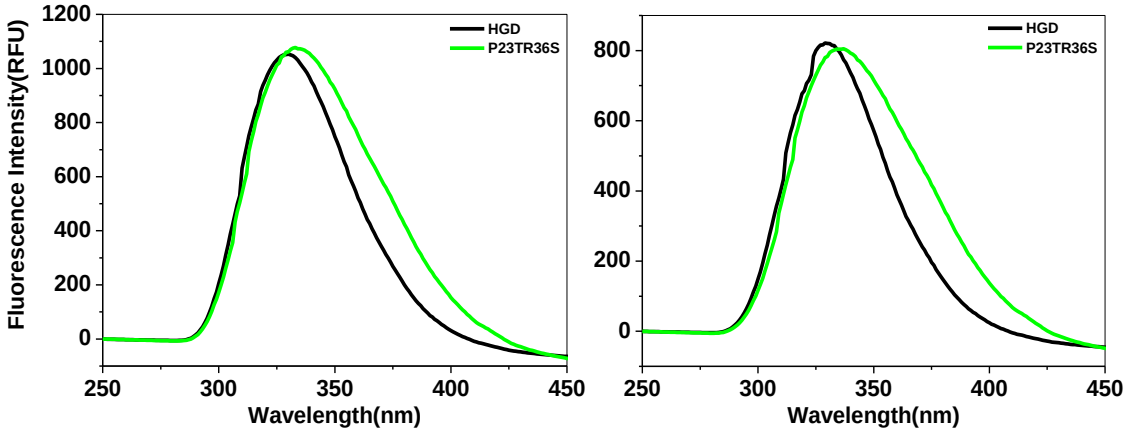

Figure 4.8: *Fluorescence emission spectra (excitation wavelength=280nm (left); excitation wavelength=295nm (right)) for P23TR36S double mutant protein in comparison with native HGD.*

4.3.2 Phase diagram for P23TR36S double mutant

The double mutant protein crystallizes spontaneously at low concentrations. Furthermore, two different crystal types are observed, depending on the temperature at which the crystals are grown (figure 4.9). A rhombic shaped crystal is formed at temperatures higher than 30°C. These crystals melt as the temperature is lowered indicating that the temperature dependence of the solubility line is inverted. At room temperature and below, a rod-shaped crystal is formed. As the solution containing these crystals is heated, the crystals melt, indicating normal solubility. The liquidus lines for both crystal types were measured (figure 4.10).

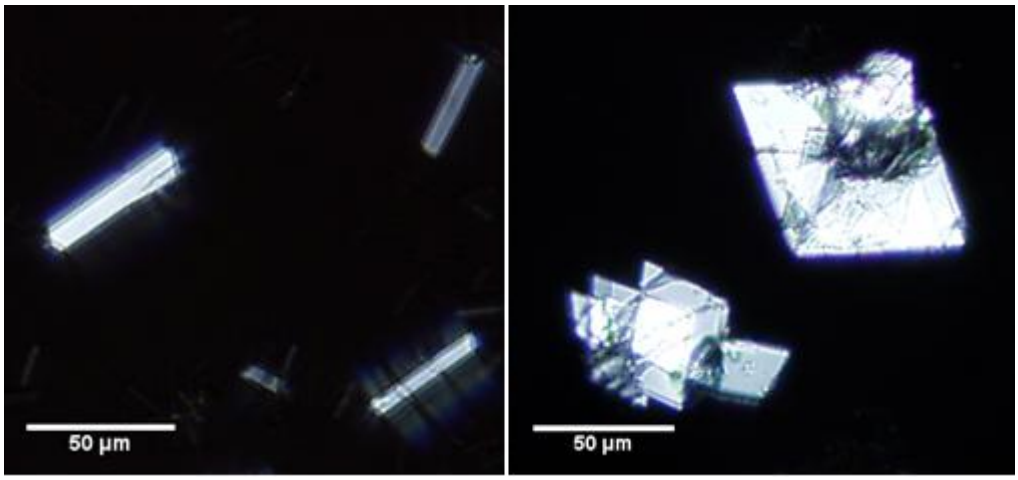

Figure 4.9: *The P23TR36S double mutant forms two different crystal morphologies for the same protein; one, a rod-shaped crystal that has normal temperature dependence and a second, with rhombic shaped crystals having inverted solubility as shown in figure 4.10 (James et al., 2015).*

There is also a temperature region ≈ 303 K (30°C) on the phase diagram (figure 4.10) of the protein where rod-shaped and rhombic-shaped crystal morphologies co-exist (figure 4.11) (James *et al.*, 2015). The P23TR36S crystals share no morphological similarities with the P23VR36S crystals. However, the P23TR36S does not form aggregates at all. This is interesting given that the P23T single mutant only formed protein *aggregates* at pH 7 (Pande *et al.*, 2005). Although a crystal structure for the P23T mutant has been described (Ji *et al.*, 2013), these crystals were formed at pH 4.5.

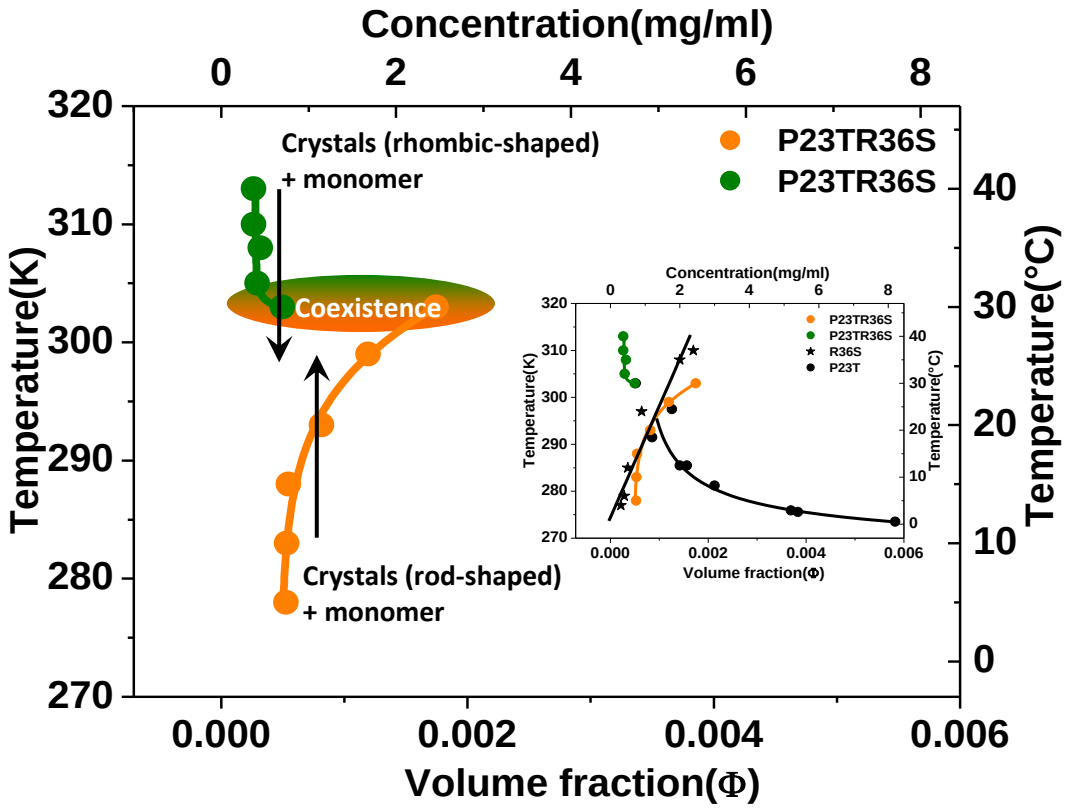

Figure 4.10: *Phase diagram showing liquidus lines for the two different crystal morphologies; rod shaped crystals with normal solubility and rhombic shaped crystals with inverted solubility, taken from (James et al., 2015). Inset shows solubility lines for the single mutants of HGD, P23T and R36S (black lines) along with the re-drawn solubility lines for the double mutant P23TR36S (green and orange lines) taken from (James et al., 2015). Data for the single mutants of HGD are taken from (Pande et al., 2000) and (Pande et al., 2005).*

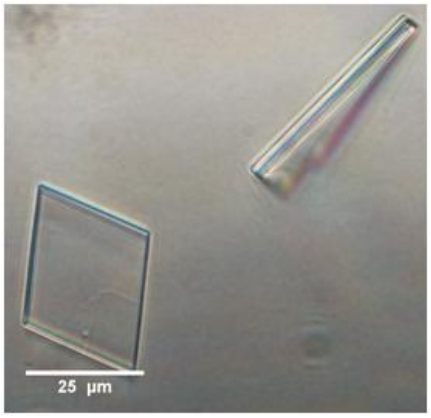

Figure 4.11: *Phase contrast microscopy image showing the coexistence of the two different crystal morphologies for P23TR36S double mutant at ≈ 303 K (30°C), taken from (James et al., 2015).*

4.3.2.1 Analysis of P23TR36S double mutant solubility data

Solubility data analysis for the P23TR36S double mutant reveals interesting information about the phase diagram for the protein. We utilize the method described

by (McManus *et al.*, 2007) to analyse the solubility data for the P23TR36S double mutant. The transfer chemical potential for the two crystal forms of the P23TR36S double mutant protein have been estimated from their liquidus lines as described previously (McManus *et al.*, 2007) (refer to section 3.3.4.2).

Quantitative estimation of the solubility data is allowed by plotting $\Delta\mu_{trans}/kT_0$ vs. T/T_0 and by determining the value of V for the proteins (Table 4.1).

Protein	S	V	K = V – S
HGD	-19	4.4	23.4
P23T	28	7.6	-20.4
R36S	-10	6.9	16.9
P23TR36S (Normal)	-7.3	6.5	13.8
P23TR36S (Inverted)	24	7.8	-16.2

Table 4.1: *Values of the parameters S, slope for $\Delta\mu_{trans}/kT_0$ vs. T/T_0 and V, the intercept at $T=T_0= 303K$ for each protein.*

The liquidus line for the rod-shaped crystals which have normal solubility is shown in figure 4.10. There is a small difference in the position of the liquidus line between the single and double mutants and the analysis indicates that the change in binding energy is ≈ $0.4kT$ at 303 K. Therefore the P23T mutation is influencing the position of the phase boundary for the rod-shaped crystals (which have a liquidus line similar to that of R36S) in the double mutant (James *et al.*, 2015).

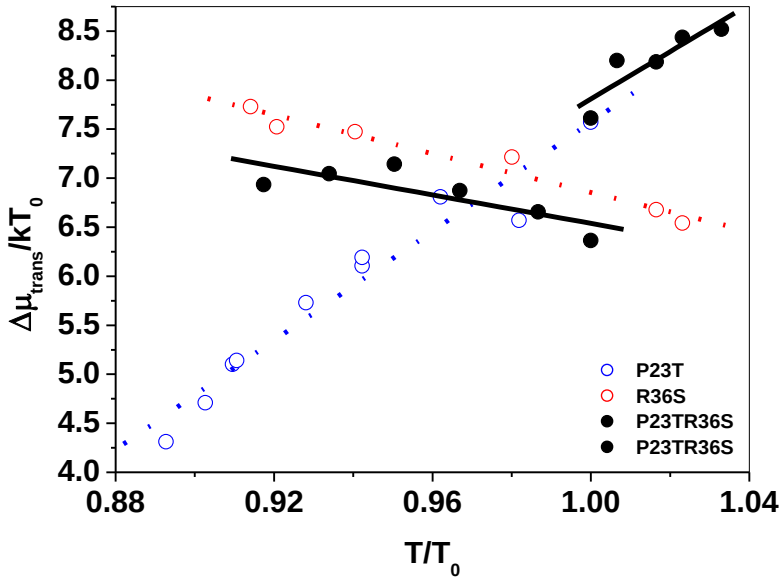

Figure 4.12: *Temperature dependence on solubility forP23T, R36S single mutants and the two crystal forms of P23TR36S double mutant of HGD, taken from (James et al., 2015).*

For the rhombic shaped crystals with inverted solubility, a small change in the binding energy ≈ 0.2 *kT* (figure 4.12) is observed when compared to the P23T single mutant. The double mutant condenses to form only crystals whereas P23T single mutant have a greater tendency to aggregate in solution. The small change in binding energy is probably insufficient to explain for the greater tendency of the double mutant to form crystals when P23T by itself has a greater tendency to form protein aggregates and suggests possible influence from the R36S substitution on crystallization behaviour of the double mutant.

4.3.2.2 Protein crystal polymorphism

A protein crystal polymorph refers to the occurrence of a protein in more than one crystalline form. Protein crystal polymorphism is the most ambiguous phenomenon frequently encountered during protein crystallization. It is influenced by a combination of various factors such as patchy protein surfaces that contribute to the formation of crystal contacts, orientation or symmetry of individual protein molecules in a unit cell space lattice, solution and temperature conditions (Pande *et al.*, 2001, James *et al.*, 2015, Lafont *et al.*, 1997, Fusco *et al.*, 2014). Polymorphic crystals that form under the same set of solution parameters (Velev *et al.*, 1998, Ataka & Asai, 1988, Vaney *et al.*, 2001, Bernardo *et al.*, 2005, Frericks Schmidt *et al.*, 2007, Carter Jr & Carter, 1979) tend to form crystal structures with similar free energies of crystallization, therefore unravelling the interactions responsible for crystallization and those required to maintain a stable polymorphic crystal form is important in understanding the occurrence of polymorphism.

Various physicochemical methods have been employed to identify solution conditions favorable for protein crystal growth and as a result many proteins have been shown to exhibit crystal polymorphism. Crystal polymorphs of aprotinin (BPTI) were reported to be stable under a variety of conditions such as pH, salt and temperature (Veesler *et al.*, 2004, Lafont *et al.*, 1997, 1994). One of the polymorphic crystals of BPTI displayed inverted temperature dependence of the solubility line (Veesler *et al.*, 2004). Proteins such as L-asparaginase have produced polymorphs that have similar space groups but different crystal lattice contacts, due to Ca^{2+} specific binding (Michalska *et al.*, 2008).

For the double mutant, P23TR36S, only a temperature change is required to transition between the two crystal polymorphs formed. Polymorphic crystal behaviour has also been previously reported in BPTI (Veesler *et al.*, 2004), the only factor controlling the transition between the two crystal polymorphs is temperature. To examine this crystal polymorphism and to analyse what factors contribute to the shifts in phase boundaries for the two crystal types of the P23TR36S double mutant in comparison to the individual single mutants, X-ray crystallographic analysis of the polymorphic crystals were performed.

