Alterations in the phase behaviour of human γD-crystallin due to mutagenesis and chemical modification of the protein surface

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

Doctor of Philosophy

by

Michelle Kathleen Quinn

Department of Chemistry

Maynooth University

January 2018

Research Supervisor: Dr. Jennifer McManus

Head of Department: Dr. Jennifer McManus

Dedicated to my parents, Mary and Martin, whose continuous encouragement, support (both financial and emotional), pleading, praying, bribing and begging over the years have finally paid dividends.

Declaration

I hereby certify that this thesis has not been submitted before, in whole or in part, to this or any university for any degree and is, except where otherwise stated, the original work of the author.

Signed: _____

Date: _____

Maynooth University

Acknowledgements

To my supervisor and Head of Department, Dr. Jennifer McManus: I am so grateful to you for your supervision and support during the last five years. I am more grateful for all the knowledge I have gained from you in that time. You were right, I did learn loads! Thank you for believing in me.

To the rest of the academic, technical and administrative staff who have provided support along the way: it is said that it takes a village to raise a child. I say it takes a department to write a thesis! Thank you all for everything. A special thanks to Noel for the technological leg-ups and corridor chats (I know we say it every year but this is Mayo's year), to Ria for keeping us safe (west coast is best coast) and to Donna and Carol who have enabled my shopping addiction for so long!

To current and former members of the McManus group: it has truly been a blast, thank you for that. A special thanks to Ruth (Lab Bestie) and Alice (Maynooth Mammy) for all the laughs, the tears (sometimes from laughing!) and everything in between.

To the rest of postgrads and postdocs, both current and of old: I am here too long at this stage to name names so I will just say that I am so glad to have met you all. Thank you for all the mornings, afternoons and evenings in the coffee room and, of course, the nights out in the village.

To my circle of friends outside of the department: thank you for the laughs, the nights out, the evenings in, the breakfasts, lunches and dinners, the adventures and any other welcome distractions that you provided.

To my family: thank you for your support and relentless teasing of my perennial student status. Bet you thought this day would never come!

To Jack: thank you for being with me in one capacity or another for most of this journey. I am so grateful that Maynooth brought us together.

Papers arising from this thesis

Quinn, M. K., N. Gnan, et al. (2015). "How fluorescent labelling alters the solution behaviour of proteins." Physical Chemistry Chemical Physics 17(46): 31177-31187. (DOI: 10.1039/C5CP04463D)

James, S., **M. K. Quinn**, et al. (2015). "The self assembly of proteins; probing patchy protein interactions." Physical Chemistry Chemical Physics 17(7): 5413-5420. (DOI: 10.1039/C4CP05892E)

Abstract

Proteins are inherently anisotropic macromolecules due to the variation in the surface-exposed amino acid side-chains on the protein surface. Protein phase diagrams describe the solution conditions where phase transitions, such as liquid-solid and liquid-liquid, occur. The position of these phase boundaries is dependent on both the intrinsic characteristics of the protein and on its environment. As a result, changes to the protein or its environment can significantly alter its phase diagram.

Human γD-crystallin (HGD) is a protein found in the eye lens and is remarkably stable at high concentrations. HGD exhibits short-ranged attractive interactions, i.e. shorter than a quarter of the protein diameter. As a result, HGD undergoes liquid-liquid phase separation (LLPS). The temperature (T_{ph}) at which LLPS occurs can be used as a measure of the strength of attractive interactions between proteins in solution.

The effects of chemically modifying specific amino acids (Lys-2 and Cys-110) on the surface of HGD with a (hydrophobic) small molecule fluorescent label were examined using both experiments and simulations. By measuring the LLPS temperature for modified proteins in mixtures with native protein, it was possible to determine how surface anisotropy and the chemical properties of the modifier changed the protein's phase behaviour. Very low modified protein compositions (as low as $x_m = 0.0001$) were sufficient to increase T_{ph} significantly (~14 K), but both the position and type of fluorescent dye used influenced $T_{ph.}$ A numerical model was designed to explain the experimental observations and revealed that the increase in LLPS in the presence of modified protein was due to a new increase in attraction in the system.

The effects of chemical modification were further examined by modifying HGD with PEGylated biotin at both the Cys and Lys positions on the protein surface. Even at high modified protein compositions ($x_m \sim 0.9$), there was little change in T_{ph} relative to unmodified HGD when the modification was performed at the Lys position. However, T_{ph} increased significantly for thiol modified protein at a similar modified protein composition (~10 K). Neither modification had any impact on the structure of

the protein relative to unmodified HGD. This study further highlighted how both the specific chemical modification and its position on the protein surface can change the impact to a phase boundary.

Finally, the nucleation and growth of protein aggregates in solution and in cells was probed using double mutants of HGD. For solution based measurements of the P23VR58H mutant of HGD, it appears that aggregate growth produces monodisperse particle sizes, which appear to grow *via* monomer addition. When measurements are made in smaller volumes, surface effects lead to heterogeneous nucleation and polydisperse aggregates emerge. However, heterogeneous nucleation can be suppressed by surface treatments. Aggregate growth in cells was demonstrated for the P23TR36S mutant of HGD. Heterogeneous aggregates of GFP labelled protein were observed in cells for up to 4 days and these particles sizes were predominantly 1-2 μm in size. While some larger particle sizes were observed, these were very few in number. Significant cell death was associated with later particle growth stages, which warrants future investigation.

List of Abbreviations

A – Absorbance
Aβ – Amyloid β
AD – Alzheimer's disease
ADC – Antibody drug conjugate
Ala – Alanine
APP – Amyloid precursor protein
APS – Ammonium persulphate
Arg – Arginine
Asp – Aspartic acid
B_{22}/B_2 – Second virial coefficient
BCS – Bovine calf serum
BSA – Bovine serum albumin
c – Concentration
CD – Circular dichroism
cDNA – Complementary deoxyribonucleic acid
CGM – Complete growth medium
Cys – Cysteine
df – dilution factor
DLVO – Derjaguin-Landau-Verwey-Overbeek
DMEM – Dulbecco's modified Eagle's medium
DMF – Dimethyl formaldehyde
DMSO – Dimethyl sulphoxide
DNA – Deoxyribonucleic acid
DNase – Deoxyribonuclease
dNTP – Deoxynucleotide triphosphate
DOS – Degree of substitution
dsDNA – Double-stranded deoxyribonucleic acid
DSC – Differential scanning calorimetry
DTT – Dithiothreitol
E. coli – Escherichia coli
EDL – Electronic double layer
EDTA – Ethyldiaminetetraacetic acid
emGFP – Emerald green fluorescent protein
ES – Electrostatic
ESI-MS – Electrospray ionisation mass spectrometry
FBS – Foetal bovine serum
FITC – Fluoroscein isothiocyanate
g – Acceleration due to gravity
Glu – Glutamate
h – Planck's constant
HbC – C haemoglobin
HbS – Sickle haemoglobin
HCl – Hydrochloric acid
HEK – Human embryonic kidney
HGD – Wild-type human gamma D crystallin
His – Histidine
HPLC – High performance liquid chromatography
Hz – Hertz
IPTG – Isopropyl-β-D-thiogalactopyranoside
K_d – Dissociation constant
***kT* –** Thermal energy
LB – Lysogeny broth
Lys – Lysine
M – Molar
MWCO – Molecular weight cut off
NaCl – Sodium Chloride
NaOH – Sodium Hydroxide
NHS – *N*-hydroxysuccinimide
OD – Optical density
P23T – Pro-23 to Thr mutant
P23V – Pro-23 to Val mutant
PAGE – Polyacrylamide gel electrophoresis
PBS – Phosphate buffered saline
PCR – Polymerase chain reaction
PDL – Poly-D-lysine
PEG – Polyethylene glycol
Phe – Phenylalanine
pI – Isoelectric point
PLL – Poly-L-lysine
POI – Protein of interest
Pro – Proline
PrP – Prion protein
PTM – Post translational modification
R36S – Arg-36 to Ser mutant
R58H – Arg-58 to His mutant
RP-HPLC – Reverse phase high performance liquid chromatography
RPM – Revolutions per minute
SDS – Sodium dodecyl sulphate
SE-HPLC – Size exclusion high performance liquid chromatography

Ser – Serine
sfDMEM – Serum free Dulbecco's modified Eagle's medium
SLS – Static light scattering
S/N – Signal-to-noise ratio
S/V – Surface-to-volume ratio
SP – Sulphopropyl
$\boldsymbol{T}$ **–** Temperature
$\boldsymbol{T_c}$ **–** Critical temperature
$\boldsymbol{T_{clear}}$ – Clearing temperature
$\boldsymbol{T_{cloud}}$ – Clouding temperature
TEMED – Tetramethylethylenediamine
Thr – Threonine
$\boldsymbol{T_{ph}}$ – Phase transition temperature
$\mathbf{t_R}$ – Retention time
Tris-HCl – Tris(hydroxymethyl) aminomethane hydrochloride
Trp – Tryptophan
Tyr – Tyrosine
UV – Ultraviolet
UV/Vis – Ultraviolet/Visible
Val – Valine
VDW – van der Waals
v/v – volume per volume
w/v – weight per volume
$\mathbf{x_L}$ – Fraction of labelled protein
$\mathbf{x_m}$ – Fraction of modified protein
$\boldsymbol{\varepsilon}$ – Extinction co-efficient
$\boldsymbol{\phi}$ – Volume fraction
$\boldsymbol{\phi_c}$ **–** Critical volume fraction
$\boldsymbol{\lambda}$ – Wavelength
$\boldsymbol{\lambda_{ex}}$ – Excitation wavelength
$\boldsymbol{\lambda_{em}}$ – Emission wavelength

Table of Contents

Chapter 1:

Introduction

1.1 Self-assembly

Self-assembly can be defined as “the autonomous organization of components into patterns or structures without human intervention” and is a ubiquitous process throughout nature and technology (Whitesides and Boncheva 2002; Whitesides and Grzybowski 2002). There are two types of self-assembly: static self-assembly involving systems at equilibrium and dynamic self-assembly which requires dissipation of energy within the system (Whitesides and Grzybowski 2002). The ability of components in a system to self-assemble arises from the characteristics of the components as well as their environment (Whitesides and Grzybowski 2002). Protein self-assembly plays a crucial biological role and understanding the mechanisms underlying the self-assembly of proteins is of considerable importance for many biological and industrial processes (McManus, Charbonneau et al. 2016).

1.2 Thermodynamics of self-assembly

Thermodynamic constraints are the driving forces leading to self-assembly (Grzybowski, Wilmer et al. 2009). Thermodynamics is the study of transformations of energy, the basic concepts of which are work, heat and energy (Atkins 1988). The First Law of Thermodynamics states that the energy of an isolated system is constant and is defined mathematically in equation 1.1:

$$\Delta U = q + w \qquad \textbf{1.1}$$

where ΔU is the change in internal energy of a closed system, q is the heat supplied to a system and w is work done on a system. At constant volume, ΔU and q are equal.

Enthalpy (H) pertains to systems where heat (q) is supplied to a system at a constant pressure and no other work is being done on the system. H is defined mathematically in equation 1.2:

$$H = U + pV \qquad \textbf{1.2}$$

where H is the enthalpy, U is the internal energy and p is pressure and V the volume of the system. At constant pressure, ΔH is equal to q.

Entropy (*S*) is a term denoting the disorder of a system and is defined by equation 1.3:

$$S = k_B \log W \qquad \textbf{1.3}$$

where k_B is the Boltzmann constant equal to 1.38 x 10^{-23}J K^{-1} and *W* denotes the number of microstates in the system.

Free energy is related to *H* and *S* through the Gibbs function (*G)* at constant temperature (*T*) given in equation 1.5:

$$\Delta G = \Delta H - T\Delta S \qquad \textbf{1.4}$$

If Δ*G* is negative, a reaction has a tendency to convert reactants into products and is said to be spontaneous. Conversely, when Δ*G* is positive, the reaction is not spontaneous. When Δ*G* is zero, a system is said to be in thermodynamic equilibrium. The chemical potential of a sample is uniform when Δ*G* is zero regardless of the number of phases present (Atkins 1988).