4.3.3 Protein crystallography for the P23TR36S double mutant crystals

High-resolution atomic structures for native-HGD and its single mutant proteins such as P23T, R36S and R58H are published and available in the protein data bank (PDB). Here we describe for the first time, the structures of the two crystal types formed by the P23TR36S double mutant of human γD-crystallin protein measured using X-ray crystallography. Both P23TR36S double mutant protein crystals provided fully measurable sets of x-ray reflections which gave rise to complete diffraction patterns for the two crystal forms as shown in figure 4.13 and figure 4.14.

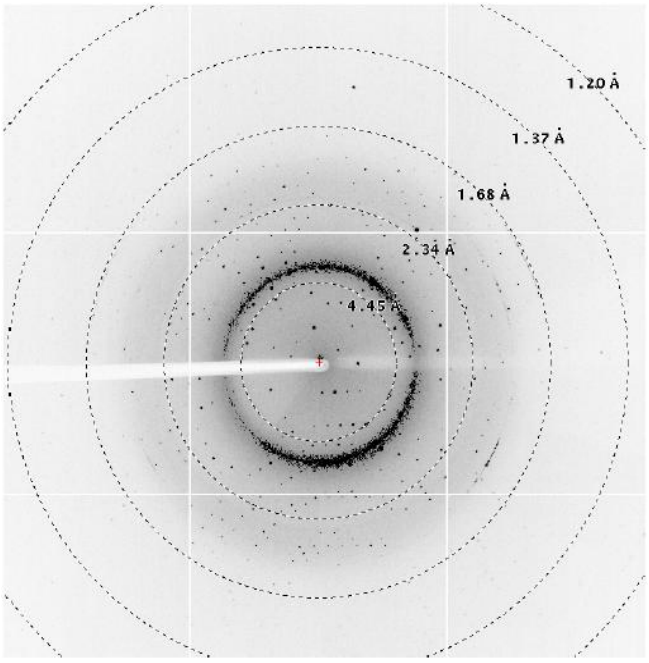

Figure 4.13: *X-ray diffraction pattern obtained for the rhombic crystal of the P23TR36S double mutant having inverted solubility.*

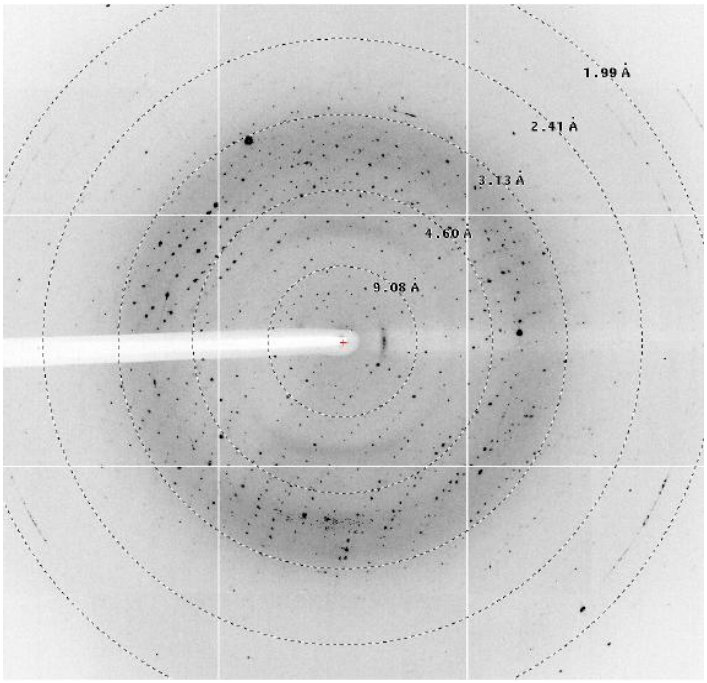

Figure 4.14: *X-ray diffraction pattern obtained for the rod-shaped crystal of the P23TR36S double mutant having normal solubility.*

4.3.3.1 Preliminary analysis

The rhombic-shaped crystal of P23TR36S double mutant with inverted solubility diffracted to a resolution of 1.24 Å and initial indexing of the diffraction pattern indicated that the rhombic crystal belonged to the P21 space group with monoclinic symmetry and unit cell dimensions of a = 44.02 Å, b = 31.7 Å, and c = 52.5 Å (α = γ = 90.00° and β = 91.29°). The rod-shaped crystal of P23TR36S double mutant with normal solubility diffracted to a resolution of 2.2 Å. Initial indexing of the diffraction patterns indicated that the rod-shaped crystal belonged to the $P2_12_12_1$ space group with orthorhombic symmetry and unit cell dimensions of a = 54.04 Å, b = 82.1 Å, and c = 106.25 Å. For the rhombic crystals with inverted solubility the space group consists of one protein molecule per asymmetric unit with an estimated solvent content of ~31% (V_m = 1.78 $Å^3$/Da) as evaluated by Matthew's coefficient (Matthews, 1968), while for rod-shaped crystals with normal solubility the space group comprises of two protein molecules per asymmetric unit with an estimated solvent content of ~57% (V_m = 2.86 $Å^3$/Da). Details of the data collected for both crystal types are shown in table 4.2.

Data Collection	Rhombic crystal	Rod-shaped crystal
Resolution range	44.01 - 1.2Å	48.16 - 2.2Å
Number of observations		
Measured	157974	164792
Unique	41264	24765
R_{merge} (within I+/I-)	0.096	0.121
< I / σI >	8.1	9.2
Completeness (%)	99.6	99.9
Multiplicity	3.8	6.7

Table 4.2: *Data collection for rhombic and rod-shaped crystals of P23TR36S double mutant.*

The corresponding data for the crystals of the single mutants (P23T, R36S) (table 4.3) is provided for comparison with the double mutant crystal types.

Data Collection	P23T crystal	R36S crystal
Resolution range	30.1 - 2.5Å	29.68 - 2.2Å
Space group	P 1 21 1	P 21 21 21
Completeness (%)	100.0	91.2
Cell dimensions	*a* = 37.55 Å	*a* = 54.38 Å
	b = 101.93 Å	*b* = 81.78 Å
	c = 41.37 Å	*c* = 106.28 Å
	α=γ= 90°,β= 96.54°	α=β=γ= 90°

Table 4.3: *Data collection for P23T and R36S single mutant crystals of HGD.*

The mosaicity value for protein crystals is another important diffraction characteristic parameter to determine crystal quality. The measurement of the rocking width of the crystal's reflections quantifies the crystal mosaicity (Wiencek, 1999). Most single protein crystals comprise a regular arrangement of the unit cells interrupted by lattice defects. The diffraction pattern for a single protein crystal can be obtained by adding the resulting diffraction patterns arising from mosaic blocks with slightly different orientations. Mosaicity values between 0.2° and 0.5° are considered ideal for good quality protein crystals as it reflects upon the success of the X-ray diffraction experimental procedure (Drenth, 2007). Both the crystal forms read low mosaicity values of 0.35° for the rhombic-shaped crystals (inverted solubility) and 0.5° for the rod-shaped crystals (normal solubility). Although a full refinement of the x-ray data has not yet been performed, some initial observations have been made. A molecular replacement with the existing crystal structures for the single mutant proteins (from the PDB) has been performed using COOT. Superposition of the double mutant structures

with their single mutant counterparts has been performed using Chimera v1.10.2 and is illustrated in figures 4.15 and 4.16.

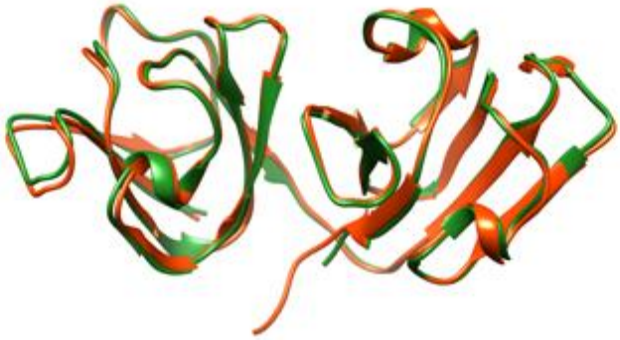

Figure 4.15: *Superposition of the structures for the P23TR36S rhombic crystal (inverted solubility) and the P23T structure from the PDB (green, PDB ID: 4JGF).*

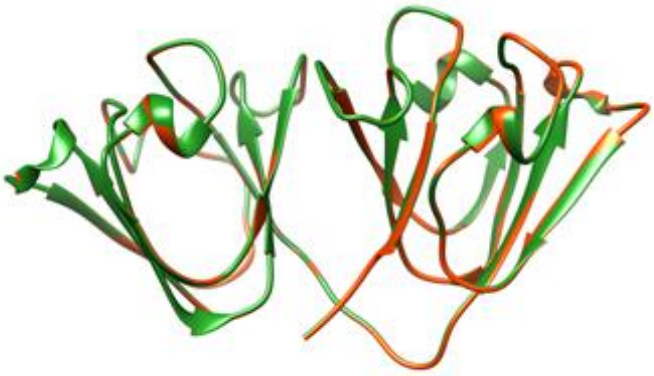

Figure 4.16: *Superposition of the structures for the P23TR36S rod-shaped crystal (normal solubility) and the R36S structures (green, PDB ID: 2G98).*

Full refinement of the crystal structures for the two crystal forms of P23TR36S double mutant are not performed here, however, initial data analysis performed by Prof. Amir Khan, Trinity College Dublin suggests small differences in the local structures between the single mutants and the two crystals formed by the P23TR36S double mutant. A symmetry map showing the closest symmetry molecules for the P23TR36S double mutant protein within a 50 Å radius is shown in figure 4.17.

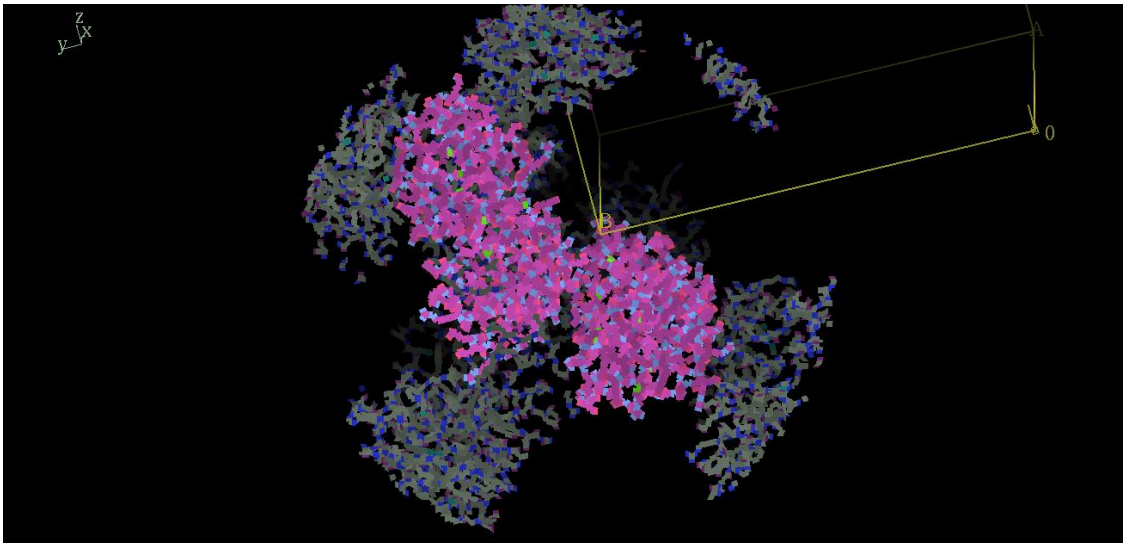

Figure 4.17: *Symmetry map showing the symmetry related molecules for the P23TR36S double mutant protein (normal solubility crystal type) within a 50 Å radius.*

4.3.3.1 Structural changes associated with the Pro to Thr substitution at site 23 for the rhombic-shaped P23TR36S double mutant crystals (inverted solubility)

The initial electron density map for the P23TR36S double mutant rhombic crystal shows well-defined electron density for the threonine residue that replaces the proline residue in the double mutant (figure 4.18). Although the equilibrium solubility line for the rhombic-shaped P23TR36S double mutant crystal with inverted solubility was shown to be qualitatively consistent with the P23T single mutant which lacks molecular contacts, the initial electron density map for the rhombic-shaped P23TR36S double mutant crystals with inverted solubility suggest the formation of a new molecular contact which is apparently absent in the P23T single mutant: the Oγ1 group of the Thr23 (due to mutation at site 23) side chain in the double mutant interacts with the NH group of the Gly129 backbone possibly via hydrogen bond formation at atomic distances of 3.28 Å. The only known crystal structure for the HGD P23V single mutant was obtained for crystals grown at pH 4.6 (Ji *et al.*, 2013). In that crystal structure, no new contact at position 23 was observed. However, the structure in that work was obtained from crystals in which the asymmetric unit was a dimer. For this new crystal structure, the crystals were formed at physiological pH and the asymmetric unit is a monomer. If the formation of a molecular contact around site 23 is confirmed after a full refinement, this could explain why the P23TR36S double mutant (rhombic-shaped with inverted solubility) only forms crystals when the P23T single mutant by itself has a greater tendency to aggregate in solution.

The structure of P23T single mutant in the PDB indicated local conformational changes such as Nε2 group of the imidazole ring of His22 forming a hydrogen bond with Tyr16 side chain and Nδ1 group of the imidazole ring of His22 interacting with the backbone carbonyl group of Ser20 through a hydrogen bond (Ji *et al.*, 2013). The above mentioned local conformational changes were almost absent in the P23TR36S double mutant rhombic crystal with inverted solubility. However, the minor structural readjustments which were part of the P23T single mutant around the Trp42 side chain were retained in the double mutant with no major changes in the molecular contact distances. Thus, it can be inferred that the formation of the new intermolecular contact around Thr23 in the double mutant has most likely contributed to a more tight crystal packing in the rhombic crystal with reduced solvent content of ~31%. This low solvent content in the double mutant crystal is noteworthy.

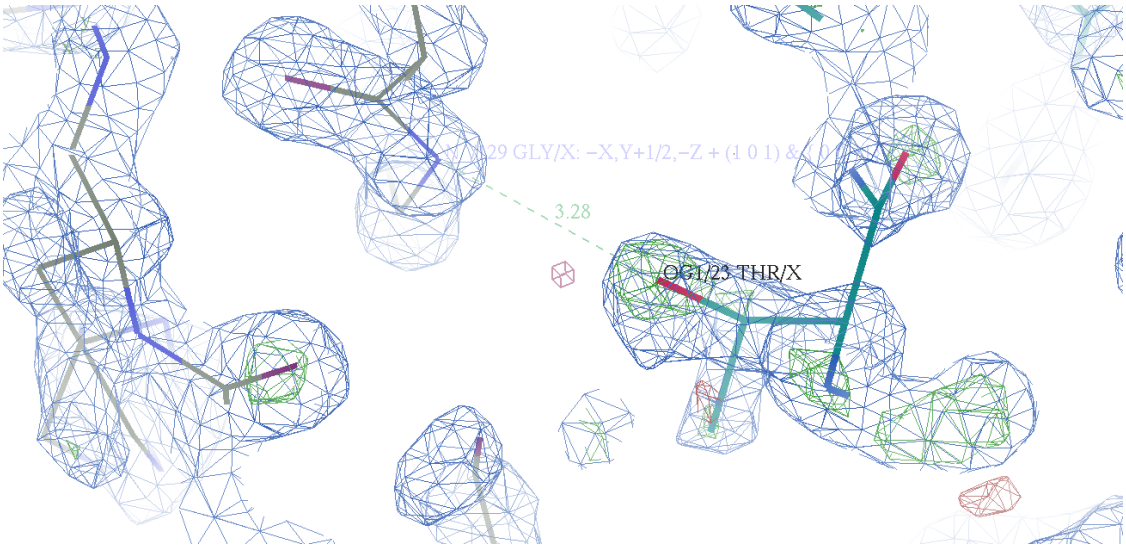

Figure 4.18: *Electron density map showing the intermolecular lattice contacts formed in the rhombic shaped P23TR36S double mutant with respect to the mutation at site 23.*

4.3.3.2 Structural changes associated with the Arg to Ser substitution at site 36 for the rhombic-shaped P23TR36S double mutant crystals (inverted solubility)

The initial electron density map around site 36 for the P23TR36S double mutant rhombic crystal suggests that electron density is poor around the region of mutation where the serine residue replaces the arginine residue in the double mutant (figure 4.19). However, the same mutation when present in the single mutant has a very well-defined electron density map. Also, unlike the single mutant for which substituting arginine by serine at site 36 results in the formation of a crystal contact, the same mutation in the P23TR36S double mutant for the rhombic crystals occur with the loss

of this specific crystal contact. The electron density map of R36S single mutant suggests that Oγ group of the Ser36 side chain of molecule B interacts with the Nδ2 group of the Asn24 side chain of molecule B at distances of 2.78 Å (Kmoch *et al.*, 2000). However, this interaction at short molecular distances is not present in the P23TR36S double mutant rhombic crystal with a mutation at site 36. For the rhombic crystal at site 36, there is likely no molecular contact as the Oγ group of the Ser36 side chain is at a long distance from the nearest NH1 group of the Arg140 side chain (figure 4.19).

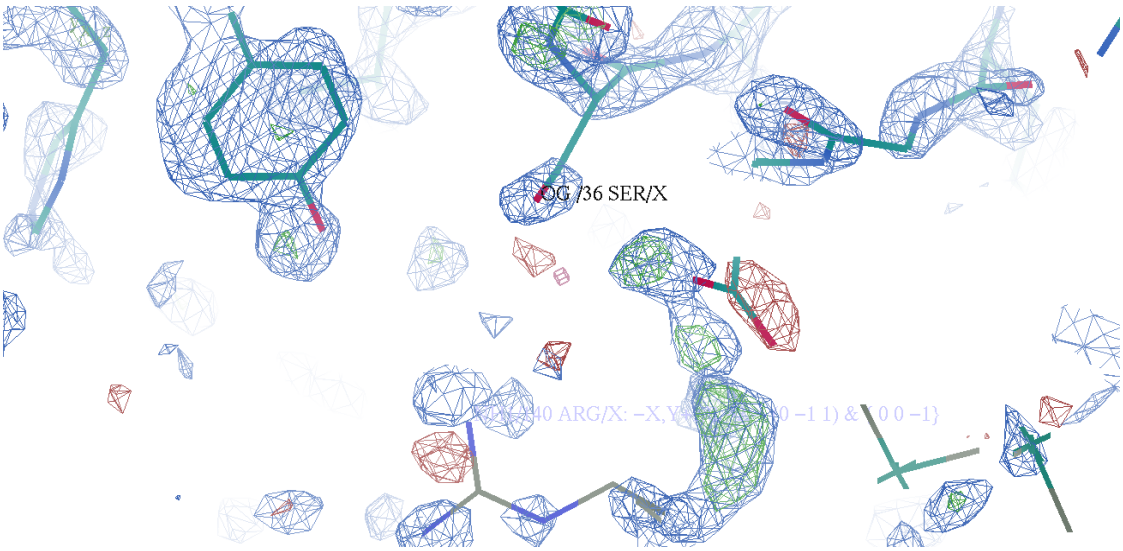

Figure 4.19: *Electron density map showing the lattice contacts formed in the rhombic shaped P23TR36S double mutant with respect to the mutation at site 36.*

4.3.3.3 Structural changes associated with the Arg to Ser substitution at site 36 for the rod-shaped P23TR36S double mutant crystals (normal solubility)

The initial electron density map for the P23TR36S double mutant rod-shaped crystal shows well-defined electron density for the serine residue that replaces the arginine residue in the double mutant (figure 4.20). The equilibrium solubility line for the rod-shaped P23TR36S double mutant crystal with normal solubility was shown to be qualitatively consistent with the R36S single mutant. The R36S single mutant forms a crystal contact around site 36 (the Oγ group of the Ser36 side chain of molecule B interacts with the Nδ2 group of the Asn24 side chain of molecule A at distances of 2.78 Å) in the single mutant (Kmoch *et al.*, 2000). In the double mutant P23TR36S which forms the rod-shaped crystal, the presence of the same crystal contact as that of the R36S single mutant around site 36 is observed except at slightly shorter molecular distances (the Oγ group of the Ser36 side chain of molecule B interacts with the Nδ2

group of the Asn24 side chain of molecule B at distances of 2.57 Å (figure 4.20)). The slightly shorter molecular distance suggest that the crystal packing around site 36 remains intact in the double mutant and a comparatively tighter packing with no major differences between the structures for the rod-shaped crystal and the R36S single mutant crystal. This confirms that the R36S substitution is the driving force for the formation of the rod-shaped crystals for the P23TR36S double mutant.

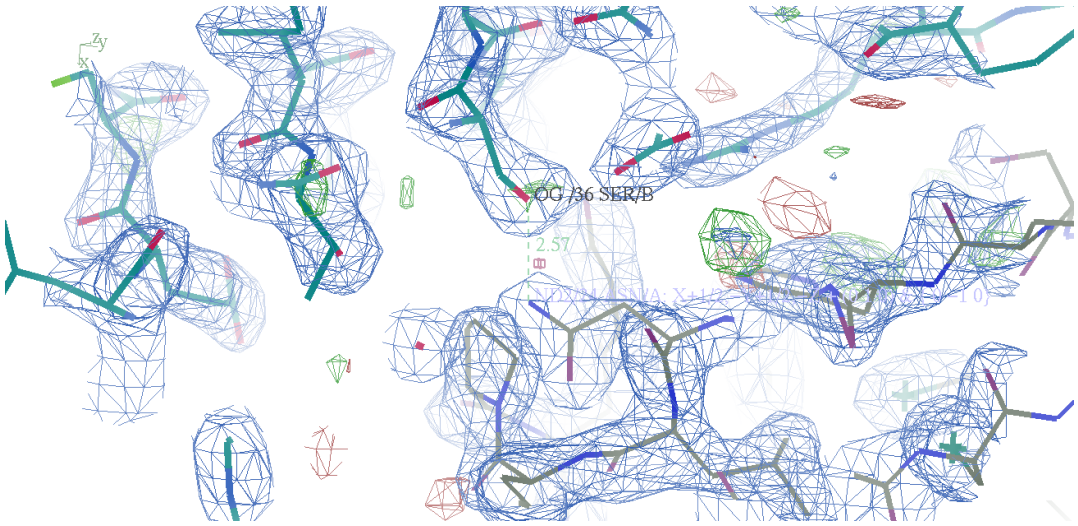

Figure 4.20: *Electron density map showing the lattice contacts formed in the rod-shaped P23TR36S double mutant with respect to the mutation at site 36 on molecule B.*

4.3.4 Conclusion

Creation of the first double mutant, P23VR36S, was the initial step towards protein phase diagram prediction from phase behaviours of the individual single mutants. However, it was understood that P23VR36S double mutant by itself was not sufficient enough to validate the concept of protein phase diagram prediction. Hence, the second double mutant, P23TR36S was created for further investigative analysis of the concept. P23TR36S double mutant proteins are soluble following extraction and purification procedures, they retain their native structure upon mutagenesis confirmed by spectroscopic techniques and most significant of all, they formed crystals with two different morphologies for the same protein; one, a rod-shaped crystal that has normal temperature dependence on solubility and a second, with rhombic shaped crystals having inverted solubility (James *et al.*, 2015).

X-ray crystallographic analysis of the two crystal morphologies reveals interesting observations about the P23TR36S double mutant. The formation of the new molecular contact around site 23 for the rhombic crystals of P23TR36S double mutant could

possibly explain for the formation of crystals when mutation around the same region in the single mutant has a greater tendency to form protein aggregates. Secondly, the formation of the P23TR36S rod-shaped crystal with normal solubility whose behaviour is purely driven by the R36S substitution can be inferred from the initial structural analysis of the mutation around site 36 for the double mutant. Thus, crystallographic evidence provides supporting evidence in addition to the phase boundaries for the dramatic phase transitions pertaining to the P23TR36S double mutant. A table summarizing the structural changes associated with the single and double mutants is provided below (table 4.4).

	P23T crystal	**R36S crystal**	**Double mutant (rhombic-shaped)**	**Double mutant (rod-shaped)**
Site 23	Crystal contact absent in the single mutant.	Crystal contact absent in the single mutant.	Molecular contact formed in the double mutant.	Molecular contact absent in the double mutant.
Site 36	Crystal contact absent in the single mutant.	Crystal contact formed in the single mutant.	Molecular contact absent in the double mutant.	Molecular contact formed in the double mutant.

Table 4.4: *Table summarizing the structural changes associated with the single and double mutants of HGD due to mutations at site 23 and 36.*

Chapter 5

Phase diagram and aggregate growth rate for the HGD P23VR58H double mutant

5.1 Introduction

5.1.1 Background

The third double mutant type, P23VR58H, was created by incorporating the P23V amino acid substitution associated with inverted solubility and the R58H substitution associated with crystallization and normal solubility. This double mutant is formed by replacing proline with valine at position 23 on the amino acid chain of native HGD protein exhibiting inverted temperature dependence on solubility *i.e.* the protein forms condensed phases (crystals/aggregates) at higher temperatures and melts as the temperature is lowered (Pande *et al.*, 2005, McManus *et al.*, 2007). The other single amino acid substitution in HGD (R58H) is formed by replacing arginine with histidine at position 58 on the amino acid chain of native HGD. The R58H single mutant, similar to the R36S single mutant crystallizes with an almost identical liquidus line (Pande *et al.*, 2001). However, the difference in their crystallization kinetics (slow crystallization) makes the R58H single mutant 100 times more soluble than the R36S single mutant (Pande *et al.*, 2001). The R58H single mutant is responsible for the formation of aculeiform cataract (Basak *et al.*, 2003, Pande *et al.*, 2001).

The R58H single mutant is a crystal contact point for the native protein but the strength of the intermolecular ion-pair interaction at this position decreases upon mutagenesis (Basak *et al.*, 2003). In the native protein, aspartate at position 156 for native HGD forms strong intermolecular ion-pair interaction with the arginine at position 58, mediated by direct hydrogen bonds (Basak *et al.*, 2003). However, substituting arginine with histidine in the single mutant, results in lessening the strength of the strong, protein-protein intermolecular interaction at the mutation site (Basak *et al.*, 2003). The solubility of the mutant protein is lower than that of the native version and the protein crystallizes, and the liquidus line for the R58H mutant is very similar to that of the R36S variant. However, the kinetics of crystallization is very different. The R58H mutant crystallizes far more slowly than the R36S and it can be concentrated to over 100mg/ml before crystallization occurs. This chapter provides a detailed description on the creation of a third double mutant P23VR58H (figure 5.1) and exploring the phase diagram for the newly formed double mutant.

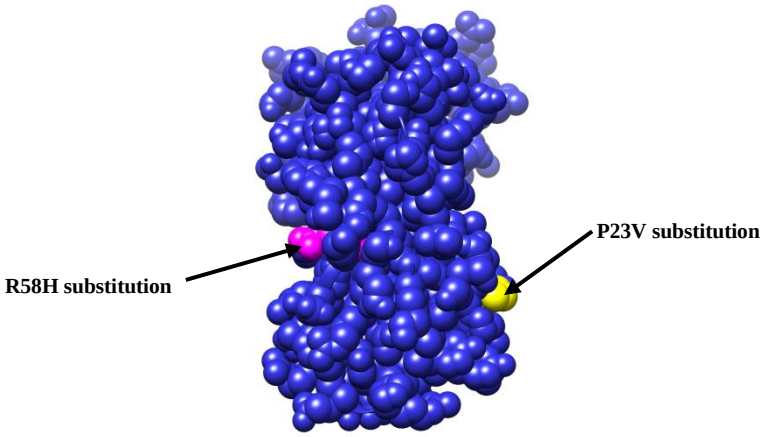

Figure 5.1: *3D representation of the two amino acid substitutions (P23V and R58H) incorporated in the native HGD protein to create the P23VR58H double mutant.*

5.2 Aim of the study

The creation of the third double mutant, P23VR58H incorporating the P23V and the R58H single mutants allowed us to access the higher concentration region of the phase diagram where the inverted solubility due to the P23V single mutant is observed which was inaccessible for the first double mutant P23VR36S (in chapter 3).

5.3 Results and Discussions

5.3.1 Production and characterisation of P23VR58H double mutant protein

The P23VR58H double mutant recombinant DNA was sequenced with the T7 promoter and confirmed to include the two single amino acid substitutions, proline (CCC) to valine (GTC) at position 23 and arginine (CGC) to histidine (CAC) at position 58 within the double mutant.

Figure 5.2: *Plasmid DNA sequence for P23VR58H double mutant of HGD aligned and compared against native HGD using BLAST.*

The purity and DNA concentration for the P23VR58H double mutant was measured to be 1.84 (A_{260nm}/A_{280nm}) and 80.8 ng.µl^{-1}, respectively.

Following size exclusion chromatography (figure 5.3), SDS-PAGE identified the mutant protein peak (figure 5.4). These fractions were pooled and further purified by ion exchange chromatography, producing a well purified product (figure 5.5).