1.3 Proteins

Proteins are polymers composed of chains of amino acids which, in many cases, fold into a unique three-dimensional structure. Proteins are synthesised in ribosomes and are generally divided into two distinct classes: globular and transmembrane. Globular proteins are water-soluble and are found in the cytosol of the cell as well as in intracellular and bodily fluids (Whitford 2005). Transmembrane proteins are found in the cellular membrane and contain a hydrophobic core region that interacts with the hydrophobic tails of the lipid bilayer which is flanked by two hydrophilic caps (Alberts, Johnson et al. 2002). Both classes of protein have functionalities that are derived from their folding and self-assembly.

1.3.1 Protein structure and folding

Amino acids are the building blocks of proteins. There are over 300 naturally occurring amino acids but proteins are comprised of only 22 (Rodwell, Botham et al. 2015), all of which consist of a central carbon atom (known as the α-carbon) connected to an amino functional group, a carboxyl functional group, a hydrogen and a side chain (R-group) (figure 1.1). Selenocysteine (Sec) and pyrolysine (Pyl) are

known as the 21st and 22nd amino acids and, though found in all three domains of life (archae, prokaryotes and eukaryotes), are extremely rare (Zhang and Gladyshev 2007). Amino acids are zwitterionic and differ only in the chemical make-up of their side-chains which vary in complexity but can only contain carbon, hydrogen, nitrogen, oxygen and sulphur (figure 1.1). Although amino acids (excluding Glycine) can have L and D forms, the vast majority of proteins contain amino acids in the L enantiomer. Amino acids can be classed according the chemical properties of their side chains as acidic, basic, hydrophobic, hydrophilic and neutral.

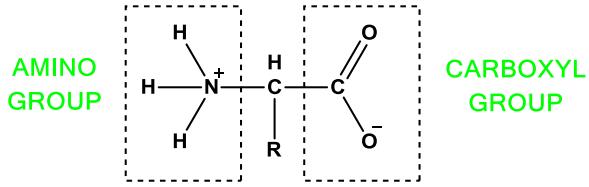

Figure 1.1: *An amino acid showing the amino group, carboxyl group and R group.*

Amino acids are covalently bonded *via* a peptide (amide) bond after a condensation reaction forming polypeptides and proteins (figure 1.2)**.** The unique sequence of the amino acids between the N- and C- terminals in a protein is called its primary structure.

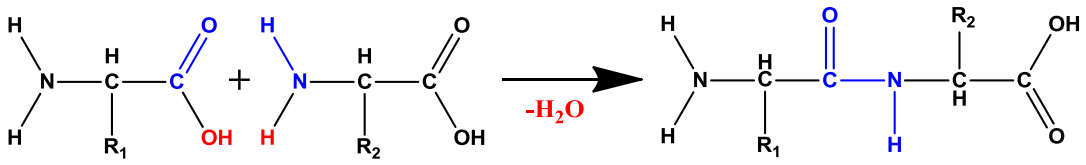

Figure 1.2: *Condensation of water to form a peptide bond between two amino acids.*

The regular repeating local structures arising from the amino acids in the primary structure in a protein give rise to its secondary structure. There are two main classifications of secondary structure: α-helices (figure 1.3) and β-sheets (figure 1.4). The α-helix is a right-handed helix with 3.6 amino acid residues per turn and a distance (known as a translation distance) of 0.15 nm between amino acids resulting in a pitch of 0.54 nm. It is the most common secondary structure motif. The backbone of the α-helix is stabilised by hydrogen bonds.

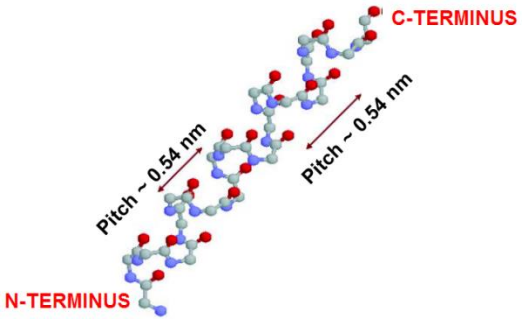

Figure 1.3: *Right handed α-helix. Taken from Whitford 2005.*

β-sheets are formed from β-strands which are elongated helical arrangements with two residues per turn, a translation distance of 0.34 nm and a pitch of 0.7 nm. These strands interact to form a sheet-like arrangement which is stabilised by hydrogen bonding.

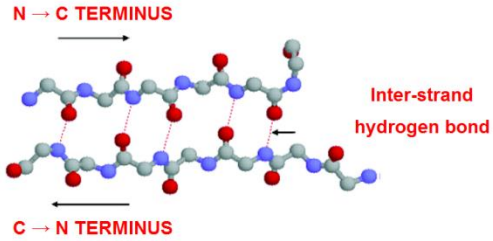

Figure 1.4: *Two adjacent β-strands interacting via hydrogen bonds. Taken from Whitford 2005.*

Protein tertiary structure arises from the folded polypeptide chain and is stabilised by hydrogen bonds (as for the secondary structure) in addition to disulphide bridges (formed by the oxidation of the thiol side chains in Cys residues), hydrophobic interactions between the non-polar residues, electrostatic interactions between charged residues and van der Waals interactions arising between dipoles. Predicting a protein's tertiary structure from its amino acid sequence is extremely difficult and, in general, protein structures are determined experimentally rather than being inferred or calculated on the basis of their sequence (Dill and MacCallum 2012). Protein folding is reversible but the native folded state is more stable than the unfolded state due to the decrease in free energy associated with protein folding. The theoretical time for a protein to explore all of the possible conformations during folding is in the region of 10^{73} years (Cooper 2011). However, proteins fold in timescales of between microseconds to minutes. This is known as the Levinthal

Paradox (Levinthal 1968). When a protein's secondary or tertiary structure is disrupted, the protein is said to be 'denatured' and protein functionality is diminished. Denaturation can occur either reversibly or irreversibly as a result of extremes of heating or pH or by the addition of denaturants such as chaotropes, solvents, cross-linkers or reducing agents (Whitford 2005).

Protein quaternary structure arises when proteins containing more than one polypeptide chain (also known as subunits) are held together by disulphide bonds, hydrophobic interactions, hydrogen bonds and salt bridges. For some proteins such as haemoglobin, the formation of its tetrameric quaternary structure is essential for its function (Gunton, Shiryayev et al. 2007).

1.4 Protein condensation

1.4.1 Background

Protein condensation can refer to the native or non-native assembly of proteins that results in the formation of condensed states such as amorphous solids, crystals, dense protein liquids, gels or polymer fibres (Gunton, Shiryayev et al. 2007). These assemblies can form either by design, which is seen with the self-assembly of viral capsids, or when something goes wrong, which is seen with the polymerization of haemoglobin S (HbS) (Eaton and Hofrichter 1990; Hagan 2014; McManus, Charbonneau et al. 2016). Understanding protein condensation is of considerable importance for both biological and industrial processes (McManus, Charbonneau et al. 2016). Protein condensation is also implicated in the pathogenesis of several diseases which are described in section 1.4.2.

1.4.2 Protein condensation diseases

The main pathological feature of protein condensation diseases, which include sickle cell anaemia, Alzheimer's disease and cataract, is the loss of protein solubility that leads to a condensed protein phase (Annunziata, Ogun et al. 2003; Chen, Vekilov et al. 2004). The insoluble protein condensates formed as a feature of the pathogenesis of these diseases have impaired biological function. The pathologies of some of these condensation diseases are discussed below.

Sickle cell disease is one of the oldest known protein condensation diseases and was first diagnosed over a century ago (Ferrone 2016). Individuals suffering from sickle cell disease have inherited two mutated genes while those who have inherited only one mutated gene are said to have sickle cell trait (Gunton, Shiryayev et al. 2007). Normal haemoglobin (HbA) is a tetrameric globular protein consisting of two α and two β subunits. Sickle haemoglobin (HbS) is a mutant of HbA associated with sickle cell anaemia. HbS accounts for over half of the haemoglobin in patients suffering from sickle cell anaemia (Steinberg 1999). Red blood cells in individuals suffering from sickle cell anaemia are distorted, with some resembling a sickle as opposed to being disc-shaped. The distortion in cell shape associated with sickle cell anaemia is caused by a two-step homogeneous/heterogeneous nucleation process that leads to polymerization of deoxygenated HbS to form rigid linear fibers that distort the shape of the cell (figure 1.5). The polymerization of HbS is a result of a single mutation of a negatively charged Glu to a hydrophobic Val at position 6 on the surface of the β subunits of HbS that, in turn, leads to its polymerization. The binodal for HbS is unusual in the sense that it has a critical point that represents the lowest temperature at which LLPS occurs as opposed to the highest (Serrano, Galkin et al. 2001; Galkin, Chen et al. 2002).

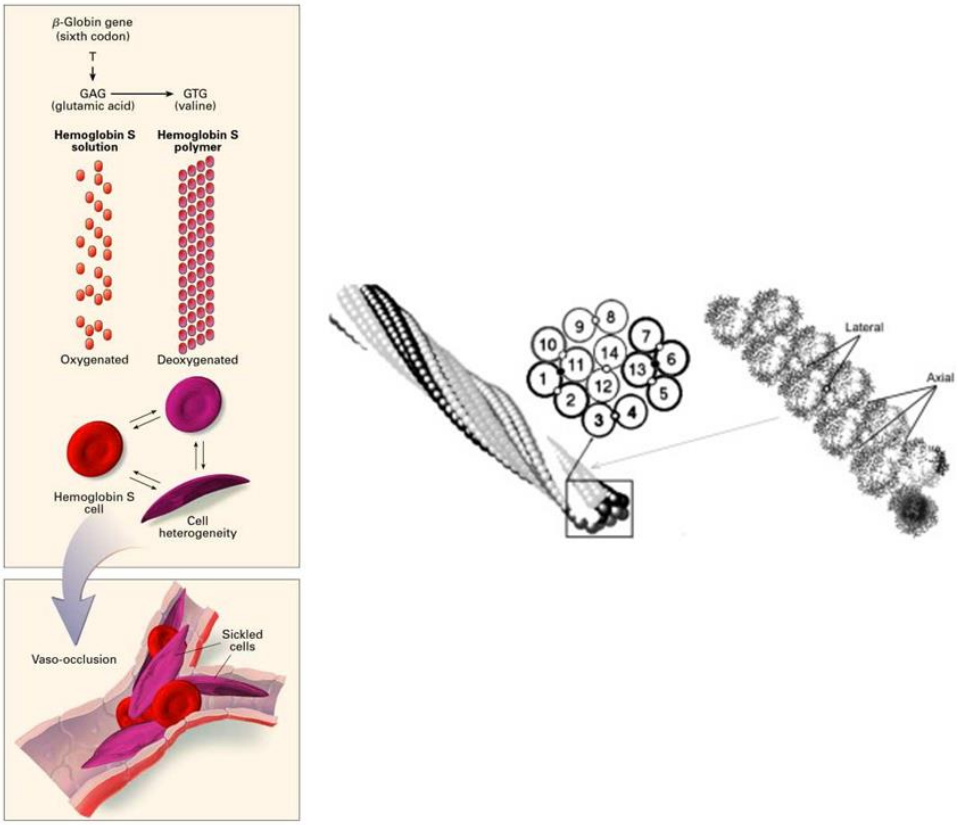

Figure 1.5: *The pathophysiology of sickle cell disease (left) and the structure of sickle cell fibers (right). Taken from Steinberg 1999 and Ferrone 2016.*

Homozygous haemoglobin C (CC) disease is a form of anaemia involving condensed protein states and is also due to a mutation at site 6 in the β subunit of the haemoglobin tetramer involving the replacement of a Glu with a positively charged Lys which leads to a decreased solubility (Hunt and Ingram 1958; Tokumasu, Nardone et al. 2009). HbC forms crystals in oxygenated red blood cells of patients presenting with CC disease as well as in patients with SC disease who are heterozygous and produce HbS and HbC (Feeling-Taylor, Yau et al. 2004). The latter is considerably more severe than the former and is potentially life threatening (Nagel and Lawrence 1991). In patients suffering from CC and SC disease, crystallization occurs due to an increase in hydrophobic interactions as a result of a conformational change which arises from the mutation at site 6 (Feeling-Taylor, Yau et al. 2004). The solubility line for HbC indicates that it has a retrograde temperature dependence i.e. the solubility decreases as temperature increases (Feeling-Taylor, Banish et al. 1999; Vekilov, Feeling-Taylor et al. 2002).