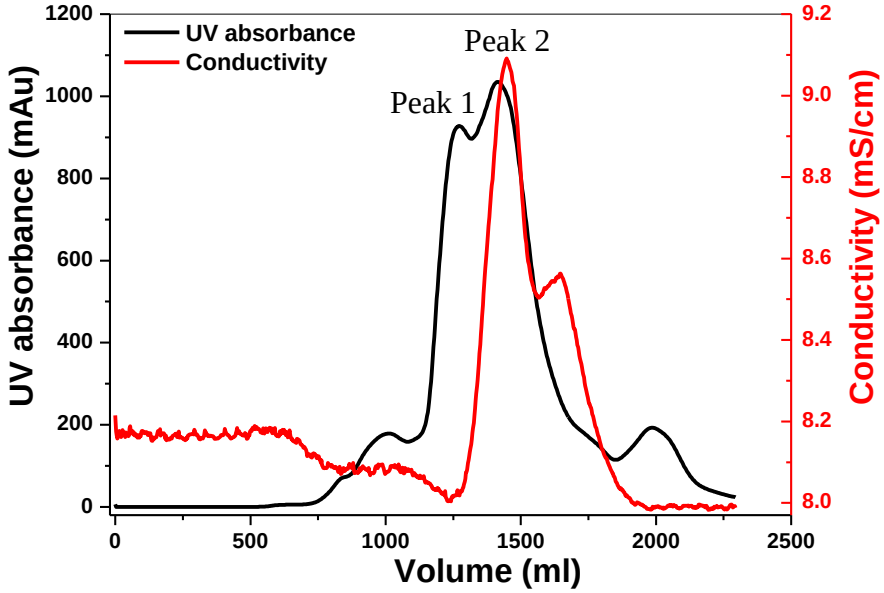

Figure 5.3: *Size-exclusion chromatogram depicting the primary purification step for P23VR58H double mutant of HGD.*

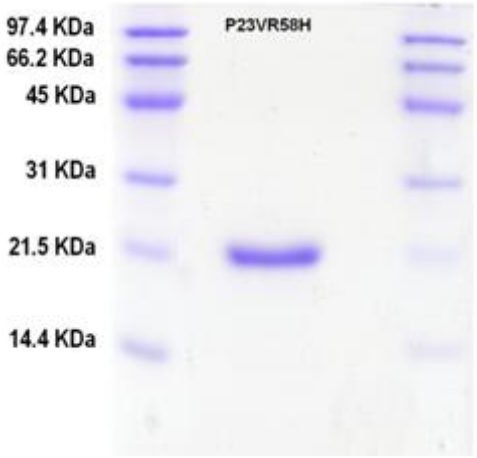

Figure 5.4: *12.6% reducing SDS-PAGE gel showing a single monomer band confirming the purity of P23VR58H double mutant protein.*

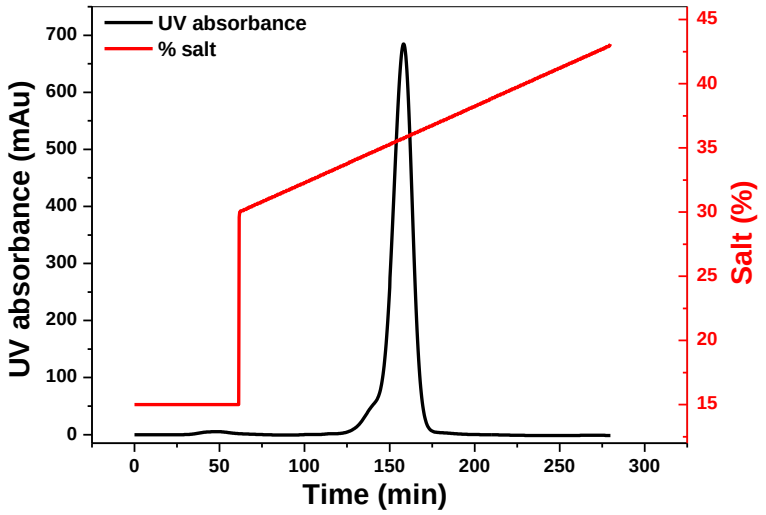

Figure 5.5: *Cation-exchange chromatogram depicting clean elution profile for P23VR58H double mutant.*

The P23VR58H double mutant protein has a molecular mass of 20,589 ± 1 Da, determined by intact molecular weight analysis by electrospray ionization mass spectrometry. This molecular mass for P23TR36S double mutant is consistent with the expected mass value for the specified amino acid substitutions. Chromatographic analysis by SE-HPLC data confirmed that the P23VR58H double mutant protein is mostly monomeric (≥ 99% purity) and shows a major peak at 12.53 min (figure 5.6). From the calibration of the Superdex 75 10/300 GL SE-HPLC column used to isolate the protein, a peak at 12.53 min indicates a protein molecular weight in the region of 20.5 KDa, which is within the range expected for P23VR58H.

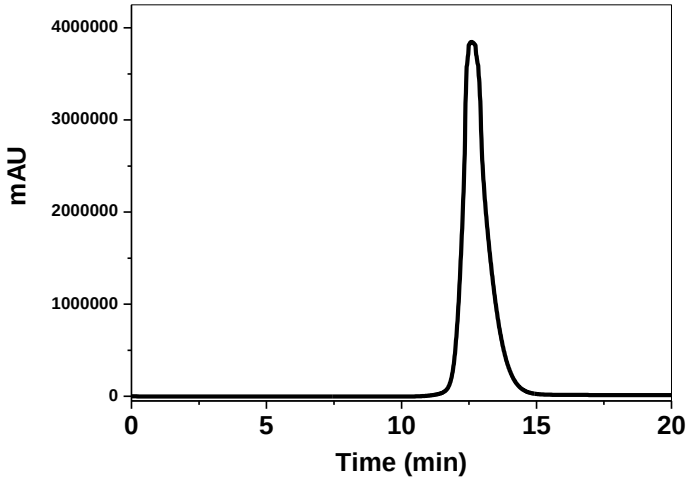

Figure 5.6: *SE-HPLC profile for the P23VR58H double mutant when freshly prepared in 0.1M sodium phosphate buffer pH 7.0 in the presence of 20mM DTT.*

5.3.1.1 Structural studies of P23VR58H double mutant protein using circular dichroism spectroscopy and fluorescence spectroscopy

Circular dichroism spectra. The CD spectral data of P23VR58H double mutant and native HGD were compared to examine if structural and conformational changes occurred upon mutagenesis. The far-UV and near-UV CD spectral analysis of the P23VR58H double mutant indicates that it has very similar secondary and tertiary structures to native HGD. Both the P23VR58H double mutant and the native HGD show a negative ellipticity at around 218 nm on the far-UV CD spectra and also display similar spectral features on the near-UV CD spectra for the proteins (figure 5.7).

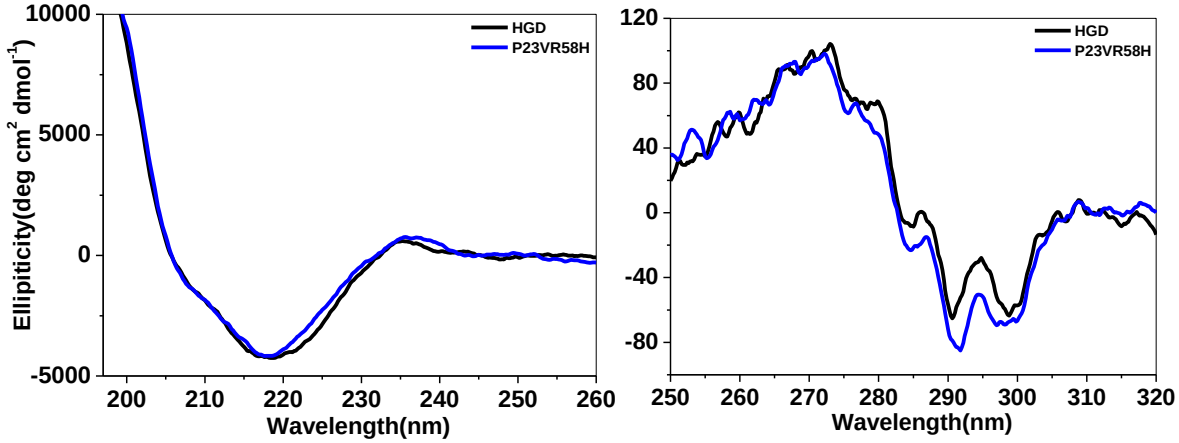

Figure 5.7: *Comparison between secondary and tertiary structures of P23VR58H double mutant and native HGD using far-UV CD spectra (left) and near-UV CD spectra (right).*

Fluorescence spectra. The fluorescence emission spectra of the P23VR58H double mutant and native HGD at excitation wavelengths of 280 nm and 295 nm were compared to confirm if protein structural change occurs upon mutagenesis. The fluorescence emission spectra were normalized with respect to protein concentration. The λ_{max} (emission maximum) value for P23VR58H double mutant protein (figure 5.8) was well within the range of those observed for all γ-crystallins (λ_{max} from 324 to 335 nm (Kono *et al.*, 1990)) which indicated that tryptophan residues in the mutant protein remain buried in the hydrophobic core (Mandal *et al.*, 1985).

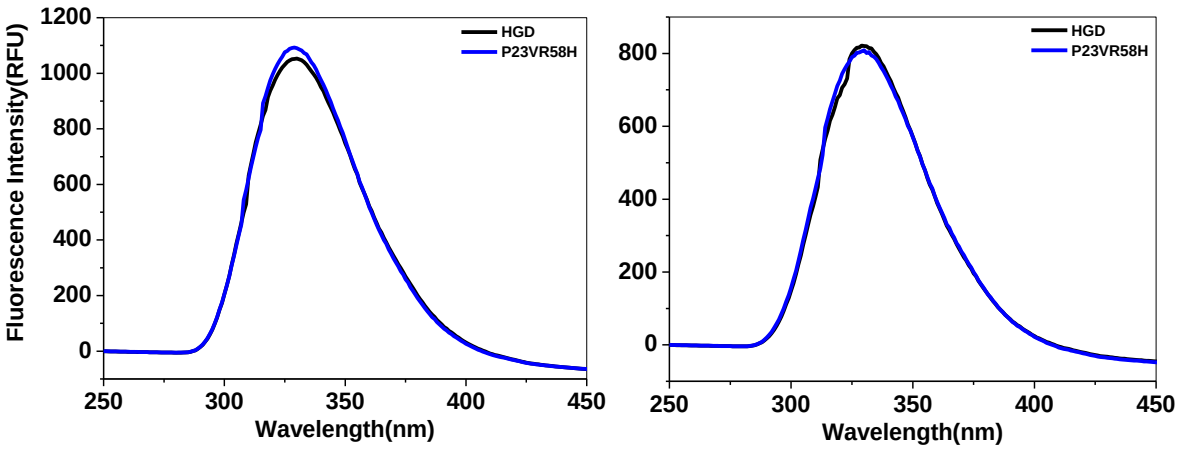

Figure 5.8: *Fluorescence emission spectra (excitation wavelength=280nm (left); excitation wavelength=295nm (right)) for P23VR58H double mutant protein in comparison with native HGD.*

5.3.2 Phase diagram (solubility line) for P23VR58H double mutant

The P23VR58H double mutant principally forms large amorphous protein aggregates as the protein is concentrated at temperatures above the expected solubility line for the P23V single mutant (James *et al.*, 2015). These aggregates have an inverted temperature dependence of the solubility line. Once the protein aggregates begin to form, the concentration of protein monomers in solution decreases over time (6-24 hours). After this period of time, crystal formation on the surface of the aggregates is observed (figure 5.9). There are no crystals formed in the bulk solution. The monomer concentration in the supernatant is sufficiently high after the aggregate formation is complete to allow crystals to form, if we compare the solution conditions to those where crystallization of the R58H single mutant occurs (Pande *et al.*, 2001).

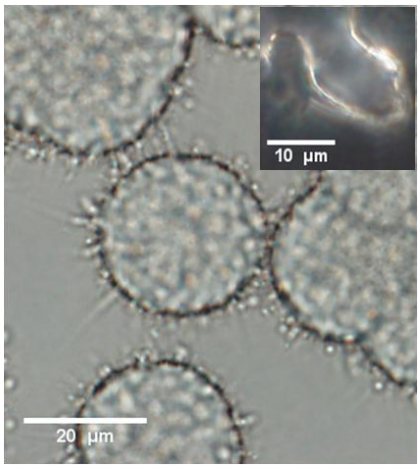

Figure 5.9: *Phase contrast microscopy image of P23VR58H protein aggregates encased in protein crystals. Inset shows polarization microscopy image of the same mutant confirming the crystalline nature of the protein surface structures, taken from (James et al., 2015).*

When we attempt to measure the solubility line for this double mutant, we find that the crystals and aggregates co-exist with the protein monomers in solution. Hence, the measured solubility line is the equilibrium established between solid (crystal + aggregates) and liquid (protein monomers). This 'equilibrium solubility line' occurs at concentrations between those expected for the two individual single mutants as shown in figure 5.10 (James *et al.*, 2015). The phase behaviour of the double mutant appears to be controlled by the kinetics of the crystallization process. The relative concentrations of the solid condensed phases *i.e.* crystals or aggregates formed are defined by the phase boundaries of the initial single mutants and the rate at which these condensed phases form (James *et al.*, 2015).

The position of the solubility line shifts depending on the temperature at which initial aggregate formation occurs. As shown in figure 5.11, a greater amount of aggregation occurs at temperatures higher than 25°C which shifts the solubility line to higher concentrations, closer to the solubility line for P23V single mutant with inverted solubility. When lesser aggregation occurs at lower temperatures, increasing numbers of crystals are formed and the solubility line moves to lower concentrations, closer to the solubility line for the R58H single mutant with normal solubility (James *et al.*, 2015).