Alzheimer's disease (AD) is the most prevalent form of age-related dementia with the majority of cases occurring in individuals aged 65 or over (Williamson, Goldman et al. 2009). Amyloid β (Aβ) is a globular protein associated with AD (Gunton, Shiryayev et al. 2007). There are two major forms of Aβ: Aβ40 and Aβ42 (Gu and Guo 2013). Proteolysis of amyloid precursor protein (APP) by γ-secretase at the C-terminus results in peptides of varying lengths (Blennow, de Leon et al. 2006). The sole difference between Aβ40 and Aβ42 is their length, which differ by two hydrophobic residues to produce peptides which are 40 and 42 residues long respectively (Gu and Guo 2013; Cukalevski, Yang et al. 2015). The two hydrophobic residues in Aβ42 make it more prone to aggregation than its more abundant Aβ40 counterpart and thus it is more commonly associated with AD (Cukalevski, Yang et al. 2015). Amyloid plaques are compacted deposits of Aβ fibrils and are a key characteristic of AD (figure 1.6) (Selkoe and Podlisny 2002). Abnormally hyper-phosphorylated tau proteins, which are associated with AD as well as other neurodegenerative diseases known as 'taupathies', form neurofibrillary tangles and also contribute to the pathology of AD (figure 1.6) (Iqbal, Liu et al. 2010; Iqbal, Liu et al. 2016).

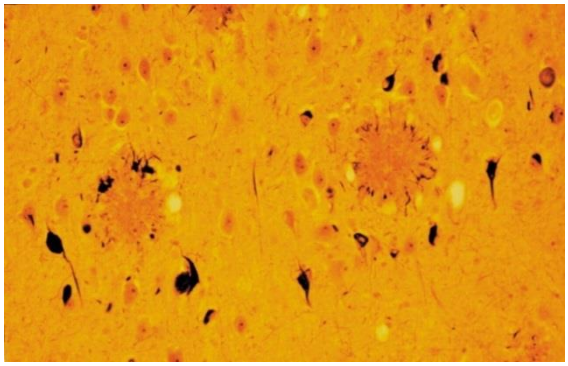

Figure 1.6: *A photomicrograph of a section of the amygdala of a patient with AD showing two amyloid plaques and numerous neurofibrillary tangles. Taken from Selkoe and Podlisny 2002.*

There are two types of AD: early onset familial AD and late onset sporadic AD. Familial AD is extremely rare and is as result of mutations in APP and presenilin genes (Blennow, de Leon et al. 2006). The more common sporadic AD is thought to occur as a result of an age-induced imbalance between the production and clearance of Aβ (Blennow, de Leon et al. 2006).

Other condensed protein diseases that result in a loss of proper protein function that in turn lead to neurological disorders include Parkinson's Disease (PD, caused by aggregation of a protein called α-synuclein (α-syn) into amyloid fibrils (Galvagnion, Buell et al. 2015)) and Huntington's Disease (HD, which involves the formation of aggregates of huntingtin protein (HTT) (Ross and Tabrizi 2011)).

Cataract, which is the leading cause of blindness worldwide, can arise due to aging which leads to modifications such as deamidation, methylation or disulphide bond formation, damage from UV radiation, an altered environment due to chronic conditions such as diabetes or be genetic in origin (Banerjee, Pande et al. 2011; Rivillas-Acevedo, Fernández-Silva et al. 2015). Cataract is defined as any opacity of the crystallin lens (Hejtmancik 2008) (figure 1.7). Lens opacity arises as a result of substantial variations in the refractive index over distances that are comparable to the wavelength of visible light (Benedek 1971; Delaye and Tardieu 1983) due to the formation of aggregates with high molecular weights within the ocular lens which scatter light and impair vision (Thurston, Hayden et al. 1997).

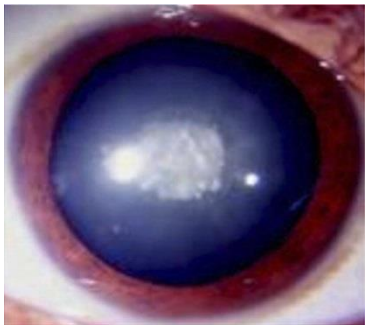

Figure 1.7: *A cataractous lens showing opacity. Taken from Messina-Bass et al. 2006.*

The high molecular weight aggregates that cause light scattering are held together by various attractive forces which include covalent interactions arising from disulphide or glycosidic linkages or non-covalent interactions arising from hydrophilic, electrostatic and van der Waals interactions (Benedek, Pande et al. 1999). Cataract can be categorized based on the age of onset: congenital cataract usually occurs in the first year of life; juvenile cataract usually occurs in the first decade of life; pre-senile cataract occurs before the age of 45 and age-related cataract occurs after this time (Shiels and Hejtmancik 2013). Congenital (or hereditary) cataract arises due to a decrease in protein solubility as a result of single mutations in the amino acid

sequence of proteins found in the eye lens (Rivillas-Acevedo, Fernández-Silva et al. 2015). Some of these mutations are discussed in section 1.9.2.

1.5 Protein interactions

The macroscopic thermodynamic properties and phase behaviour of proteins in solution that are critical to their function are determined by intermolecular protein-protein interactions (Quang, Sandler et al. 2014). The intermolecular forces between proteins that govern its behaviour are no different from those between any other types of molecules (Leckband and Israelachvili 2001). Non-covalent protein interactions can be either specific or non-specific. Specific protein interactions (also known as complementary, lock-and-key and ligand-receptor and recognition interactions) arise synergistically from geometric, steric, ionic and directional forces to give a strong bond e.g. the avidin-biotin complex which boasts the strongest known non-covalent bond with a dissociation constant $K_d = 10^{-15}$ M (which translates to a binding energy of ~35 *kT* or ~20 kcal mol^{-1} at room temperature) (Leckband and Israelachvili 2001; General, Dragomirova et al. 2012). In general, the sum of the non-covalent attractive and repulsive interactions between proteins gives rise to a protein interaction potential. Attractive interactions arise from the hydrophobic effect, van der Waals interactions and hydrogen bonding, while repulsive interactions are due to electrostatics and hydration interactions. These are outlined below (table 1.1 and figure 1.8). Protein-protein interactions arising from non-covalent forces release substantially less energy than for the formation of a typical covalent bond (1–5 kcal mol^{-1} for a non-covalent interaction versus ~50–200 kcal mol^{-1} for covalent bond formation) and, so, protein-protein interactions are mediated by weak intermolecular forces at room and physiological temperatures (Lodish, Berk et al. 2000). However, multiple non-covalent interactions can work together to stabilise a protein and are also essential in mediating the binding of enzymes with substrates and antibodies with antigens (Lodish, Berk et al. 2000). The sum of the attractive and repulsive contributions arising from the amino acid side chains of a protein gives rise to its net interaction potential which is also known as the potential of mean force (PMF) (Israelachvili 1992). The energy (*E*) of the interaction potential is distance (*D*) dependent and related to the force (*F)* by equation 1.5 (Leckband and Israelachvili 2001):

$$F(D) = \frac{-\Delta E(D)}{\Delta D} \qquad \textbf{1.5}$$

Table 1.1: *Non-covalent interactions and their relevant interaction type, strength and range.*

Interaction	Attractive (–)/ Repulsive (+)	Strength	Range
van der Waals	–	Weak	Short → long
Hydration	+	Strong	Short
Hydrophobic	–	Strong	Long
Steric	+	Strong	Short
Hydrogen bonding	–	Weak	Short
Electrostatic	+/ –	Weak → strong	Short → long

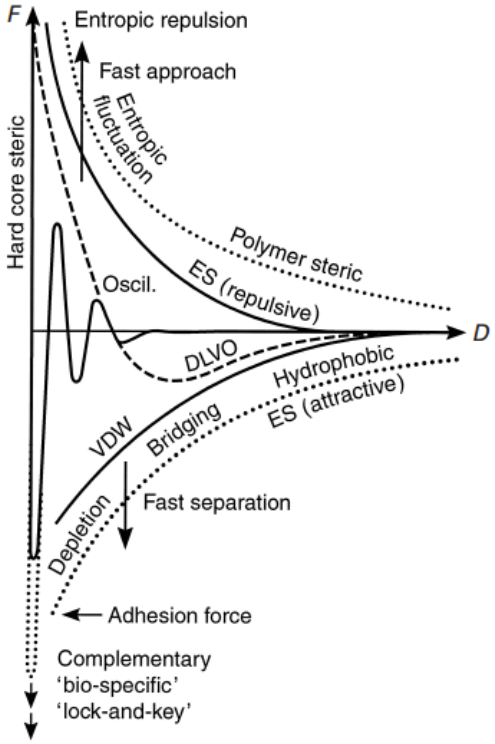

Figure 1.8: *Generic attractive and repulsive interaction potentials between two protein molecules. Taken from Leckband and Israelachvili 2001.*

van der Waals (VDW) forces act between all atoms and molecules and arise due to fluctuations in the electric dipole moments of molecules which give rise to an attractive force as the atoms or molecules approach one another (Israelachvili 1992;

Leckband and Israelachvili 2001). The strength of VDW forces has a power law dependence on the separation distance (*D*) between the two bodies on which it is acting (equation 1.6)

$$E(D) = \frac{-C_{VDW}}{D^6} \qquad \textbf{1.6}$$

where C_{VDW} is dependent on the optical properties and geometries of the acting bodies (Leckband and Israelachvili 2001).

The hydrophobic effect is a strong, long-range, entropy-driven attractive force between non-polar molecules, the strength of which is dependent on the hydrophobicity of the interacting non-polar molecules (Leckband and Israelachvili 2001). When a hydrophobic molecule comes into contact with water (or another polar solvent), it forces the water molecules to form a rigid cage-like structure around the hydrophobic molecule (Lodish, Berk et al. 2000). The restricted motion of the polar solvent surrounding non-polar molecules is energetically unfavourable (Israelachvili 1992). The long-range attraction that exists between non-polar molecules causes them to come together and this leads to an increase in entropy by releasing the rigid polar molecules in the solvent, thereby reducing free energy (Israelachvili 1992). The hydrophobic effect is important for protein stability and contributes to the tendency of proteins to bury non-polar amino acid side-chains in the proteins core and thus exclude water (General, Dragomirova et al. 2012).

Hydration forces (also known as solvation forces) are repulsive interactions arising between molecules with surfaces that have water tightly bound to chemical groups with hydrophilic nature (Israelachvili 1992). The repulsive nature of these forces is two-fold and is energetically driven in terms of both entropy and enthalpy (Leckband and Israelachvili 2001). An increased excluded volume around the hydrated molecule as a result of the surface-bound water is an entropy driven process while a decrease in the adhesion strength of the surface is enthalpy driven (Leckband and Israelachvili 2001).

Steric interactions are repulsive forces that exist between molecules that contain a hydrophilic polymer (e.g. PEG) bound to the surface as they approach one another in solution (Leckband and Israelachvili 2001). However, if the attached polymer

contains functional groups that can attractively interact with another molecule at once within range, **bridging** is said to occur (Leckband and Israelachvili 2001). When polymers are free in solution they can cause a **depletion** force which is an entropy driven event that leads to an attractive force between proteins that interact *via* a depletion zone that is free of the hydrophilic polymer as water is driven out into the bulk solution which decreases the free energy of the system (Leckband and Israelachvili 2001; Thurston 2006).