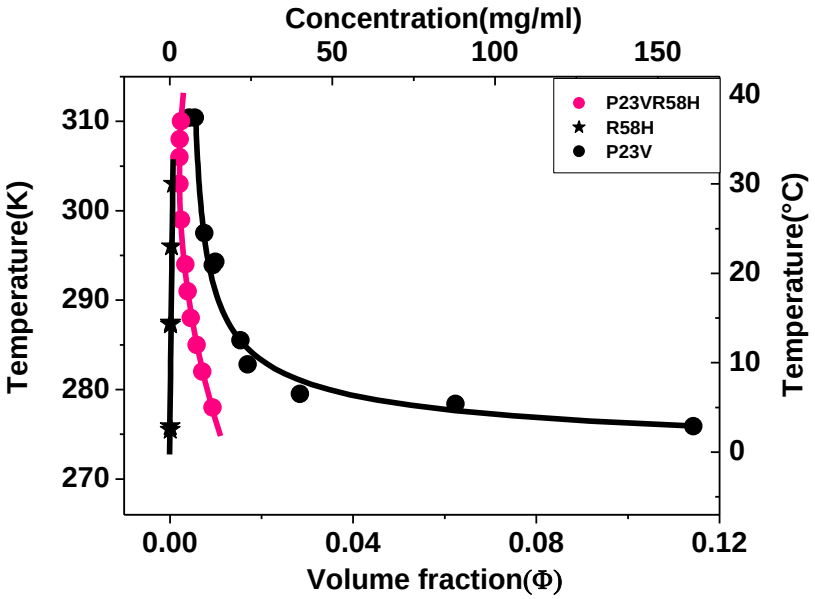

Figure 5.10: *Phase diagram showing equilibrium solubility line for the co-existing aggregates and crystals with the P23VR58H protein monomers, taken from (James et al., 2015). Data for the single mutants of HGD are taken from (Pande et al., 2000) and (Pande et al., 2005).*

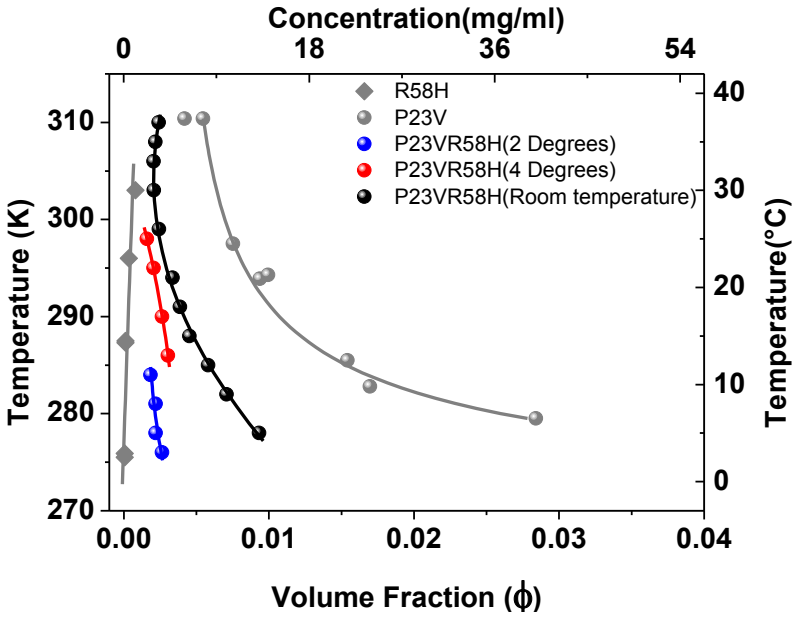

Figure 5.11: *Shifts in the solubility line depending on the temperature at which initial aggregate/crystal formation occurs (James et al., 2015).*

5.3.2.1 Aggregation and crystallization in the P23VR58H double mutant

The solubility line for the P23VR58H double mutant represents the equilibrium between three co-existing protein phases; monomers, aggregates and crystals. Given this complexity, this 'equilibrium solubility line' can only be used for a qualitative analysis. The most interesting observation from the microscopy experiments reveals that both crystals and aggregates co-exist with protein monomers in solution (figure 5.9). The formation of crystals was confirmed by polarization microscopy (figure 5.9). According to basic thermodynamic principles, crystals should have the lowest free energy and form the most stable solid phase. However, our experiments suggest that amorphous aggregates formed by the P23VR58H double mutant are as thermodynamically stable as the crystals formed by this same mutant (James *et al.*, 2015). The solubility boundary for the P23VR58H double mutant is established from the initial condensed phase that forms at a particular temperature and there is no change in the position of the phase boundary once initial aggregate formation has occurred and has been established over a few days. Nonetheless, the solubility boundary can be shifted to lower concentrations by decreasing the temperature at which initial aggregate formation occurs, thereby decreasing the concentration of aggregated material in the solid condensed phase before crystallization begins on the surface of the aggregated material as shown in figure 5.11. Higher temperatures result in a greater number of aggregates formed (thereby sequestering a greater proportion of

the protein available in solution), whereas lower temperatures suppress aggregate formation and allows greater numbers of crystals to form, resulting in phase boundaries shifting to lower concentrations approaching the phase boundary of R58H single mutant as shown in figure 5.11 (James *et al.*, 2015). Crystallization does not occur in the bulk of the protein solution; however, nucleation leading to crystallization appears to occur heterogeneously on the surface of these pre-formed amorphous aggregates. This sequence of events is possible because crystallization in the R58H mutant is delayed, and allows a range of concentrations to be reached where the aggregation process can precede crystallization when the solution conditions favour this. In this way, the kinetics of the crystallization process due to the R58H substitution drives the phase behaviour of the double mutant. Therefore the rate of formation of each condensed phase affects the position of the phase boundary for the P23VR58H mutant protein.

5.3.3 Kinetics of Aggregate Growth

The aggregation for the P23VR58H double mutant at a given concentration begins in a monomeric protein solution, devoid of aggregates, incubated at temperatures above the phase boundary for the P23V single mutant. The morphology of the aggregates formed is very unusual. The aggregates are very large and spherical, both for the single and double mutant proteins (figure 5.12 and figure 5.13). To probe the growth of these aggregated particles, we took time resolved images during the growth of the aggregates, up to and beyond the formation of crystals on the surface of the particles. Given the diffraction limit of optical microscopy, only particles of 1μm and larger were measured.

The growth profile for the P23VR58H double mutant protein aggregates show amorphous aggregates that grow as three dimensional spherical structures as shown in figure 5.12. We track this growth at regular time points after they reach an initial size of 1μm, until there is no further increase in the diameter of the particles. After this growth is complete, crystal growth occurs on the surface of the pre-formed aggregates. The different growth stages of P23VR58H double mutant aggregate for a protein concentration of 10 $mg.ml^{-1}$ are as displayed in figure 5.12. These aggregates were grown from monomeric protein solution to which 20mM DTT (Dithiothreitol) was added prior to inducing aggregation by heating to 30°C (to suppress disulfide linked dimer formation). For these solution conditions, at time t = 127min, the initial clusters

with a size of 5.7µm appeared followed by rapid growth into large spherical amorphous aggregates with a size of 42.3µm at time t = 489 minutes, which doesn't increase any further up to 1439 minutes. Crystallization on the surface of the protein aggregates only occurs after the aggregates have stopped growing beyond the time point of t = 489 min. This suggests that the aggregation and crystallization processes occur independently and that the growth of the aggregate is not halted by the onset of crystallization.

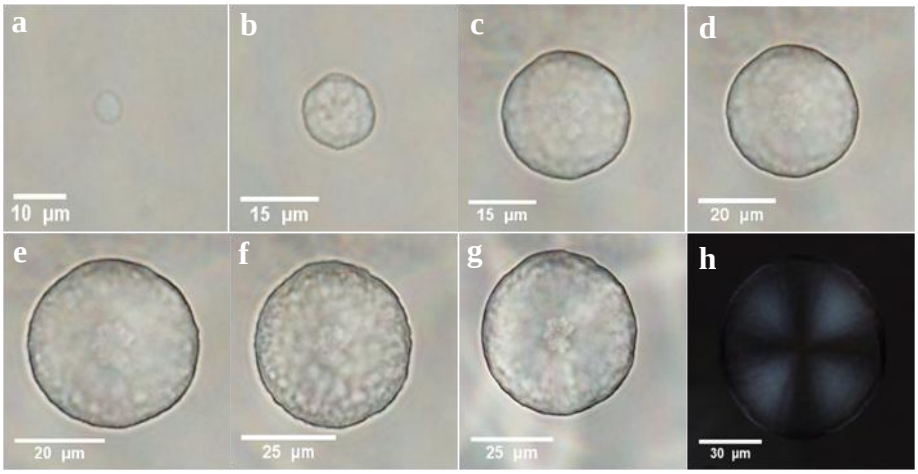

Figure 5.12: *Growth of P23VR58H aggregates monitored over time at 30°C using phase contrast (a-g) and polarization microscopy (h). Each panel indicates a different time point: 127 min (a), 193 min (b), 282min (c), 332 min (d), 383 min (e), 489 min (f) and 1439 min (g and h). Polarization microscopy image indicates optical birefringence, indicative of ordered structures, most probably due to crystallization on the surface of the P23VR58H protein aggregate beyond 489 mins.*

To confirm this, we have also compared the growth of the P23VR58H double mutant aggregates with the growth of P23V single mutant aggregates. Both the double and single mutants give rise to the spherical aggregates which are similar in their morphological appearances and both exhibit inverted solubility. As shown in figure 5.13, the P23V aggregates resemble the P23VR58H double mutant aggregates except for the absence of crystallization on the surface of pre-formed aggregates as confirmed by the polarization microscopy image (figure 5.13(e and h)). This suggests that the crystallization behaviour on the surface of P23VR58H aggregates is only present due to the R58H substitution in the double mutant protein (James *et al.*, 2015). The P23V single mutant is known to form both crystals or aggregates and the solubility/liquidus lines for both condensed phases overlap (McManus *et al.*, 2007).

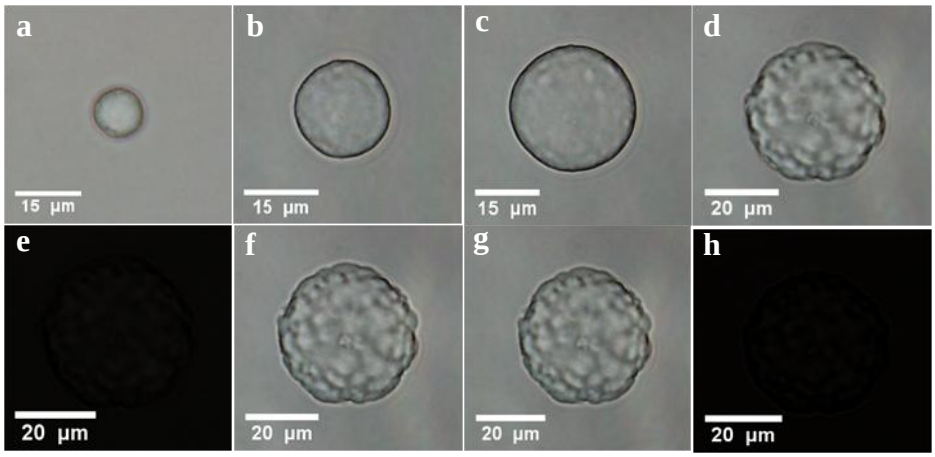

Figure 5.13: *Growth of the P23V aggregate monitored over time at 30°C from a solution at 10 mg/ml in 0.1M sodium phosphate buffer using phase contrast (a-g) and polarization microscopy (h). 187 min (a), 272 min (b), 434min (c), 1204 min (d and e), 1336 min (g and h). Polarization microscopy image (e and h) confirms absence of surface crystallization.*

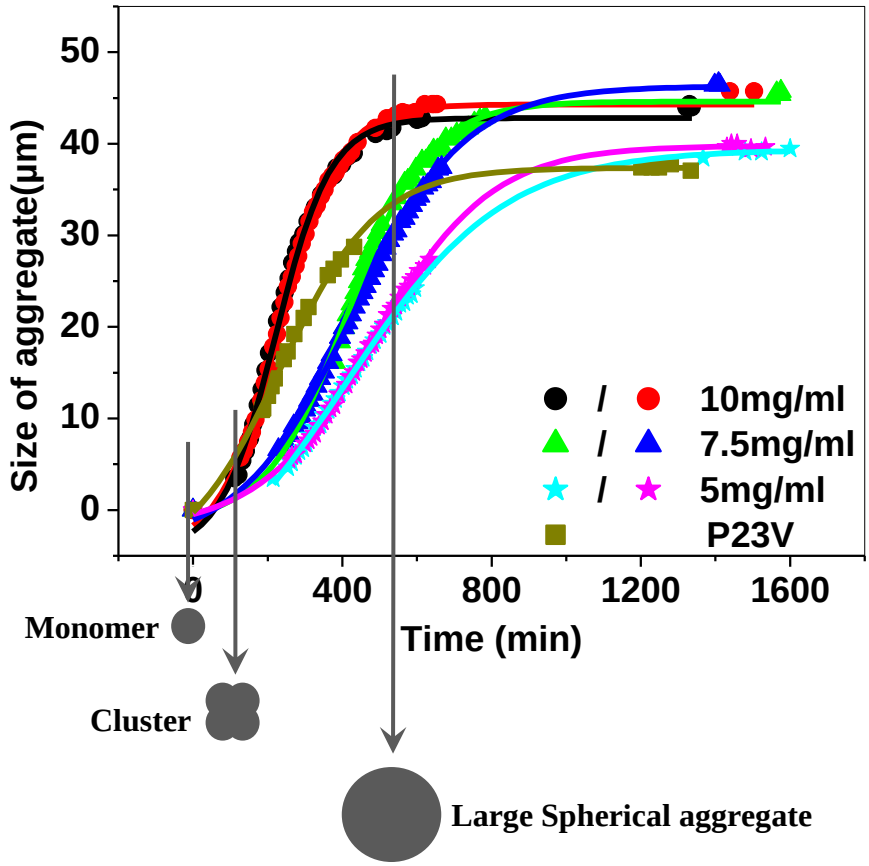

Figure 5.14: *Graphical representation for the sigmoidal growth pattern of P23VR58H double mutant aggregates.*

The evolution of the growth of amorphous aggregates at different concentrations of the double mutant protein P23VR58H and also for the single mutant protein P23V at 10 $mg.ml^{-1}$ is shown in figure 5.14.

The data indicates that, as for most protein aggregation phenomena, the growth of protein aggregates follow a sigmoidal curve (Frieden & Goddette, 1983, Flyvbjerg *et al.*, 1996, Fink, 2006, Kamihira *et al.*, 2000). While the growth kinetics of amyloid forming protein aggregates have been studied in great detail (Lomakin, Chung, *et al.*, 1996, DuBay *et al.*, 2004, Cohen *et al.*, 2013, Knowles *et al.*, 2007), the growth of amorphous protein aggregates has received far less attention (Stranks *et al.*, 2009). As such, there are fewer kinetic models describing the growth process for amorphous protein aggregation. Furthermore, amorphous protein aggregates usually form fractal structures (Aymard *et al.*, 1997), whereas the structures formed here are unusually spherical. Therefore detailed information about the aggregation pathway that can be extracted from our initial experimental measurements is somewhat limited.