Hydrogen bonding is a type of weak electrostatic attraction that occurs between a hydrogen atom (H) that is covalently bonded to an electronegative atom (known as a donor atom (D)) to give a polar molecule that forms a weak attraction with an electronegative acceptor atom (A) (figure 1.9) (Lodish, Berk et al. 2000; Leckband and Israelachvili 2001).

$$D^{\delta-} - H^{\delta+} + : A^{\delta-} \leftrightarrow D^{\delta-} - H^{\delta+} \cdots\cdots : A^{\delta-}$$

Figure 1.9: *Attractive interaction between a covalently bound hydrogen and a polar molecule.*

The strength of the hydrogen bond in water is in the region of 4 to 5 kcal mol^{-1} and between 1 and 2 kcal mol^{-1} for biomolecules (Lodish, Berk et al. 2000). Hydrogen bonding between amino acids is responsible for stabilising the alpha helix and beta sheet components of protein secondary structure and is one of the major contributors to a protein stability (Pace, Scholtz et al. 2014)

Electrostatic (ES) interactions can be both attractive and repulsive and arise due to an interaction between a negatively charged atom/molecule (anion) and a positively charged atom/molecule (cation) where atoms/molecules of opposite charge attract one another and atoms/molecules of the same charge repel one another. In an ionic species, bonding electrons are never shared. A salt-bridge (or ion pair) contributes to protein stability and is formed between side-chain groups in a protein that are of opposite charge and within 5 Å of each other (General, Dragomirova et al. 2012). Zeta potential (also known as electrokinetic potential) is a measure of the effective electrostatic interaction between two charged surfaces (such as in a protein). A charged particle in buffer will form an adsorbed layer called an electric double layer (EDL) around itself (figure 1.10). The inner layer, known as the Stern layer, consists of particles with an opposite charge to that of the particle surface. A diffuse, variable

outer layer comprising of charged atoms/molecules of both signs surrounds the Stern layer and is separated from the bulk solution by a hypothetical interface known as the slipping plane. The Zeta potential is the potential at this interface (Bhattacharjee 2016; An, Huang et al. 2017).

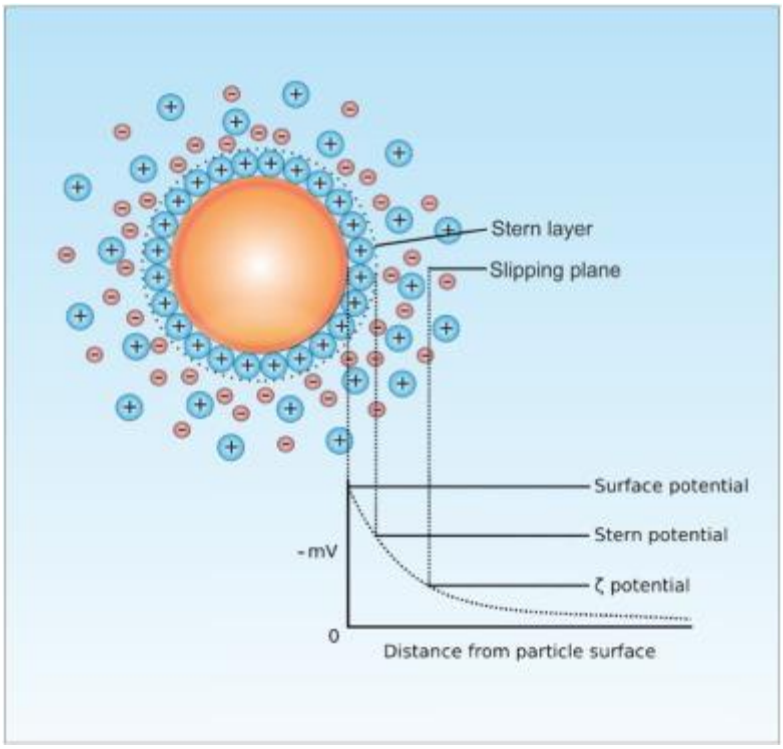

Figure 1.10: *Schematic representation of electric double layer. Taken from An, Huang et al. 2017.*

Despite these individual non-covalent interactions being well characterised, their collective contribution to protein stability and self-assembly is far less well understood (Israelachvili 1992; Vekilov and Chernov 2003; Gunton, Shiryayev et al. 2007; Fusco and Charbonneau 2016). Protein stability and self-assembly are discussed in more detail in section 1.7.

1.6 Colloids

1.6.1 Colloid systems

While there is no defined distinction between colloidal and non-colloidal systems, the term ‘colloid science’ refers to systems where one or more of the components has at least one dimension that is between a nanometre and a micron in length scale (Shaw 1992). There are three unique classes of colloidal system: colloidal

dispersions, colloidal solutions and association colloids. Colloidal dispersions are subsequently classified according to the two phases involved. A *sol* is a dispersion of a solid in a liquid or a solid in a solid; An *aerosol* is a dispersion of liquid or a solid in a gas; An *emulsion* describes a dispersion of a liquid in a liquid; A *foam* is a dispersion of a gas in a liquid. Colloids can be further classified according to how well they attract (lyophilic or hydrophilic) or repel (lyophobic or hydrophobic) a solvent or water (Atkins 1988).

1.6.2 Isotropic colloids

The hard sphere (HS) model was initially developed by theorists to understand liquids (Jones 2002). Hard spheres are thought of as perfect spheres with an interparticle potential equal to zero unless they overlap in which case the potential becomes infinite (Jones 2002). Hard sphere (HS) systems are one of the simplest forms of interacting systems and neither repel nor attract at distances greater than the particle diameter (Royall, Poon et al. 2013). Theoretically, the particle volume fraction (ϕ) is the only parameter that determines the behaviour of a HS system. In reality, however, colloids will display a degree of 'softness' (Pusey and van Megen 1986; Royall, Poon et al. 2013).

1.6.3 Anisotropic colloids

Patchy particles are colloids that exhibit anisotropy in their shape and/or surface properties. The self-assembly of patchy particles, designed to interact in an anisotropic manner is of great interest in the field of soft matter (Bianchi, Largo et al. 2006; Bianchi, Blaak et al. 2011). In addition, patchy particle models have been developed where an isotropic interaction potential has been replaced with an anisotropic one to explain their behaviour (Glotzer and Solomon 2007; Pellicane, Smith et al. 2008; Fusco, Headd et al. 2014).

1.6.4 Colloidal interactions

The net interaction potential between atoms or molecules in a system is given by the sum of its attractive and repulsive contributions (Everett 1988). However, forces between particles have a much longer interaction range than atoms or molecules which must be taken into account (Everett 1988). In a colloidal system, there are several physicochemical factors that account for the behaviour of the system: colloid

size, colloid shape, surface properties, colloid-colloid interactions and colloid-solvent interactions (Shaw 1992). Colloids also display phase transitions between gas, liquid, solid and liquid crystalline phases depending on the range of the inter-particle attraction (figure 1.10) (Anderson and Lekkerkerker 2002). These phase transitions have been observed for globular protein solutions (Rosenbaum, Zamora et al. 1996; Asherie 2004). These properties allow for direct parallels to be drawn between colloids and proteins. Indeed, the tuneable electrostatic interactions of proteins made possible by varying pH and ionic strength as well as the monodisperse size distribution of a single protein system make proteins ideal candidates to understand equilibrium and non-equilibrium phase behaviour of colloidal solutions (Dorsaz, Thurston et al. 2008). Since protein solutions can be classed as colloidal fluids due to the their submicron size, colloidal systems with defined regions of attraction have been successfully used as a means of gaining insights into protein phase behaviour (Dorsaz, Thurston et al. 2008).

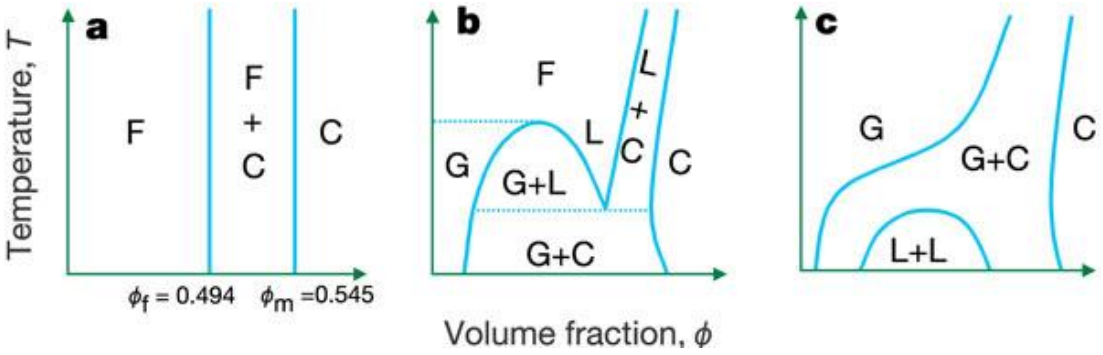

Figure 1.10: *Phase diagrams for (a) hard-sphere system showing fluid (F) and crystal (C) phases, (b) hard-spheres with long range attraction showing gas (G), liquid and fluid phases and (c) a system incorporating short-range attraction. Taken from Anderson and Lekkerkerker 2002.*

1.6.5 Modelling colloidal interactions

Simplified coarse-grained models that are representative of protein-protein interactions are the preferred simulation method because direct molecular simulations are computationally demanding (Quang, Sandler et al. 2014). Several potentials such as adhesive hard sphere (Rosenbaum, Zamora et al. 1996), square-well (Liu, Garde et al. 2005), modified Lennard-Jones (Wolde and Frenkel 1997), Yukawa (Foffi, McCullagh et al. 2002) and Derjaguin, Verway, Landau and Overbeek (DVLO) (Pellicane, Costa et al. 2004) potentials treat protein molecules as

hard spheres with an implicit solvent potential and parameters for interaction strength, interaction range and particle diameter (Quang, Sandler et al. 2014) and can qualitatively predict certain features of the phase diagram. However, to quantitatively reproduce the phase diagram for a protein, it is imperative to account for the anisotropy of the protein surface. Proteins are inherently anisotropic as a result of their unique surface, varying charge distribution and, to a lesser extent, their non-spherical shape (Quang, Sandler et al. 2014; McManus, Charbonneau et al. 2016). This anisotropy affects the way a protein behaves in solution. The chemical anisotropy of the surface plays a more crucial role in determining phase behaviour than the shape anisotropy (McManus, Charbonneau et al. 2016). The seminal work of Lomakin, Asherie et al. (1999) incorporated an anisotropic (or aeolotopic) potential into a simple model and accurately reproduced the liquidus and binodal lines (discussed in more detail in 1.7.4) of the experimental phase diagram for bovine γ_{IIIb} crystallin. Other models have been developed since then that also account for the anisotropy dependent phase behaviour of globular proteins (Sear 1999; Romano, Sanz et al. 2011; Dorsaz, Filion et al. 2012; Quinn, Gnan et al. 2015).

1.7 Protein stability, phase transitions and phase diagrams

1.7.1 Protein stability

Aggregation can be defined as the aberrant assembly of two or more protein monomers (either folded or partially/fully unfolded or a combination of both). Aggregates may be reversible (which tends to be thermodynamically driven or irreversible which tends to be kinetically driven (Roberts 2014) and can range in size between nanometres and micrometres in size (figure 1.11) (McManus, Charbonneau et al. 2016).

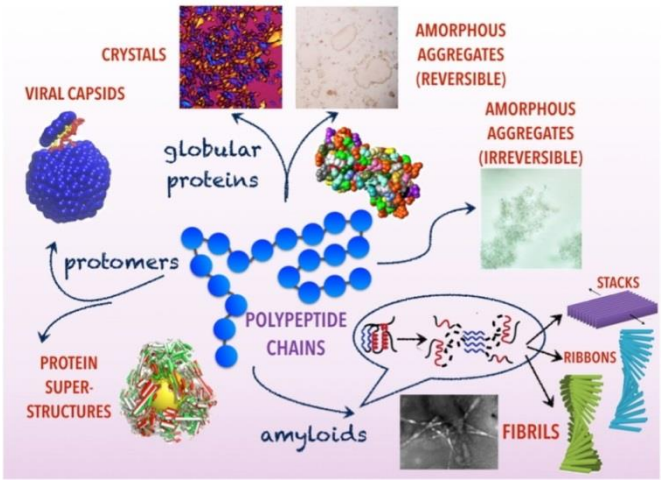

Figure 1.11: *Various protein and peptide self-assemblies.*
Taken from McManus, Charbonneau et al. 2016.

One way to think about protein stability is the ability of a protein to maintain its native state after folding and is of considerable importance in the biopharmaceutical and food industries since proteins in solution need to remain soluble during production, transport and storage (Amin, Barnett et al. 2014). Structural stability is the ability of a protein to resist denaturation and solution (or colloidal) stability is the resistance of the protein to self-associate in solution (James and McManus 2012; Blumlein and McManus 2013; Quigley and Williams 2015). Both the structural and solution stability of a protein can depend on the solution conditions (such as ionic strength, temperature, pH, salt type and concentration, and the presence of excipients, ligands and cosolutes, etc.) as well as the inherent properties of the protein itself (James and McManus 2012; Blumlein and McManus 2013; Wang, Latypov et al. 2014). Protein aggregation can occur as a result of either colloidal or thermal instability and is a major biopharmaceutical regulatory concern since the formation of aggregates in a biotherapeutic can impact efficacy as well as patient safety and minimizing the rate of aggregation is of particular interest in the development of biotherapeutics (Wang and Roberts 2010).

1.7.1.1 Aggregate morphology

Protein aggregates are usually categorised as amorphous or fibrillar. Amorphous aggregates lack long range order and are composed of native and/or unfolded protein (Wang and Roberts 2010). Fibrillar aggregates, so called due to their long fibrous

shape, have long range order and have been shown to contain mainly β-sheet structure (Wang and Roberts 2010).

1.7.1.2 Mechanisms of protein aggregation

There are five common mechanisms associated with protein aggregation. Although these mechanisms are distinct, they are not mutually exclusive and can occur simultaneously within the same system (Philo and Arakawa 2009). Understanding these mechanisms can help to develop methods to prevent or diminish the undesirable formation of aggregates (Philo and Arakawa 2009).

1.7.1.2.1 Mechanism 1: Reversible association of the native monomer

Reversible association of native monomer (i.e. monomer in its natively folded state) to form oligomers occurs due to the inherent patchiness and resulting complementarity of the protein surface (Fink 1998; Philo and Arakawa 2009). Irreversible aggregation can be mediated by inter-protein disulphide bridges. As protein concentration increases and larger oligomers are formed, these aggregates may become irreversible over time (Philo and Arakawa 2009).

1.7.1.2.2 Mechanism 2: Aggregation of conformationally altered monomer

Proteins that are conformationally altered (i.e. partially unfolded) due to heat or shear stress can form strong oligomers in a manner similar to the mechanisms associated with reversible association of native monomer (Roberts 2007; Philo and Arakawa 2009). Proteins that associate *via* Mechanism 2 have a low propensity to aggregate in their native states and the principal difference between these two mechanisms is the state of the associating protein (i.e. folded or unfolded) (Philo and Arakawa 2009).

1.7.1.2.3 Mechanism 3: Aggregation of chemically modified product

Mechanism 3 is similar to mechanism 2 in the sense that a conformational change is required to induce aggregation. The origin of the structural change is usually chemical degradation caused by oxidation, deamidation, photo-decomposition or proteolysis which change the interaction potential of the protein (Philo and Arakawa 2009; Wang, Nema et al. 2010). Aggregates formed *via* Mechanism 3 can consist of

both chemically modified monomer as well as native monomer (Philo and Arakawa 2009).

1.7.1.2.4 Mechanism 4: Nucleation controlled aggregation

Nucleation controlled aggregation involves monomer addition of native protein and can proceed either homogeneously *via* monomer addition to a critical nucleus composed of the native monomer or heterogeneously where an impurity or contaminant has acted as a critical nucleus (Philo and Arakawa 2009; Wang and Roberts 2010).

1.7.1.2.5 Mechanism 5: Surface induced aggregation

Surface induced aggregation begins with reversible binding of native monomer to a surface which then triggers a conformational change in the protein that in turn drives aggregation *via* a mechanism similar to Mechanism 2 (Philo and Arakawa 2009; Wang, Nema et al. 2010). Mechanism 4 can be considered a special case of Mechanism 5 since the surface of the critical nucleus is driving aggregation (Philo and Arakawa 2009).

1.7.2 Protein phase transitions and phase diagrams

Thermodynamically-driven protein phase transitions proceed without a change in protein structure and can give rise to a number of condensed protein phases and states such as crystals, concentrated liquids *via* liquid-liquid phase separation (LLPS) and gels (Dumetz, Chockla et al. 2008; Vekilov 2010). For a phase transition to occur, the change in Gibbs free energy in the system ($\Delta G = \Delta H - T\Delta S$) must be negative (Wentzel and Gunton 2008; Haaga, Pemberton et al. 2016). A system is said to be in thermodynamic equilibrium when ΔG is zero (Vekilov 2010).

The second virial coefficient B_{22} (also known as B_2 and A_{22}) is a measure the strength of a two body (e.g. protein-protein) interaction in a dilute solution and has been used as a means of measuring the tendency of a protein-protein to self-associate i.e. its colloidal stability (Gunton, Shiryayev et al. 2007). It is related to the interaction potential $W(r)$ by equation 1.7:

$$B_{22}M^2 = 2\pi N_A \int_a^{\infty} (1 - e^{-W(r)/k_B T}) r^2 dr \qquad \textbf{1.7}$$

where M is the molecular mass, N_A is Avogadro's Number, $W(r)$ is the interaction potential, k_B is Boltzmann's constant, T is the temperature and r is the centre to centre separation. A positive B_{22} value indicates that interactions between the protein and solvent are favoured so a repulsive protein-protein interaction dominates and the solvent is said to be a "good solvent". A negative B_{22} value indicates that an attractive protein-protein interaction dominates and the solvent is said to be a "bad solvent". A "theta solvent" is said to be an ideal solvent and is one where the B_{22} value is zero (George and Wilson 1994). George and Wilson (1994) found that crystallization of proteins occurs within a narrow range of slightly negative B_{22} values. As negativity increases (i.e. attraction increased), so does the propensity of a protein solution to form an amorphous precipitate. Understanding the conditions under which crystallization occurs has long been the subject of interest for both the elucidation of protein structure by means of X-ray crystallography and as a method of protein purification (Giegé 2013). Some proteins can form crystals with more than one morphology which can coexist (Vekilov 2012; James, Quinn et al. 2015). In this case, the solubility of each crystal type differs.

Liquid-liquid phase separation (LLPS), also known as coacervation or fluid-fluid phase separation, is a phase transition where, under a given set of conditions, a solution (which can be a simple liquid, a colloidal suspension, or a protein solution) separates into two liquid phases that can be macroscopically separated: a dense phase and a dilute phase (figure 1.12). The opacity associated with LLPS is due to variations in refractive index between the two phases (Malfois, Bonneté et al. 1996). At equilibrium, the chemical potential (consisting of an entropic component and a binding energy component) of the dense and dilute phases are equal (Vekilov 2010; Wang, Latypov et al. 2014). The phenomenon of LLPS occurs as a result of short-ranged (less than a quarter of the protein diameter) attractive interactions. The range of hydration, hydrophobic and van der Waals forces is ~10 Å (Velev, Kaler et al. 1998; Leckband and Israelachvili 2001) whereas salt bridges and electrostatic forces have a shorter range (Muschol and Rosenberger 1995; Velev, Kaler et al. 1998; McPherson and Cudney 2006). LLPS is metastable with respect to crystallization and can only be observed in protein solutions where the kinetics of nucleation required for aggregation or crystallization to proceed are sufficiently slow and the range of the

interaction is much smaller than the size of the protein (Vekilov and Chernov 2003). The metastability of the dense protein phase following LLPS in protein solutions arises due to the restriction of the spatial position of each protein in the condensed phase as result of its nearest neighbouring proteins. A protein in the condensed phase has fewer neighbours when compared to the solid phase and so it is not caged by surrounding proteins to the same extent as in the solid phase. This gives rise to a dense protein phase with a slightly lower entropy when compared to the solid phase which, in turn, results in a dense protein phase that is metastable with respect to crystallization (Wang, Latypov et al. 2014).

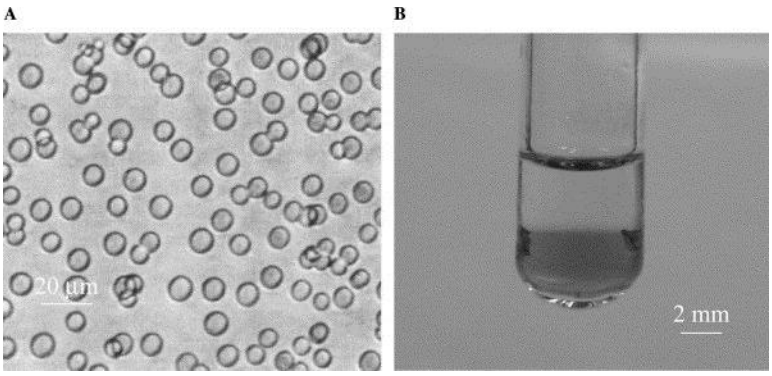

Figure 1.12: *Liquid-liquid phase separation of bovine γE crystallin showing dense protein droplets surrounded by the dilute protein (A) and complete liquid-liquid phase separation of thaumatin (B). Taken from Asherie 2004.*

Above certain volume fractions, native proteins in solution may condense to form a reversible gel (Vekilov 2010). Gelation is not a true equilibrium state and is instead a kinetically trapped state similar to the formation of colloidal glasses (Gunton, Shiryayev et al. 2007) Although the mechanisms governing gelation are poorly understood, several theories have attributed the mechanism to long-range attractive interactions (Noro, Kern et al. 1999; Sear 1999; Kulkarni, Dixit et al. 2003) or to formation of weak networks of protein molecules with limited lifetimes (Pan, Filobelo et al. 2009).

1.7.3 Factors affecting the liquid-liquid phase separation temperature

Temperature-induced LLPS of proteins was first measured for a solution of lysozyme in salt water (Ishimoto and Tanaka 1977). Protein solution conditions influence the

temperature at which liquid-liquid phase separation (T_{ph}) occurs. The effect of varying salt concentration, salt type and pH on the T_{ph} of lysozyme has been previously determined (Taratuta, Holschbach et al. 1990; Muschol and Rosenberger 1997). The effect of adding another protein on T_{ph} has also been determined for different protein mixtures (Liu, Lomakin et al. 1995; Asherie, Pande et al. 1998; Thurston 2006; Wang, Lomakin et al. 2010; Banerjee, Pande et al. 2011; Wang, Lomakin et al. 2011; Quinn, Gnan et al. 2015). The impact point mutations on a protein surface have on T_{ph} when compared to the native protein has also been determined for a number of mutant proteins (Pande, Annunziata et al. 2005; McManus, Lomakin et al. 2007).

1.7.4 Protein phase diagrams

Protein phase diagrams describe the physical state of a protein in two or three dimensions for a given set of parameters such as temperature, protein concentration (or volume fraction), pH or salt concentration. The phase diagram for a protein is determined by the range of interaction (Asherie, Lomakin et al. 1996; Rosenbaum, Zamora et al. 1996). A temperature-concentration (T,c) phase diagram (taken from Vekilov 2010) for lysozyme in pH 4.5 0.5 M sodium acetate with 4% salt is shown in figure 1.13. This phase diagram shows several regions of coexistence and stability, metastability and instability. A solubility line (or liquidus line) separates regions of monomer and monomer in equilibrium with a solid protein phase. Two solubility lines are shown in figure 1.13 because lysozyme can form two distinct crystal types: tetragonal and rhombohedral, each with its own solubility line (Muschol and Rosenberger 1997; Vekilov 2010). Of the protein phase transitions, LLPS is particularly important in determining colloidal stability (Wang, Latypov et al. 2014). There are two phase lines associated with LLPS; namely the liquid-liquid coexistence line (also known as the binodal) and the spinodal (figure 1.13). LLPS occurs between the binodal and spinodal by nucleation of protein droplets in the dense phase that grow until they reach equilibrium whereas spinodal decomposition occurs instantaneously below the spinodal without a nucleation barrier (Cahn and Hilliard 1958; Thomson, Schurtenberger et al. 1987; Dumetz, Chockla et al. 2008; Vekilov 2010). The critical point of a protein solution is the point at which the binodal and spinodal converge (Vekilov 2010). Normally, the critical point

represents the highest temperature that liquid coexistence is possible. However, for haemoglobin S (HbS), the critical point is the lowest temperature of liquid coexistence (Serrano, Galkin et al. 2001). The region where LLPS occurs for lysozyme (and indeed other proteins) is below the solubility line which is in contrast to the phase diagram for small molecules (Vekilov 2010). The position and the shape of the binodal is directly related to the range of attraction which must be significantly reduced i.e. much smaller than the size of the protein for LLPS to be metastable with respect to crystallization (Hagen and Frenkel 1994; Asherie, Lomakin et al. 1996; Lomakin, Asherie et al. 1996; Malfois, Bonneté et al. 1996). This in turn makes it directly related to the colloidal stability of a protein solution.