If we follow the growth of the aggregates, there is a period of time for which we do not see particles (although we know that growth is occurring during this time period). This is not a lag phase, but rather a period of time for which we do not have data. Once particles of 1μm in size appear, there is an exponential increase in the size of the aggregates and a subsequent plateau where the growth ceases and the aggregates remain at the same size. Once the aggregate growth has terminated, for all concentrations of protein the maximum diameter of the spherical aggregates was 40-50μm. Depending on the initial monomer protein concentration, there is a shift in the time it takes to see the appearance of 1μm sized particles. This concentration dependence on the initial growth of aggregates is consistent with a diffusion limited mechanism of particle growth (Witten & Sander, 1983).

5.3.3.1 Rate of formation of initial clusters

We define the second stage of growth to be the formation of 1 micron-sized particles with size ranging from 1-5 μm. This is initiated in a supersaturated solution, at temperatures above the phase boundary for the mutant (≥ 30°C) (figure 5.10). We speculate that protein monomers of the double mutant protein, P23VR58H, begin to form small nuclei due to concentration fluctuations (via diffusion) and that by a process of monomer addition grow into larger clusters over time. The complexity of studying nucleation event arises from the fact that nuclei typically comprises of 100-1000 atoms which are inaccessible to current experimental methods (Schüth *et al.*, 2001). A few microscopic techniques allow detection without distinction of the sub-microscopic nuclei (Yau & Vekilov, 2001). The formation of nuclei from a metastable

solution is referred to as nucleation and represents a first-order phase transition. The growth stage leading up to the formation of aggregates is governed by diffusion of protein monomers and addition of growth units to the surface of the existing critical nuclei (Stranks *et al.*, 2009).

The rate of formation of clusters with a size of 1μm was investigated for the different P23VR58H double mutant concentrations (figure 5.15). This is taken as the time between the change in temperature to 30°C and the point at which the first particles of 1μm in size appear. What is particularly interesting is that although only single particles are followed throughout the growth process, all of the particles in a particular solution grow at the same rate. There is remarkable homogeneity in the distribution of particle sizes throughout the aggregating solution. With increasing concentration, the time taken for a 1μm sized particle to appear decreases (again consistent with a diffusion driven nucleation mechanism). At lower protein concentrations, it takes longer to achieve thermodynamic stability by the formation of these micron-sized clusters and this is not surprising. As the protein concentration is lowered, there are fewer protein monomers randomly moving around in the solution volume to add onto the surface of the nuclei. Thus, it takes longer to both form clusters at lower protein concentrations and longer for the aggregates to grow (figure 5.15).

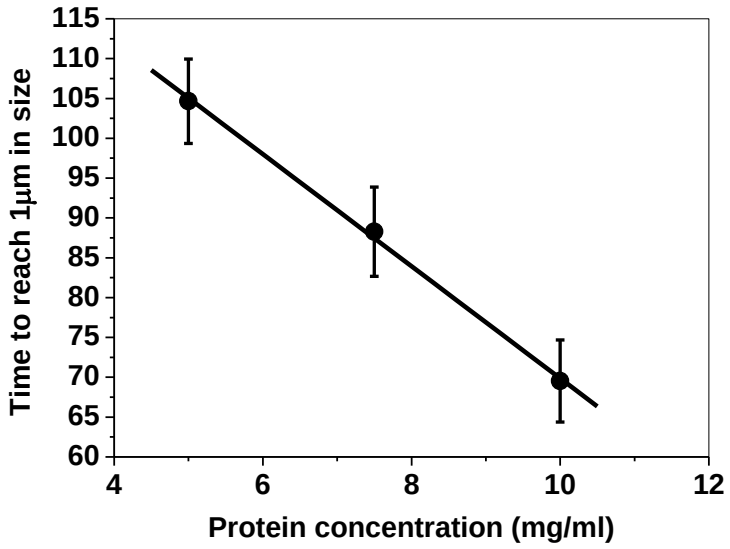

Figure 5.15: *Plot of initial lag phase time vs. protein concentration for the P23VR58H double mutant clusters.*

5.3.3.3 Growth rate for the formation of large spherical aggregates in the transitional phase

Once the formation of clusters is initiated in solution, they may act as centres for growth leading to the formation of large spherical aggregates whose growth follows a sigmoidal pattern. Recently, a Gompertz function was utilized to understand the particle growth in kinetic studies of protein aggregation (Skrdla, 2012). This function has long been described to study population mortality (Kozusko & Bajzer, 2003) and malignant tumour formation (Laird, 1964). We have utilized the equation for Gompertz function given by 5.1 to fit the growth curve for the P23VR58H double mutant aggregate,

$$y = ae^{-e^{(-k(x-x_c))}} \qquad 5.1$$

where a is a point on the sigmoidal curve where it intersects the horizontal asymptote (the maximum aggregate size), x_c , is the mid-point of the exponential phase of aggregate growth and k is the particle growth rate.

Concentration of P23VR58H (1) or P23V (2)	a (μm)	x_C (min)	k (min^{-1})
(1) 10 $mg.ml^{-1}$	43.34 44.37	199.81 206.76	0.01 0.01
(1) 7.5 $mg.ml^{-1}$	49.4 45.04	386.65 376.08	0.005 0.0075
(1) 5 $mg.ml^{-1}$	39.26 40.33	420.93 423.94	0.004 0.004
(2) 10$mg.ml^{-1}$	37.39	214.79	0.0066

Table 5.1: *Values of the parameters a (maximum particle size), x_c(mid-point of particle growth) and k (growth rate) extracted from the Gompertz fit to the growth curve for the various concentrations of the P23VR58H double mutant and for P23V single mutant at 10 $mg.ml^{-1}$.*

A Gompertz fit to the experimental data provides values for a, x_c and k at each protein concentration for both the P23VR58H double mutant protein and the P23V single mutant (table 5.1). The growth rates plotted for the different P23VR58H double mutant concentrations give rise to a slope with a linear relationship established between growth rate and protein concentration (figure 5.16).

This linear relationship between growth rate (k) and P23VR58H protein concentration suggests that the P23VR58H aggregates follow a first-order growth process. The concentration dependence on the growth rate for the P23VR58H double mutant follows a linear regression given by equation 5.2,

$$k = 1.96 + 818.48(c) \qquad 5.2$$

This indicates that the growth rate for the P23VR58H aggregates depend only on the initial concentration of the P23VR58H protein solution prior to the temperature change that leads to aggregation. This is consistent with diffusion driven growth mechanism where concentration is the major driving factor in aggregate growth.

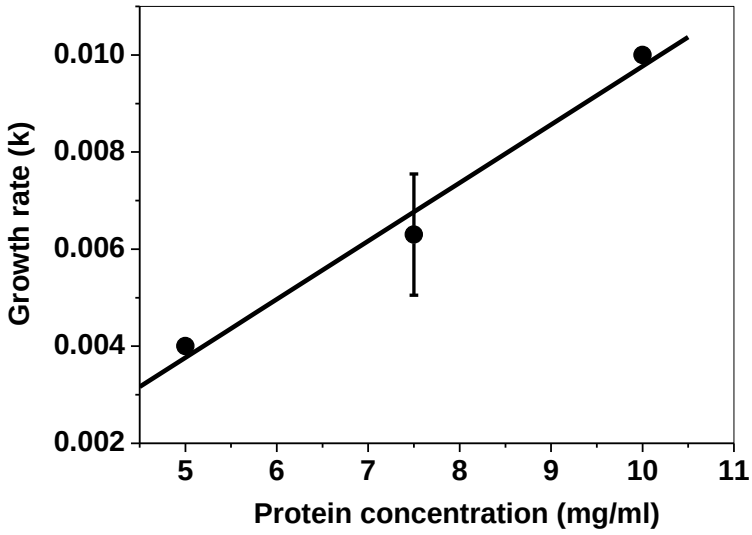

Figure 5.16: *Plot of growth rate vs. protein concentration for the P23VR58H double mutant aggregates.*

5.3.3.4 Termination of aggregate growth in the plateau phase

The exponential growth phase ceases once the aggregates attain a maximum diameter in the range between 40-50 μm for all the concentrations studied in here. This is consistent with diffusion limited aggregation *i.e.* as time elapses there are fewer protein monomers available to add on to the surface of the existing aggregate. Hence, beyond a certain point of time for each concentration the growth of these large spherical aggregates enter a plateau phase. One such aggregation studies on α-synuclein aggregates shows a similar plateau phase (Fink, 2006).

5.3.3.5 Crystallization on the surface of the amorphous aggregate

The P23VR58H aggregate growth is terminated due to the diffusion limited availability of protein monomers to add onto the surface of existing aggregates. However, the concentration of protein monomers in the solution is still sufficiently high enough to allow protein crystal formation on the surface of these pre-formed spherical amorphous aggregates as shown in figure 5.12(h). Comparative studies of the P23VR58H double mutant with the P23V single mutant suggests that crystallization on the surface of P23VR58H aggregate occur independent of the diffusion limited aggregation phenomena and that the processes are not coupled (i.e. crystallization does not stop aggregate growth).

5.4 Conclusion

The third double mutant P23VR58H protein incorporating the P23V substitution associated with inverted solubility and the R58H substitution associated with crystallization are soluble following extraction and purification procedures, they also maintain their native conformation post mutagenesis confirmed by circular dichroism spectroscopy. The P23VR58H double mutant primarily forms aggregates when concentrated at temperatures above the solubility line for the P23V single mutant and crystallization occurs heterogeneously with time on the surface of these pre-formed aggregates which is highly temperature dependent. The solubility line which is the equilibrium between protein monomers and protein solid phase (crystals + aggregates) appears between the solubility lines for the two individual single mutants. However, when the temperature for the formation of initial aggregate material is lowered the solubility line shifts to lower concentrations approaching the normal solubility line for the R58H single mutant. The P23VR58H double mutant gives rise to large three dimensional spherical aggregates whose nucleation, growth and termination of growth is consistent with a diffusion limited aggregation mechanism. The growth of these aggregates follow a sigmoidal pattern and using a Gompertz function to fit the growth curve, the growth rate is found to have a linear dependence on protein concentration suggesting a first-order mechanism for the growth of P23VR58H aggregates.

Summary and final conclusions

This work focuses on the phase diagrams for double mutants of human γD-crystallin (HGD). The three double mutants of HGD used in this study are P23VR36S, P23TR36S and P23VR58H. All the three double mutants are soluble following extraction and purification procedures. They also maintain their native structure upon mutagenesis as confirmed by spectroscopic techniques.

The first double mutant used in our study, P23VR36S, forms a solid condensed phase (crystals) that resemble one of the single parent mutants (R36S). Furthermore, the liquidus line for the P23VR36S double mutant overlaps with the liquidus line for the R36S single mutant. The R36S single mutant forms a new crystal contact and analysis of the microscopy images and liquidus lines suggests that this contact has been preserved in the double mutant. The behaviour of the P23VR36S double mutant by itself would more likely reflect upon the crystal contact formed by the R36S substitution that dominates the behaviour of the double mutant. The kinetics of the R36S crystallization is fast and therefore, it is not possible to explore the higher concentration region of the phase diagram where the solubility line due to the P23V aggregates occur. The most interesting observation for the second double mutant, P23TR36S, is the formation of crystals with two different solubility behaviours: one, a rod-shaped crystal with normal solubility and second, a rhombic-shaped crystal with inverted solubility. For the point at which the individual liquidus lines for each single mutant overlaps, we get co-existence of the two crystal forms. To understand this polymorphism, X-ray crystallographic analysis was performed on both crystal types of the double mutant. Initial crystallographic analysis suggests the formation of a new molecular contact around position 23 of the double mutant responsible for the formation of rhombic-shaped crystals with inverted solubility.

The phase diagram for the third double mutant, P23VR58H, suggests that the solubility lines formed due to equilibrium between protein monomers and protein solid phase (crystals + aggregates) may be kinetically controlled by influencing the temperature for initial aggregate formation. As the temperature for the formation of initial aggregates is lowered, the solubility line shifts to lower concentrations approaching the normal solubility line for the R58H single mutant. The double mutant forms large amorphous aggregates via nucleation and growth which is consistent with diffusion limited aggregation mechanism. The growth curve is sigmoidal and follows a first-

order growth mechanism. Although the double mutant primarily forms large spherical aggregates, the concentration of protein monomers is sufficiently high enough to allow crystallization on the surface of these pre-formed aggregates. The table shown below summarizes the properties exhibited by all the three double mutant proteins used in this study.

Double mutant protein	**Nature of condensed phase**	**Solubility type**
P23VR36S	Protein crystals formed	Normal solubility
P23TR36S	Rhombic-shaped crystals formed	Inverted solubility
	Rod-shaped crystals formed	Normal solubility
P23VR58H	Co-existence of protein aggregates and crystals observed.	Normal solubility

These observations suggest that self-assembly of proteins is controlled by a number of factors ranging from the formation of new crystal contacts, the orientation of protein molecules that are in close proximity with each other along with solution conditions that influence directional anisotropic interactions and dictate the formation of protein condensed phases. This has been qualitatively and semi-quantitatively confirmed in the explored double mutant proteins.

References

Adams, M., Dogic, Z., Keller, S. L. & Fraden, S. (1998). *Nature* **393**, 349-352.

Alivisatos, A. P., Johnsson, K. P., Peng, X., Wilson, T. E., Loweth, C. J., Bruchez Jr, M. P. & Schultz, P. G. (1996). *Nature* **382**, 609-611.

Anderson, V. J. & Lekkerkerker, H. N. (2002). *Nature* **416**, 811-815.

Arakawa, T. & Timasheff, S. N. (1982). *Biochemistry* **21**, 6545-6552.

Arakawa, T. & Timasheff, S. N. (1985). *Biochemistry* **24**, 6756-6762.

Asenjo, J. A. & Andrews, B. A. (2011). *Journal of Chromatography A* **1218**, 8826-8835.

Asherie, N. (2004). *Methods* **34**, 266-272.

Asherie, N. (2012). *Protein and peptide letters* **19**, 708-713.

Asherie, N., Ginsberg, C., Greenbaum, A., Blass, S. & Knafo, S. (2008). *Crystal Growth and Design* **8**, 4200-4207.

Asherie, N., Lomakin, A. & Benedek, G. B. (1996). *Physical review letters* **77**, 4832.

Asherie, N., Pande, J., Lomakin, A., Ogun, O., Hanson, S. R., Smith, J. B. & Benedek, G. B. (1998). *Biophysical chemistry* **75**, 213-227.

Ataka, M. & Asai, M. (1988). *Journal of Crystal Growth* **90**, 86-93.

Avbelj, F. & Fele, L. (1998). *Journal of molecular biology* **279**, 665-684.

Aymard, P., Durand, D., Nicolai, T. & Gimel, J. (1997). *Fractals* **5**, 23-43.