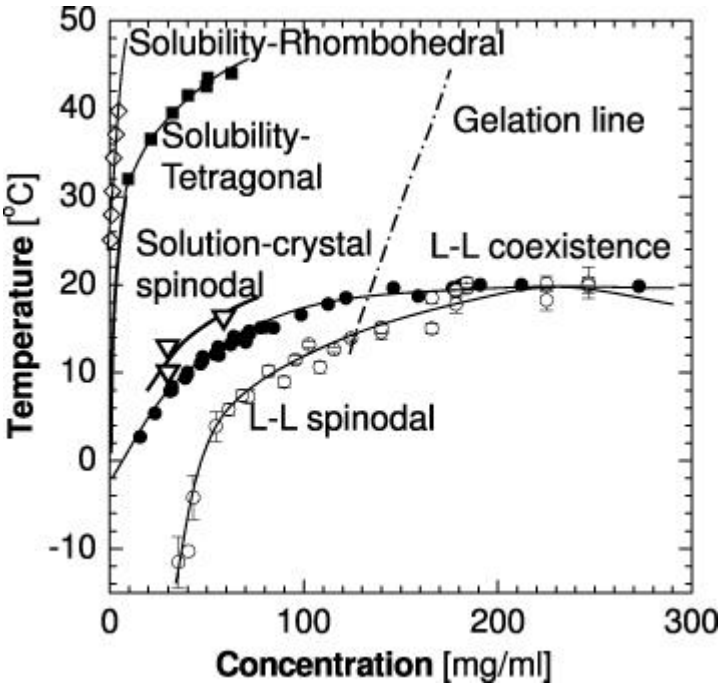

Figure 1.13: *Temperature, concentration (T,c) phase diagram for lysozyme in 0.5 M sodium acetate buffer pH4.5 with 4% salt showing two solubility lines for both types of crystal, liquid-liquid coexistence and spinodal curves, gelation line and the solution-crystal spinodal line. Taken from Vekilov 2010.*

1.8 Protein modifications

1.8.1 Post translational modifications

Covalent post translational modifications (PTMs) are chemical modifications of proteins that occur either *in vivo* or *in vitro* after their biosynthesis and serve to change the protein's structure or function by means of covalent attachment of a

functional group or protein, proteolytic (or, less commonly, autocatalytic) cleavage or degradation (Walsh, Garneau-Tsodikova et al. 2005; Baslé, Joubert et al. 2010). Common naturally occurring PTMs include acetylation, acylation, ADP-ribosylation, amidation, γ-carboxylation, β-hydroxylation, disulphide bond formation, methylation, ubiquitylation, glycosylation, phosphorylation, proteolytic processing and sulphation (Whitford 2005; Walsh and Jefferis 2006). Out of the 22 proteinogenic amino acids, 15 are subject to diversification *via* PTM (there are no known modifications of Leu, Ile, Val, Ala, Phe, Sec and Pyl) (Walsh, Garneau-Tsodikova et al. 2005). In almost all cases, naturally occurring PTMs are crucial for the function and structure of proteins (Whitford 2005).

1.8.2 Artificial chemical modification

Given the diversity of structure and function of proteins as a result of naturally occurring PTMs, the ability to chemical modify proteins by artificial means is highly attractive (Baslé, Joubert et al. 2010). The artificial chemical modification of a protein surface (which will be referred to here simply as 'chemical modification') was first reported in 1904 and has, since then, been a mainstay in the study of proteins (Stephanopoulos and Francis 2011). Of the 22 proteinogenic amino acids, lysine (Lys) and cysteine (Cys) (figure 1.14) are the most commonly used sites for chemical modification (Sletten and Bertozzi 2009; Polakis 2016) owing to their inherent nucleophilic character (Baslé, Joubert et al. 2010; Spicer and Davis 2014; Boutureira and Bernardes 2015). Other amino acids can be chemically modified but can be more synthetically challenging to achieve (Sletten and Bertozzi 2009).

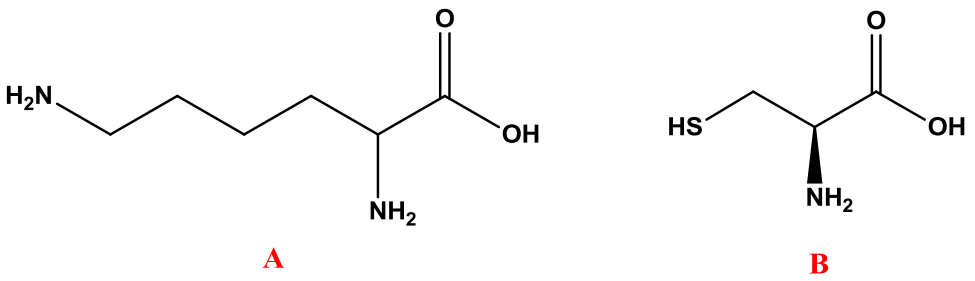

Figure 1.14: *Structures of lysine (A) and cysteine (B).*

Lysine is one of the most abundant amino acids in proteins and there are many established protocols to preferentially chemically modify the amine of its side chain which is typically the most nucleophilic amine on a protein surface (Sletten and

Bertozzi 2009; Spicer and Davis 2014; Boutureira and Bernardes 2015; Turecek, Bossard et al. 2016). Preferential modification of Lys requires the use of 'harder' nucleophiles such as activated esters, sulfonyl chlorides and isothiocyanates or modification can be achieved through reductive alkylation using aldehydes in the presence of cyanoborohydride (Spicer and Davis 2014). The abundance of Lys can lead to heterogeneous mixtures of modified protein which may be useful in some applications where multiple conjugates are required but undesirable in others such as the development of some biopharmaceutics where a more homogeneous mixture is preferred (Spicer and Davis 2014). Cys is regarded as the most important naturally occurring amino acid for chemical modification because, in addition to its highly nucleophilic nature (it is much more nucleophilic than Lys), it has a low abundance (<2%) which allows for chemical modification with a greater degree of specificity and, therefore, less heterogeneity (Chalker, Bernardes et al. 2009; Sletten and Bertozzi 2009; Spicer and Davis 2014). There are a number of electrophiles available for specific labelling of Cys residues such as α-halocarbonyls, iodoacetamides and maleimides, with maleimides prevailing as the most commercialised and widely used derivative for Cys modification as a result of their ease of use and straightforward synthesis (Spicer and Davis 2014). Indeed, a selenocysteine residue, which is more nucleophilic than Cys, can be inserted into a protein for highly specific modification also using maleimide derivatives (Hofer, Thomas et al. 2008).

Tyrosine (Tyr) and tryptophan (Trp) (figure 1.15) are rarely found on the surface of a protein (due to their hydrophobic nature) which has made them the subject of focus as sites for chemical modification since single site modification is advantageous (Joshi, Whitaker et al. 2004; Baslé, Joubert et al. 2010). The modification of Tyr and Trp is generally performed using water-soluble transition metal catalysis (Antos and Francis 2006). Trp has the lowest natural abundance of all the canonical amino acids which can allow for Trp modification specificity provided the protein being modified has no surface-bound Tyr residues.

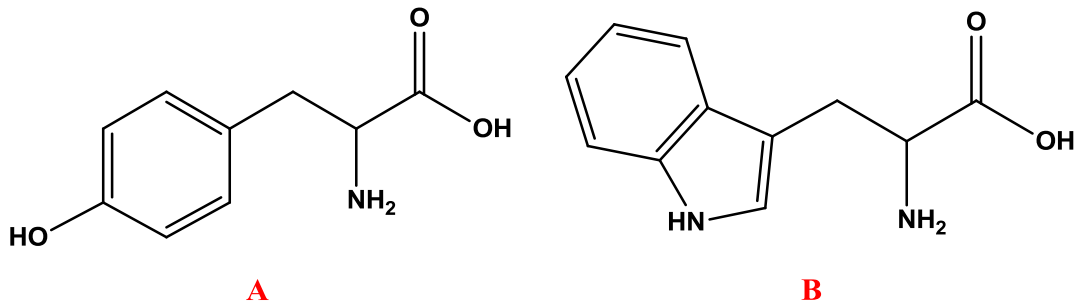

Figure 1.15: *Structures of tyrosine (A) and tryptophan (B)*

Chemical modification of histidine (His), arginine (Arg), glutamic acid (Glu) and aspartic acid (Asp) is also possible as each has a nucleophilic residue that can be used as a site for modification but their use is much rarer than modification of Lys, Cys, Tyr and Trp (Boutureira and Bernardes 2015). His (figure 1.16A) is a target for nucleophilic attack and is highly selective towards vinyl sulphone which allows for site specific labelling with molecules containing a vinyl sulphone group as an acetylating agent (del Castillo, Marales-Sanfrutos et al. 2014). While Arg (figure 1.16B) has a similar abundance to Lys, it has a lower tendency to be found on the surface of the protein which makes it more favourable for site specific modification (Gauthier and Klok 2008). Specific modification of Arg involves the formation of a pyrimidine ring using two out of the three nitrogen atoms found in its side chain (Oya, Hattori et al. 1999). Although highly selective, modification of Arg requires a 14 day reaction time, which may make it an unsuitable choice of site for some studies that require specificity as well as a short reaction time. Modification of the carboxylate groups in Glu (figure 1.16C) and Asp (1.16D) involves amidation using ethylenediamine (Díaz-Rodríguez and Davis 2011). However, ethylenediamine indiscriminately labels both Glu and Asp which make these sites unfavourable as sites for modification.

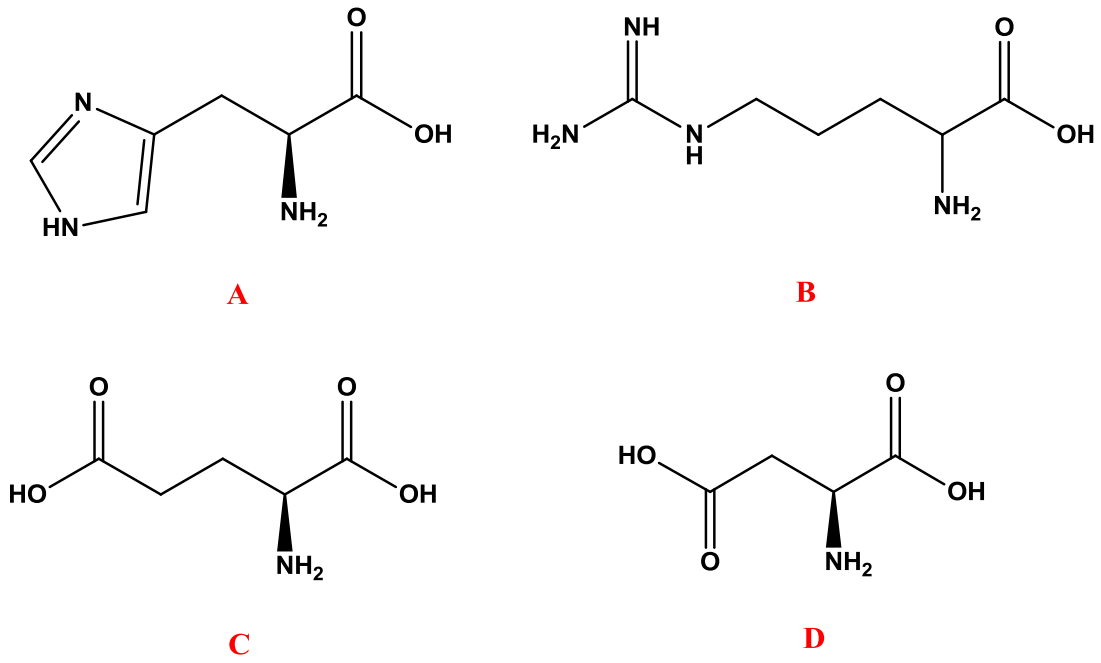

Figure 1.16: *Structures of histidine (A), arginine (B), glutamic acid (C) and aspartic acid (D)*

Antibody drug conjugates (ADCs) are a promising emergent biotherapeutic for targeted treatment of cancer that combine the specificity of an antibody with the potency of a small molecule (Agarwal and Bertozzi 2015; Akkapeddi, Azizi et al. 2016; Diamantis and Banerji 2016; Polakis 2016). ADCs have three parts: a humanised antibody for an antigen uniquely expressed or overexpressed in cancer cells, a potent cytotoxic drug that can be tethered to the antibody and a stable linker between the two that prevents premature release of the cytotoxic drug and can also serve to increase the solubility of the ADC (Akkapeddi, Azizi et al. 2016) (figure 1.17).

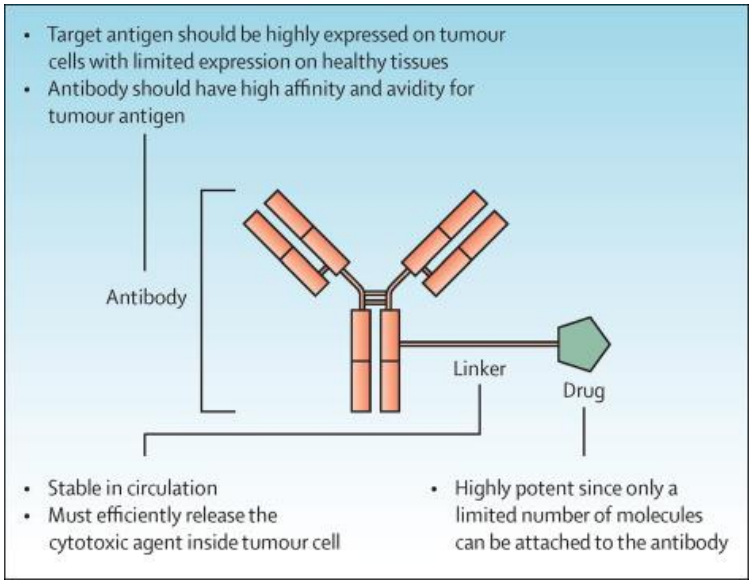

Figure 1.17: *ADC structure and component requirements. Taken from Polakis 2016.*

There are currently two ADCs approved by the US Food and Drug Administration (FDA) available for cancer treatment: Kadcyla (approved in 2013) and Adcetris (approved in 2011) and over 40 currently undergoing clinical trials (Polakis 2016). Ado-trastuzumab emtansine (brand name Kadcyla) is manufactured by Roche/Genentech and used for the treatment of HER2 positive metastatic breast cancer and uses succinimidyl-4-[N-maleimidomethyl]- cyclohexane-1-carboxylate (SMCC) for Lys modification. Given the abundance of Lys residues, modification at this site gives a heterogeneous drug product with an average drug-to-antibody ratio (DAR) of between 3 and 3.5 (Akkapeddi, Azizi et al. 2016). Brentuximab vedotin (brand name Adcetris) is manufactured by Seattle Genetics/Millennium Pharmaceuticals and is used for the treatment of relapsed Hodgkin Lymphoma. Modification of Cys is achieved using *N*-hydroxysuccinimide (NHS) ester and produces a more heterogeneous drug product given the low abundance of Cys residues available for modification (Agarwal and Bertozzi 2015).

1.8.3 Mutagenesis

Mutagenesis also alters the surface chemistry of proteins and, as opposed to proteins that are modified after translation, mutant proteins are coded for by mutated codons, insertions and deletions in the DNA sequence (Carter 1986). Site-directed mutagenesis, was first successfully used 35 years ago to specifically alter the amino acid sequences of 2 proteins, namely tyrosyl–transfer RNA synthetase (TyrRS) and

beta-lactamase (Dalbadie-McFarland, Cohen et al. 1982; Irving, Harwood et al. 1982; Winter, Fersht et al. 1982; Brannigan and Wilkinson 2002). Since then, it has become an invaluable tool in molecular biology that provides insight into the structure and function of proteins (Hsieh and Vaisvila 2013). For the purposes of site specificity during chemical modification, proteins can be mutated with either canonical (e.g. Cys) or non-canonical amino acids (Gauthier and Klok 2008; Stephanopoulos and Francis 2011). Naturally occurring mutants, such as those associated with sickle cell anaemia and congenital cataract, are discussed in more detail in sections 1.4.2 and 1.9.2 respectively.

1.9 The eye lens

1.9.1 Background

The eye lens (figure 1.18) is an avascular tissue whose function is to focus light on the retina in order to see (Sharma and Santhoshkumar 2009). Crystallins, which were discovered nearly 125 years ago and are divided into two separate families (α and βγ), are the main structural proteins of the vertebrate lens and account for the vast majority (about 90%) of the water soluble proteins found in the eye lens (Andley 2007; Graw 2009; Sharma and Santhoshkumar 2009).

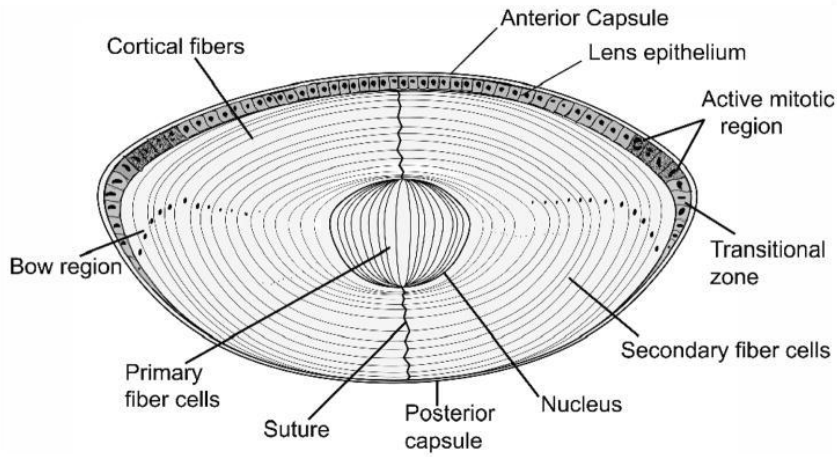

Figure 1.18: *A representation of the cross-section of the lens indication the location of the nucleus. Taken from Sharma and Santhoshkumar 2009.*

These proteins remain soluble at high concentrations in excess of 450 mg ml^{-1} (Hejtmancik 2008). While both crystallin families are structural proteins, α crystallins (consisting of αA and αB which are large polydisperse multimeric

subunits) account for 40% of the soluble lens proteins in humans and act as molecular chaperones to maintain lens transparency (Horwitz, Bova et al. 1999; Bloemendal, de Jong et al. 2004). Oligomeric β crystallins have four key Greek motifs organised into two domains and are divided into acidic β crystallins (βA1, βA2, βA3 and βA4) and basic β crystallins (βB1, βB2 and βB3) (Bloemendal, de Jong et al. 2004). Monomeric γ crystallins are also comprised of four Greek key motifs organised into two domains. The mammalian genome codes for seven γ crystallins (γA, γB, γC, γD, γE, γF and γS). In humans, γA, γB, γC, γD and γS are all expressed but only γC and γD are expressed in appreciable quantities, with γD having the most predominant expression (Andley 2007; Wistow 2012). The γ crystallins interact *via* short-ranged attractive interactions to contribute to lens transparency due to this short-ranged order (Benedek 1971; Delaye and Tardieu 1983). Of all the proteins that undergo liquid-liquid phase separation, the γ crystallins are the most widely studied (Malfois, Bonneté et al. 1996; Muschol and Rosenberger 1997).

Human γD-crystallin (HGD), formerly known as γ_{IIIb} (Liu, Lomakin et al. 1995), is an extremely stable (soluble at concentrations in excess of 400 mg ml^{-1}), monomeric, globular protein consisting of 174 amino acids with a critical point at a volume fraction (ϕ) of approximately 0.21 and corresponding temperature of ~277 K in 0.1 M sodium phosphate at pH 7.0 (Liu, Asherie et al. 1996; Lomakin, Asherie et al. 1996; Wang, Lomakin et al. 2010). As a result of these short-ranged attractive interactions, it undergoes LLPS. It can also form crystals. Point mutations in the amino acid sequence of HGD have been linked to congenital cataracts (Basak, Bateman et al. 2003; Pande, Annunziata et al. 2005; Pande, Gillot et al. 2009; Banerjee, Pande et al. 2011; Ji, Jung et al. 2013). Modifications to the surface of HGD either by chemical modification or mutagenesis have been used to investigate the effect anisotropy has on determining the phase behaviour of proteins in solution (McManus, Lomakin et al. 2007; James, Quinn et al. 2015; Quinn, Gnan et al. 2015).

1.9.2 Human γD-crystallin mutants

A number of congenital cataract phenotypes are associated with several point mutations in the amino acid sequence of HGD and are briefly discussed below. Although mutations are known to occur in both the α and βγ crystallin families

(Wistow 2012), those occurring in the γ crystallin family, specifically mutations associated with human γD-crystallin (HGD) are the most relevant here.

Arg-58 to His (R58H) and Arg-36 to Ser (R36S) mutations lead to a loss of transparency due to a decrease in the solubility of the mutated protein relative to its native form and its increased propensity to form crystals (Pande, Pande et al. 2001). The formation of these crystals in the eye lens leads to increased light scattering due to variations in refractive index.

Pro-23 to Thr (P23T) mutation also presents with a decrease in solubility relative to the native protein but form non-covalent self-associated aggregates with a retrograde temperature dependence that lead to blindness (Evans, Wyatt et al. 2004). Other mutations at and around this site (Pro-23 to Ser (P23S), Pro-23 to Val (P23V), P23TInsP24 and P23TAsn-24 to Lys (P23TN24K)) have been explored in a bid to better understand the mechanisms of aggregation associated with this mutation (Pande, Annunziata et al. 2005). P23S and P23V both led to the formation of aggregates that were more soluble than the naturally occurring P23T mutant but less soluble than the native form. Like the naturally occurring P23T mutant, the aggregates formed due to the P23V and P23S mutations also maintained an inverted temperature dependence. Interestingly, the P23TInsP24 and P23TN24K mutants restored protein solubility to native values (Pande, Annunziata et al. 2005).

Like P23T, Arg-14 to Cys (R14C) leads to the formation of aggregates resulting in blindness. Unlike the P23T mutation, the R14C mutation results in disulphide cross-linked oligomers as a result of presence of another Cys residue (Pande, Pande et al. 2000). In the native protein, there is only one surface exposed Cys residue which can form only one disulphide bridge with another HGD protein. In the mutant protein, two Cys residues are available to form disulphide bridges and the mutant protein can form higher weight oligomers.

A Trp-42 to Arg (W42R) mutation has been shown to have a reduced solubility with respect to the native form (Wang, Yu et al. 2011). The basis of the decreased solubility arises from a partial unfolding as a result of this mutation which mimics the photo-damaged HGD involved in age related cataract (Ji, Jung et al. 2013).

Like the point mutations mentioned previously, the Glu-107 to Ala (E107A) mutation is also responsible for a decrease in lens protein solubility but, unlike the above mutations, cataract is not caused due to an altered γ-γ interaction alone (Messina-Baas, Gonzalez-Huerta et al. 2006; Banerjee, Pande et al. 2011). Instead, light scattering arises due an increased net attraction between the negatively charged α-crystallins and the mutant γD-crystallin as a result of an increase in positive charge due to the replacement of a negative charged Glu with an uncharged Ala.