Banatao, D. R., Cascio, D., Crowley, C. S., Fleissner, M. R., Tienson, H. L. & Yeates, T. O. (2006). *Proceedings of the National Academy of Sciences* **103**, 16230-16235.

Basak, A., Bateman, O., Slingsby, C., Pande, A., Asherie, N., Ogun, O., Benedek, G. B. & Pande, J. (2003). *Journal of Molecular Biology* **328**, 1137-1147.

Bassnett, S., Shi, Y. & Vrensen, G. F. (2011). *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 1250-1264.

Baud, F. & Karlin, S. (1999). *Proceedings of the National Academy of Sciences* **96**, 12494-12499.

Benedek, G. (1971). *Applied Optics* **10**, 459-473.

Berg, J. M. & Goffeney, N. W. (1997). *Methods in enzymology* **276**, 619-627.

Berland, C. R., Thurston, G. M., Kondo, M., Broide, M. L., Pande, J., Ogun, O. & Benedek, G. B. (1992). *Proceedings of the National Academy of Sciences* **89**, 1214-1218.

Bernardo, A., Calmanovici, C. & Miranda, E. (2005). *Brazilian Journal of Chemical Engineering* **22**, 331-339.

Bianchi, E., Blaak, R. & Likos, C. N. (2011). *Physical Chemistry Chemical Physics* **13**, 6397-6410.

Bloemendal, H., de Jong, W., Jaenicke, R., Lubsen, N. H., Slingsby, C. & Tardieu, A. (2004). *Progress in biophysics and molecular biology* **86**, 407-485.

Bolen, D. (2004). *Methods* **34**, 312-322.

Bookchin, R. & Nagel, R. (1981). *Contemporary hematology/oncology* **2**, 31-77.

Bowden, N. B., Weck, M., Choi, I. S. & Whitesides, G. M. (2001). *Accounts of chemical research* **34**, 231-238.

Breed, D. R. (2007). *Engineered colloids: Patchy particles with reversible, directional interactions*. ProQuest.

Broide, M. L., Berland, C. R., Pande, J., Ogun, O. O. & Benedek, G. B. (1991). *Proceedings of the National Academy of Sciences* **88**, 5660-5664.

Broide, M. L., Tominc, T. M. & Saxowsky, M. D. (1996). *Physical Review E* **53**, 6325.

Burdon, K. P., Wirth, M. G., Mackey, D. A., Russell-Eggitt, I. M., Craig, J. E., Elder, J. E., Dickinson, J. L. & Sale, M. M. (2004). *British Journal of Ophthalmology* **88**, 79-83.

Byrne, B. & Iwata, S. (2002). *Current opinion in structural biology* **12**, 239-243.

Carter Jr, C. W. & Carter, C. W. (1979). *The Journal of Biological Chemistry* **254**, 12219-12223.

Carver, J. A., Lindner, R. A., Lyon, C., Canet, D., Hernandez, H., Dobson, C. M. & Redfield, C. (2002). *Journal of molecular biology* **318**, 815-827.

Casjens, S. & King, J. (1975). *Annual review of biochemistry* **44**, 555-611.

Chaikin, P. M. & Lubensky, T. C. (2000). *Principles of condensed matter physics*. Cambridge Univ Press.

Chayen, N., Boggon, T., Cassetta, A., Deacon, A., Gleichmann, T., Habash, J., Harrop, S., Helliwell, J., Nieh, Y. & Peterson, M. (1996). *Quarterly reviews of biophysics* **29**, 227-278.

Chayen, N. E. (1998). *Acta Crystallographica Section D: Biological Crystallography* **54**, 8-15.

Chayen, N. E. (2005). *Progress in biophysics and molecular biology* **88**, 329-337.

Chayen, N. E. & Saridakis, E. (2008). *Nature methods* **5**, 147-153.

Chayen, N. E., Shaw Stewart, P., Maeder, D. & Blow, D. (1990). *Journal of applied crystallography* **23**, 297-302.

Cheng, Z., Chaikin, P., Russel, W., Meyer, W., Zhu, J., Rogers, R. & Ottewill, R. (2001). *Materials & Design* **22**, 529-534.

Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. (2003). *Pharmaceutical research* **20**, 1325-1336.

Cieślik, M. & Derewenda, Z. S. (2009). *Acta Crystallographica Section D: Biological Crystallography* **65**, 500-509.

Clark, J. (1994). *Principle and practice of ophthalmology:* Saunders College Publishing, Philadelphia, PA.

Coen, C. J., Newman, J., Blanch, H. W. & Prausnitz, J. M. (1996). *Journal of colloid and interface science* **177**, 276-279.

Cohen, S. I., Linse, S., Luheshi, L. M., Hellstrand, E., White, D. A., Rajah, L., Otzen, D. E., Vendruscolo, M., Dobson, C. M. & Knowles, T. P. (2013). *Proceedings of the National Academy of Sciences* **110**, 9758-9763.

Cohn, E., Hughes Jr, W. & Weare, J. (1947). *Journal of the American Chemical Society* **69**, 1753-1761.

Collaborative, C. P. (1994). *Acta crystallographica. Section D, Biological crystallography* **50**, 760.

Cutler, P. (2008). *Molecular Biomethods Handbook*, edited by J. Walker & R. Rapley, pp. 719-729: Humana Press.

Daanoun, A., Tejero, C. & Baus, M. (1994). *Physical Review E* **50**, 2913.

Darby, N. J. & Creighton, T. E. (1993). *Protein structure*. IRL Press at Oxford University Press Oxford, UK.

Delaye, M. & Tardieu, A. (1983). *Nature* **302**, 415-417.

DeLucas, L. & Bugg, C. (1987). *Trends in Biotechnology* **5**, 188-193.

Derewenda, Z. S. (2004). *Structure* **12**, 529-535.

Derewenda, Z. S. (2010). *Acta Crystallographica Section D: Biological Crystallography* **66**, 604-615.

Derewenda, Z. S. (2011). *Acta Crystallographica Section D: Biological Crystallography* **67**, 243-248.

Derewenda, Z. S. & Vekilov, P. G. (2006). *Acta Crystallographica Section D: Biological Crystallography* **62**, 116-124.

Dorsaz, N., Filion, L., Smallenburg, F. & Frenkel, D. (2012). *Faraday discussions* **159**, 9-21.

Doye, J. P., Louis, A. A. & Vendruscolo, M. (2004). *Physical biology* **1**, P9.

Drenth, J. (2007). *X-Ray Crystallography*. Wiley Online Library.

DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M. & Vendruscolo, M. (2004). *Journal of molecular biology* **341**, 1317-1326.

Ducruix, A. & Giegé, R. (1992). *Crystallisation of nucleic acids and proteins.*

Dumetz, A. C., Chockla, A. M., Kaler, E. W. & Lenhoff, A. M. (2008). *Biophysical journal* **94**, 570-583.

Dumetz, A. C., Lewus, R. A., Lenhoff, A. M. & Kaler, E. W. (2008). *Langmuir* **24**, 10345-10351.

Dumetz, A. C., Snellinger-O'Brien, A. M., Kaler, E. W. & Lenhoff, A. M. (2007). *Protein Science* **16**, 1867-1877.

Eaton, W. A. & Hofrichter, J. (1990). *Advances in protein chemistry* **40**, 263-279.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132.

Englard, S. & Seifter, S. (1990). *Methods in enzymology* **182**, 285.

Erdemir, D., Lee, A. Y. & Myerson, A. S. (2009). *Accounts of chemical research* **42**, 621-629.

Everett, D. H. (1988). *Basic principles of colloid science*. Royal Society of Chemistry.

Feeling-Taylor, A. R., Banish, R. M., Hirsch, R. E. & Vekilov, P. G. (1999). *Review of scientific instruments* **70**, 2845-2849.

Feeling-Taylor, A. R., Yau, S.-T., Petsev, D. N., Nagel, R. L., Hirsch, R. E. & Vekilov, P. G. (2004). *Biophysical journal* **87**, 2621-2629.

Fink, A. L. (2006). *Accounts of chemical research* **39**, 628-634.

Finkelstein, A. V. & Janin, J. l. (1989). *Protein Engineering* **3**, 1-3.

Finkelstein, A. V. & Ptitsyn, O. (2002). *Protein physics: a course of lectures*. Academic Press.

Flyvbjerg, H., Jobs, E. & Leibler, S. (1996). *Proceedings of the National Academy of Sciences* **93**, 5975-5979.

Frericks Schmidt, H. L., Sperling, L. J., Gao, Y. G., Wylie, B. J., Boettcher, J. M., Wilson, S. R. & Rienstra, C. M. (2007). *The Journal of Physical Chemistry B* **111**, 14362-14369.

Frieden, C. & Goddette, D. W. (1983). *Biochemistry* **22**, 5836-5843.

Fu, D., Li, Y. & Wu, J. (2003). *Physical Review E* **68**, 011403.

Fusco, D., Headd, J. J., De Simone, A., Wang, J. & Charbonneau, P. (2014). *Soft matter* **10**, 290-302.

Galkin, O. & Vekilov, P. G. (2000). *Proceedings of the National Academy of Sciences* **97**, 6277-6281.

Garcia-Ruiz, J. M., Gonzalez-Ramirez, L. A., Gavira, J. A. & Otalora, F. (2002). *Acta Crystallographica Section D* **58**, 1638-1642.

Gast, A. P. & Russel, W. B. (1998). *Physics Today* **51**, 24-31.

Geerts, N. & Eiser, E. (2010). *Soft Matter* **6**, 4647-4660.

George, A., Chiang, Y., Guo, B., Arabshahi, A., Cai, Z. & Wilson, W. W. (1997). *Methods in enzymology*, 100-110.

Gilliland, G. L. (1988). *Journal of Crystal Growth* **90**, 51-59.

Glotzer, S. C. (2004). *Science* **306**, 419-420.

Glotzer, S. C. & Solomon, M. J. (2007). *Nature materials* **6**, 557-562.

Greenfield, N. J. (2006). *Nature protocols* **1**, 2876-2890.

Grigsby, J., Blanch, H. & Prausnitz, J. (2001). *Biophysical chemistry* **91**, 231-243.

Gsponer, J. & Vendruscolo, M. (2006). *Protein and peptide letters* **13**, 287-293.

Hagen, M. & Frenkel, D. (1994). *The Journal of chemical physics* **101**, 4093-4097.

Hamada, H., Arakawa, T. & Shiraki, K. (2009). *Current pharmaceutical biotechnology* **10**, 400-407.

Hammes, G. G. (2005). *Spectroscopy for the Biological Sciences*.

Hammes, G. G. (2007). *Physical Chemistry for the Biological Sciences*. Wiley.

Hansen, C. & Quake, S. R. (2003). *Current opinion in structural biology* **13**, 538-544.

Hanson, S., Smith, D. & Smith, J. (1998). *Experimental eye research* **67**, 301.

Hejtmancik, J., Kaiser-Kupfer, M. & Piatigorsky, J. (2001). *The metabolic and molecular basis of inherited disease* **8**, 6033-6062.

Hejtmancik, J. F. (2008). *Seminars in cell & developmental biology*, pp. 134-149. Elsevier.

Hirsch, R., Raventos-Suarez, C., Olson, J. & Nagel, R. (1985). *Blood* **66**, 775-777.

Hunt, J. & Ingram, V. (1958). *Nature* **181**, 1062-1063.

Ilett, S. M., Orrock, A., Poon, W. & Pusey, P. (1995). *Physical Review E* **51**, 1344.

Ingham, K. C. (1990). *Methods in enzymology* **182**, 301-306.

Ingram, V. (1956). *Nature* **178**, 792-794.

Irvine, G. B., El-Agnaf, O. M., Shankar, G. M. & Walsh, D. M. (2008). *Molecular Medicine* **14**, 451.

James, S. & McManus, J. J. (2012). *The Journal of Physical Chemistry B* **116**, 10182-10188.

James, S., Quinn, M. K. & McManus, J. J. (2015). *Physical Chemistry Chemical Physics*. **17**, 5413-5420.

Ji, F. L., Koharudin, L. M. I., Jung, J. & Gronenborn, A. M. (2013). *Proteins-Structure Function and Bioinformatics* **81**, 1493-1498.

Jones, M. N. & Chapman, D. (1995). *Micelles, monolayers and biomembranes.*

Jones, M. R., Macfarlane, R. J., Lee, B., Zhang, J., Young, K. L., Senesi, A. J. & Mirkin, C. A. (2010). *Nature materials* **9**, 913-917.

Jones, M. R., Macfarlane, R. J., Prigodich, A. E., Patel, P. C. & Mirkin, C. A. (2011). *Journal of the American Chemical Society* **133**, 18865-18869.

Jones, R. A. (2002). *Soft condensed matter:* IOP Publishing.

Kamihira, M., Naito, A., Tuzi, S., Nosaka, A. Y. & Saito, H. (2000). *Protein Science* **9**, 867-877.

Kazakevich, Y. V. & Lobrutto, R. (2007). *HPLC for pharmaceutical scientists*. John Wiley & Sons.

Kelly, J. W. (1997). *Structure* **5**, 595-600.

Kelly, S. M., Jess, T. J. & Price, N. C. (2005). *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1751**, 119-139.

Keren, D. (2003). *Protein electrophoresis in clinical diagnosis*. CRC Press.

Kmoch, S., Brynda, J., Asfaw, B., Bezouska, K., Novak, P., Rezacova, P., Ondrova, L., Filipec, M., Sedlacek, J. & Elleder, M. (2000). *Human Molecular Genetics* **9**, 1779-1786.

Knowles, T. P., Shu, W., Devlin, G. L., Meehan, S., Auer, S., Dobson, C. M. & Welland, M. E. (2007). *Proceedings of the National Academy of Sciences* **104**, 10016-10021.

Kono, M., Sen, A. C. & Chakrabarti, B. (1990). *Biochemistry* **29**, 464-470.

Kozusko, F. & Bajzer, Ž. (2003). *Mathematical biosciences* **185**, 153-167.

Kraft, D. J., Ni, R., Smallenburg, F., Hermes, M., Yoon, K., Weitz, D. A., van Blaaderen, A., Groenewold, J., Dijkstra, M. & Kegel, W. K. (2012). *Proceedings of the National Academy of Sciences* **109**, 10787-10792.

Kuehner, D. E., Heyer, C., Rämsch, C., Fornefeld, U. M., Blanch, H. W. & Prausnitz, J. M. (1997). *Biophysical journal* **73**, 3211.

Kulkarni, A. M., Dixit, N. M. & Zukoski, C. F. (2003). *Faraday discussions* **123**, 37-50.

Kuszak, J., Zoltoski, R. & Sivertson, C. (2004). *Experimental eye research* **78**, 673-687.

Lafont, S., Veesler, S., Astier, J. P. & Boistelle, R. (1994). *Journal of crystal growth* **143**, 249-255.

Lafont, S., Veesler, S., Astier, J. P. & Boistelle, R. (1997). *Journal of crystal growth* **173**, 132-140.

Laird, A. K. (1964). *British journal of cancer* **18**, 490.

Lakowicz, J. R. (2007). *Principles of fluorescence spectroscopy*. Springer Science & Business Media.

Leckband, D. & Israelachvili, J. (2001). *Quarterly reviews of biophysics* **34**, 105-267.

Lehn, J.-M. (2004). *Reports on progress in physics* **67**, 249.

Lekkerkerker, H., Poon, W.-K., Pusey, P., Stroobants, A. & Warren, P. (1992). *EPL (Europhysics Letters)* **20**, 559.

Leslie, A. G. & Powell, H. R. (2007). *Evolving methods for macromolecular crystallography*, pp. 41-51: Springer.

Lessin, L. S., Jensen, W. N. & Ponder, E. (1969). *The Journal of experimental medicine* **130**, 443-466.

Li, L. & Ismagilov, R. F. (2010). *The Annual Review of Biophysics* **39**, 139-158.

Lin, M. J., Nagel, R. L. & Hirsch, R. E. (1989). *Blood* **74**, 1823-1825.

Lomakin, A., Asherie, N. & Benedek, G. B. (1996). *The Journal of chemical physics* **104**, 1646-1656.

Lomakin, A., Asherie, N. & Benedek, G. B. (1999). *Proceedings of the National Academy of Sciences* **96**, 9465-9468.

Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A. & Teplow, D. B. (1996). *Proceedings of the National Academy of Sciences* **93**, 1125-1129.

Long, D.-L., Burkholder, E. & Cronin, L. (2007). *Chemical Society Reviews* **36**, 105-121.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Crystallographica Section D: Biological Crystallography* **57**, 679-688.

Mackay, D. W., Andley, U. P. & Shiels, A. (2004). *Molecular Vision* **10**, 155-162.

Mandal, K., Bose, S. K., Chakrabarti, B. & Siezen, R. J. (1985). *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **832**, 156-164.

Manoharan, V. N. (2004). *Building materials with colloidal spheres*.

Mao, Y., Cates, M. E. & Lekkerkerker, H. N. W. (1995). *Physica A: Statistical Mechanics and its Applications* **222**, 10-24.

Matthews, B. W. (1968). *Journal of molecular biology* **33**, 491-497.

Matthews, B. W. (2009). *Protein Science* **18**, 1135-1138.

McCoy, A. J. (2007). *Acta Crystallographica Section D: Biological Crystallography* **63**, 32-41.

McManus, J. J., Lomakin, A., Ogun, O., Pande, A., Basan, M., Pande, J. & Benedek, G. B. (2007). *Proceedings of the National Academy of Sciences* **104**, 16856-16861.

McPherson, A. (1976). *Journal of Biological Chemistry* **251**, 6300-6303.

McPherson, A. (1999). *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press.

McPherson, A. & Cudney, B. (2006). *Journal of structural biology* **156**, 387-406.

McQuarrie, D. (1976). *Statistical Mechanics. New York.*

Michalska, K., Borek, D., Hernández-Santoyo, A. & Jaskolski, M. (2008). *Acta Crystallographica Section D: Biological Crystallography* **64**, 309-320.

Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. (1996). *Nature* **382**, 607-609.

Morris, A. M., Watzky, M. A. & Finke, R. G. (2009). *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1794**, 375-397.

Muschol, M. & Rosenberger, F. (1995). *The Journal of chemical physics* **103**, 10424-10432.

Muschol, M. & Rosenberger, F. (1997). *The Journal of chemical physics* **107**, 1953-1962.

Nandrot, E., Slingsby, C., Basak, A., Cherif-Chefchaouni, M., Benazzouz, B., Hajaji, Y., Boutayeb, S., Gribouval, O., Arbogast, L., Berraho, A., Abitbol, M. & Hilal, L. (2003). *Journal of Medical Genetics* **40**, 262-267.

Neal, B., Asthagiri, D. & Lenhoff, A. (1998). *Biophysical journal* **75**, 2469-2477.

Noro, M., Kern, N. & Frenkel, D. (1999). *EPL (Europhysics Letters)* **48**, 332.

Oesterhelt, D. & Stoeckenius, W. (1973). *Proceedings of the National Academy of Sciences* **70**, 2853-2857.

Paliwal, A., Asthagiri, D., Abras, D., Lenhoff, A. & Paulaitis, M. (2005). *Biophysical journal* **89**, 1564-1573.

Pan, W., Filobelo, L., Pham, N. D., Galkin, O., Uzunova, V. V. & Vekilov, P. G. (2009). *Physical review letters* **102**, 058101.

Pande, A., Annunziata, O., Asherie, N., Ogun, O., Benedek, G. B. & Pande, J. (2005). *Biochemistry* **44**, 2491-2500.

Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J. & Benedek, G. B. (2001). *Proceedings of the National Academy of Sciences* **98**, 6116-6120.

Pande, A., Zhang, J., Banerjee, P. R., Puttamadappa, S. S., Shekhtman, A. & Pande, J. (2009). *Biochemical and biophysical research communications* **382**, 196-199.

Pande, J., Pande, A., Asherie, N., Lomakin, A., Ogun, O., King, J. A., Lubsen, N. H., Walton, D. & Benedek, G. B. (2000). *Investigative Ophthalmology & Visual Science* **41**, S748-S748.

Petsko, G. A. & Ringe, D. (2004). *Protein Structure and Function*. New Science Press.

Philo, J. S. & Arakawa, T. (2009). *Current pharmaceutical biotechnology* **10**, 348-351.

Poon, W. (1997). *Physical Review E* **55**, 3762.

Prausnitz, J. & Foose, L. (2007). *Pure and applied chemistry* **79**, 1435-1444.

Primer, A. (1996). *Fundamentals of UV-Visible Spectroscopy.* Copyright Hewlett-Packard Company, Hewlett-Packard publication.

Pusey, P. & Van Megen, W. (1986). *Nature* **320**, 340-342.

Rascón, C., Navascués, G. & Mederos, L. (1995). *Physical Review B* **51**, 14899.

Romano, F., Sanz, E. & Sciortino, F. (2010). *The Journal of Chemical Physics* **132**, 184501.

Rosenbaum, D., Zamora, P. & Zukoski, C. (1996). *Physical review letters* **76**, 150.

Rosenbaum, D. & Zukoski, C. (1996). *Journal of crystal growth* **169**, 752-758.

Russell, P. J. (2011). *iGenetics: A Molecular Approach*. Pearson Education.

Sacanna, S. & Pine, D. J. (2011). *Current Opinion in Colloid & Interface Science* **16**, 96-105.

Santhiya, S. T., Manohar, M. S., Rawlley, D., Vijayalakshmi, P., Namperumalsamy, P., Gopinath, P. M., Loster, J. & Graw, J. (2002). *Journal of Medical Genetics* **39**, 352-358.

Saridakis, E. & Chayen, N. E. (2009). *Trends in biotechnology* **27**, 99-106.

Schall, C. A., Arnold, E. & Wiencek, J. M. (1996). *Journal of crystal growth* **165**, 293-298.

Schmid, F.-X. (2001). *Biological Macromolecules: UV-visible Spectrophotometry eLS*: John Wiley & Sons, Ltd.

Schoenmakers, J. G., Dunnen, J. T. D., Moormann, R. J., Jongbloed, R., Van Leen, R. W. & Lubsen, N. H. (1984). *Ciba Foundation Symposium 106-Human Cataract Formation*, pp. 208-217. Wiley Online Library.

Schüth, F., Bussian, P., Ågren, P., Schunk, S. & Lindén, M. (2001). *Solid state sciences* **3**, 801-808.

Sciortino, F. (2002). *Nature materials* **1**, 145-146.

Sear, R. P. (1999). *The Journal of chemical physics* **111**, 4800-4806.

Selkoe, D. J. (1994). *Journal of Neuropathology & Experimental Neurology* **53**, 438-447.

Shah, M., Galkin, O. & Vekilov, P. G. (2004). *The Journal of chemical physics* **121**, 7505-7512.

Shentu, X. C., Yao, K., Xu, W., Zheng, S., Hu, S. N. & Gong, X. H. (2004). *Molecular Vision* **10**, 233-239.

Shi, L., Lu, X., Fenollosa, R., Rodriguez, I., Harris, J. T., Ramiro-Manzano, F., Korgel, B. A. & Meseguer, F. (2013). *Protocol exchange.*

Skrdla, P. J. (2012). *Langmuir* **28**, 4842-4857.

Slingsby, C., Wistow, G. J. & Clark, A. R. (2013). *Protein Science* **22**, 367-380.

Song, X. & Zhao, X. (2004). *The Journal of chemical physics* **120**, 2005-2009.

Stauber, M., Jakoncic, J., Berger, J., Karp, J. M., Axelbaum, A., Sastow, D., Buldyrev, S. V., Hrnjez, B. J. & Asherie, N. (2015). *Biological Crystallography* **71**, 427-441.

Stranks, S. D., Ecroyd, H., Van Sluyter, S., Waters, E. J., Carver, J. A. & Von Smekal, L. (2009). *Physical Review E* **80**, 051907.

Summers, L., Slingsby, C., White, H., Narebor, M., Moss, D., Miller, L., Mahadevan, D., Lindley, P., Driessen, H. & Blundell, T. (1984). *Ciba Foundation Symposium 106-Human Cataract Formation*, pp. 219-236. Wiley Online Library.

Talanquer, V. & Oxtoby, D. W. (1998). *The Journal of chemical physics* **109**, 223-227.

Tan, L. H., Xing, H., Chen, H. & Lu, Y. (2013). *Journal of the American Chemical Society* **135**, 17675-17678.

Tardieu, A., Vérétout, F., Krop, B. & Slingsby, C. (1992). *European biophysics journal* **21**, 1-12.

ten Wolde, P. R. & Frenkel, D. (1997). *Science* **277**, 1975-1978.

Thomson, J. A., Schurtenberger, P., Thurston, G. M. & Benedek, G. B. (1987). *Proceedings of the National Academy of Sciences* **84**, 7079-7083.

Tidor, B. & Karplus, M. (1994). *Journal of molecular biology* **238**, 405-414.

Valeur, B. & Berberan-Santos, M. N. (2012). *Molecular fluorescence: principles and applications*. John Wiley & Sons.

Van Megen, W. & Underwood, S. (1993). *Nature* **362**, 616-618.

Vaney, M., Broutin, I., Retailleau, P., Douangamath, A., Lafont, S., Hamiaux, C., Prange, T., Ducruix, A. & Ries-Kautt, M. (2001). *Acta Crystallographica Section D: Biological Crystallography* **57**, 929-940.

Vanita, V. & Singh, D. (2012). *Molecular and Cellular Biochemistry* **368**, 167-172.

Veesler, S., Fertè, N., Costes, M.-S., Czjzek, M. & Astier, J.-P. (2004). *Crystal growth & design* **4**, 1137-1141.

Vekilov, P. G. (2005). *Journal of crystal growth* **275**, 65-76.

Vekilov, P. G. (2007). *British journal of haematology* **139**, 173-184.

Vekilov, P. G. (2010a). *Nanoscale* **2**, 2346-2357.

Vekilov, P. G. (2010b). *Soft Matter* **6**, 5254-5272.

Vekilov, P. G. (2012). *Journal of Physics: Condensed Matter* **24**, 193101.

Vekilov, P. G. & Chernov, A. A. (2002). *Solid State Physics* **57**, 1-147.

Vekilov, P. G., Feeling-Taylor, A., Yau, S.-T. & Petsev, D. (2002). *Acta Crystallographica Section D: Biological Crystallography* **58**, 1611-1616.

Vekilov, P. G., Feeling-Taylor, A. R., Petsev, D. N., Galkin, O., Nagel, R. L. & Hirsch, R. E. (2002). *Biophysical journal* **83**, 1147-1156.

Velev, O., Kaler, E. & Lenhoff, A. (1998). *Biophysical journal* **75**, 2682-2697.

Verduin, H. & Dhont, J. K. (1995). *Journal of colloid and interface science* **172**, 425-437.

Voet, D. & Voet, J. G. (2011). *Biochemistry*.

Wang, S., Wu, J. W., Yamamoto, S. & Liu, H.-S. (2008). *Biotechnology journal* **3**, 165.

Wang, W. (2005). *International journal of pharmaceutics* **289**, 1-30.

Wang, Y., Hollingsworth, A. D., Yang, S. K., Patel, S., Pine, D. J. & Weck, M. (2013). *Journal of the American Chemical Society* **135**, 14064-14067.

Wang, Y., Wang, Y., Breed, D. R., Manoharan, V. N., Feng, L., Hollingsworth, A. D., Weck, M. & Pine, D. J. (2012). *Nature* **491**, 51-55.

Weiss, W. F., Young, T. M. & Roberts, C. J. (2009). *Journal of pharmaceutical sciences* **98**, 1246-1277.

Westermeier, R. (2005). *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations*.

Wiencek, J. (1999). *Annual review of biomedical engineering* **1**, 505-534.

Williams, A. & Frasca, V. (2001). *Current Protocols in Protein Science*, 8.2. 1-8.2. 30.

Wistow, G. (2012). *Human Genomics* **6**, 26.

Witten, T. A. & Sander, L. M. (1983). *Physical Review B* **27**, 5686.

Xia, J., Aerts, T., Donceel, K. & Clauwaert, J. (1994). *Biophysical journal* **66**, 861-872.

Yau, S.-T. & Vekilov, P. G. (2001). *Journal of the American Chemical Society* **123**, 1080-1089.

Yeates, T. O. & Kent, S. B. (2012). *Annual review of biophysics* **41**, 41-61.

Young, K. L., Personick, M. L., Engel, M., Damasceno, P. F., Barnaby, S. N., Bleher, R., Li, T., Glotzer, S. C., Lee, B. & Mirkin, C. A. (2013). *Angewandte Chemie International Edition* **52**, 13980-13984.

Zemansky, M. W. & Dittman, R. H. (1997). *Heat and thermodynamics*. McGraw-Hill.

Zhang, Z. & Glotzer, S. C. (2004). *Nano Letters* **4**, 1407-1413.

Zheng, B., Roach, L. S. & Ismagilov, R. F. (2003). *Journal of the American chemical society* **125**, 11170-11171.

Zigler, J. (1994). *The Principles and Practice of Ophthalmology: Basic Sciences.*

Zimm, B. H. (1946). *The Journal of Chemical Physics* **14**, 164-179.