1.10 Thesis motivation

The motivation of the work presented in this thesis was to investigate how physical and chemical properties of a protein and its environment affect its phase behaviour. Phase behaviour is intrinsically linked with the solution stability of the protein. Understanding and optimising protein stability is of particular importance in the food and biopharmaceutical sectors.

Proteins are inherently anisotropic macromolecules as a result of their unique surface chemistries. The sum of the attractive and repulsive interactions from the surface-bound amino acid side chains give rise to a net interaction potential which affects a protein's solution behaviour and can be altered by changing the surface properties of the protein. The environment surrounding a protein in solution also contributes to its phase behaviour.

The first results chapter examines the effects of chemically modifying the surface of a protein with a (hydrophobic) small molecule fluorescent label. The second chapter continues in the vein of chemical modification of a protein's surface but explores chemical modification using PEGylated biotin (which is hydrophilic). This chapter also investigates the phase behaviour of a mixture of a native and mutant protein that is known to contain a hydrophobic patch on its surface. The third chapter explores the impact the environment of the protein has on its phase behaviour. This is investigated in two ways; firstly, the impact different surfaces that are in contact with the protein solution have on protein phase behaviour are explored. Secondly, the impact of intracellular milieu on protein phase behaviour is examined.

This thesis aims to show that protein phase behaviour is not only determined by the properties of the protein itself but is also dependent on its environment.

Chapter 2:

Materials and Methods

2.1 Preparation of buffers and reagents

2.1.1 Buffers

Buffers were prepared using either ultrapure (Type 1, Milli-Q) or pure (Type 2, Elix) water and analytical or HPLC grade solvents and reagents. When required, buffer pH was adjusted using either a concentrated hydrochloric acid solution (HCl, M_w = 36.46 g mol^{-1}) (Fisher Scientific, UK) or a concentrated sodium hydroxide solution (NaOH, M_w = 40.00 g mol^{-1}) (Fisher Scientific, UK), unless otherwise stated. To inhibit microbial growth, 0.02% (w/v) sodium azide (NaN_3, M_w = 65.01 g mol^{-1}) (Fisher Scientific, UK) was added to each buffer, unless otherwise stated. All buffers were filtered using either vacuum filtration through a 0.45 μm nylon membrane filter (Millipore, Ireland) or a 0.45 μm Millex-HV syringe driven filter (Millipore, Ireland) prior to use. Buffers were degassed under vacuum with stirring where necessary. Where filter sterilization was required, solutions were filtered through a 0.22 μm Millex-GV syringe driven filter (Millipore, Ireland).

2.1.1.1 Sodium acetate

Sodium acetate buffer (0.275 M) was prepared either with or without salt at pH 4.5 and pH 4.8 respectively. Sodium acetate buffer without salt at pH 4.5 was prepared using 16.50 g glacial acetic acid (CH_3COOH, M_w = 60.05 g mol^{-1}) (Fisher Scientific, UK) and 4.48 g NaOH (Fisher Scientific, UK) per litre of pure water. Sodium acetate buffer at pH 4.8 with salt was prepared using 19.02 g of sodium chloride (NaCl, M_w = 58.44 g mol^{-1}) (Fisher Scientific, UK), 16.50 g CH_3COOH and 6.70 g NaOH per litre of water. Caution was taken not to mix CH_3COOH and NaOH directly.

$$CH_3COOH_{(l)} + NaOH_{(s)} \rightarrow CH_3COONa_{(aq)} + H_2O_{(l)} \qquad \textbf{2.1}$$

$$CH_3COOH_{(l)} + NaOH_{(s)} + NaCl_{(s)} \rightarrow CH_3COONa_{(aq)} + H_2O_{(l)} + NaCl_{(aq)} \qquad \textbf{2.2}$$

2.1.1.2 Sodium phosphate

0.1 M sodium phosphate buffer was prepared using 14.20 g sodium phosphate dibasic heptahydrate ($Na_2HPO_4.H_2O$, M_w = 141.96 g mol^{-1}) (VWR, Belgium) and 15.60 g sodium phosphate monobasic dihydrate ($NaH_2PO_4.2H_2O$, M_w = 156.01 g

mol^{-1}) (Fisher Scientific, UK) per litre of water. The pH was adjusted to 7.0 or 8.0 as required.

1 M sodium phosphate buffer at pH 6.0 was prepared using 156.01 g $NaH_2PO_4.2H_2O$ and 149.2 g $Na_2HPO_4.7H_2O$ per litre of water.

2.1.1.3 Sodium borate

50 mM of sodium borate pH 8.5 was prepared using 12.37 g boric acid (H_3BO_3, M_w = 61.83 g mol^{-1}) (Sigma Aldrich, USA) and 4.00 g NaOH per litre of water (2.3).

$$4H_3BO_{3(s)} + 2NaOH_{(s)} \rightarrow Na_2B_4O_{7(aq)} + 7H_2O_{(l)} \qquad \textbf{2.3}$$

2.1.2 Buffers for fluorescent labelling

2.1.2.1 Reducing buffer

1 ml of reducing buffer was prepared by mixing 100 µl of 1 M sodium phosphate buffer at pH 6.0 and 5 µl of 0.5 M EDTA with 895 µl ultrapure water.

2.1.2.2 Conjugation buffer

Conjugation buffer was prepared using 20 µl of 0.5 M EDTA per 10 ml of 0.1 M sodium phosphate buffer at pH 7.0.

2.1.3 Buffers and reagents for *E. coli* cell culture

Where indicated, reagents for *Escherichia coli* (*E. coli)* cell culture were autoclaved at 121 °C and 0.212 MPa pressure for 20 minutes in a SX-500E TOMY autoclave (Seiko, Japan).

2.1.3.1 Ampicillin stock solution

A 50 µg ml^{-1} solution of ampicillin sodium salt ($C_{16}H_{18}N_3NaO_4S$, M_w = 371.39 g mol^{-1}) (Fisher Scientific, USA) was prepared in ultrapure water, filter sterilised and stored at -20 °C.

2.1.3.2 LB preparation

Lysogeny broth (LB) was prepared using 25 g of LB Miller powder (Fisher Scientific, USA) per litre of ultrapure water and autoclaved. Ampicillin was added to the cooled broth after autoclaving to give a final concentration of 100 µg ml^{-1}.

2.1.3.3 LB agar plates

LB Agar plates were prepared using 37 g of LB Agar granules (Fisher Scientific, UK) per litre of ultrapure water which was then autoclaved on a liquid cycle. After autoclaving, agar was cooled to below 55 °C before adding ampicillin to a final concentration of 100 µg ml^{-1}. Agar was poured into sterile 10 cm petri dishes and allowed to set before use. Plates were sealed with Parafilm (Bemis, USA) and stored at 4 °C for a maximum of 1 week.

2.1.3.4 NZY^+ broth

NZY^+ broth was prepared using 10 g NZ amine (Sigma Aldrich, USA), 5 g yeast extract (Oxoid, England) and 5 g NaCl per litre of ultrapure water. The pH of the broth was adjusted to 7.5 and the broth was autoclaved. After autoclaving, the broth was supplemented with 12.5 ml filter sterilised 1 M magnesium chloride hexahydrate (Fisher Scientific, USA) ($MgCl_2.6H_2O$, M_w = 203.30 g mol^{-1}), 12.5 ml filter sterilised 1 M magnesium sulphate heptahydrate (Fisher Scientific, USA) ($MgSO_4.7H_2O$, M_w = 246.45 g mol^{-1}) and 10 ml filter sterilised 2 M glucose ($C_6H_{12}O_6$, M_w = 180.16 g mol^{-1}).

2.1.3.5 Lysis buffer

Lysis buffer was prepared immediately before use by dissolving 0.0788 g (50 mM) tris(hydroxymethyl)aminomethane hydrochloride (Tris HCl, $C_4H_{11}NO_3$, M_w = 121.14 g mol^{-1}) (Merck, Germany), 0.0146 g (25 mM) NaCl and 0.0074 g (2 mM) ethylenediaminetetraacetic acid (EDTA, $C_{10}H_{16}N_2O_8$, M_w = 292.23 g mol^{-1}) (Fisher Scientific, USA) in 10 ml ultrapure water.

2.1.4 Reagents for mammalian cell culture

The reagents used for mammalian cell culture were prepared using sterile reagents and solvents in a Class II biological safety cabinet using aseptic technique. Where indicated, reagents for mammalian cell culture were autoclaved at 121 °C and 0.212

MPa pressure for 20 minutes in a SX-500E TOMY autoclave and filter sterilised using 0.22 μm Millex-GV syringe driven filter.

2.1.4.1 Complete growth medium

Complete growth medium (CGM) was prepared using HyClone Dulbecco's modified Eagle's medium (DMEM) (4 mM L-glutamine and 4.5 g L^{-1} glucose either with or without Phenol Red) (GE Healthcare Life Sciences, USA) supplemented with 10% (v/v) bovine calf serum (BCS) (Sigma Aldrich, USA), 1 mM sodium pyruvate (Sigma Aldrich, USA), 1.5 mg ml^{-1} sodium bicarbonate (autoclaved at 121 °C for 20 minutes and filter sterilised) (Sigma Aldrich, Germany) and 2.5 ml penicillin (10,000 units ml^{-1})-streptomycin (10 mg ml^{-1}) (Sigma Aldrich, USA). CGM was stored at 4 °C following preparation.

2.1.4.2 Phosphate-buffered saline

10 ml aliquots of 1X phosphate-buffered saline (PBS) were prepared in two ways; 1 PBS tablet (Fisher Scientific, UK) was dissolved in 100 ml ultrapure water and autoclaved. Alternatively, 10X PBS (Sigma Aldrich, USA) was diluted to 1X PBS using ultrapure water and filtered. Aliquots of 1X PBS were stored at -20 °C following preparation.

2.1.4.3 Trypsin-EDTA solution

3 ml of 1X trypsin-EDTA was prepared by adding 300 μl 10X trypsin-EDTA (5.00 g/l porcine trypsin, 2 g/l EDTA-4Na, 0.9% NaCl) (Sigma Aldrich, USA) to 2.7 ml 1X PBS.

2.1.4.4 Poly-D-lysine solution

5 ml aliquots of 100 μg ml^{-1} solution of poly-D-lysine (PDL) were prepared using 5 mg (70–150 kDa, lyophilised, γ-irradiated) poly-D-lysine hydrobromide (Sigma Aldrich, USA) per 50 ml autoclaved ultrapure water. After aliquoting, the PDL solution was stored at 4 °C.

2.1.4.5 Cryopreservation medium

Cryopreservation medium was prepared before use using 95% complete growth medium (CGM) and 5% dimethyl sulphoxide (DMSO) (Fisher Scientific, USA).

2.2 Bacterial cell culture

2.2.1 HGD transformation in *E. coli*

Human γD-crystallin (HGD) plasmid containing an ampicillin resistance gene and an isopropyl-β-D-thiogalactopyranoside (IPTG) promoter was transformed in BL21-Gold(DE3) competent *E. coli* cells (Agilent, USA) as per manufacturer's guidelines. Competent cells were thawed on ice and gently mixed. 100 µl of the cells were aliquoted into a pre-chilled 15 ml polypropylene centrifuge tube. 50 ng of HGD plasmid was added to the tube containing the competent cells and the contents swirled gently. The transformation reaction was incubated on ice for 30 minutes and then heat-pulsed at 42 °C for 20 seconds and incubated for a further 2 minutes on ice. 0.9 ml pre-heated NZY^+ broth was added to the transformation reaction and the mixture was incubated at 37 °C with shaking at 250 RPM for 1 hour in an Innova 42 incubator shaker (New Brunswick Scientific, USA). The transformation reaction was plated on a LB agar plate containing 100 µg ml^{-1} ampicillin. The plate was incubated overnight at 37 °C. Stocks of transformed cells were prepared by inoculating 100 ml LB broth containing 100 µg ml^{-1} ampicillin with 1 colony picked from the streaked plate using a 1 µl inoculation loop and incubating overnight at 37 °C with shaking at 225 RPM. Cells were transferred to Nalgene cryogenic vials (ThermoScientific, USA) and stored at -80 °C (New Brunswick Scientific, USA).

2.2.2 Bacterial cell culture for recombinant protein production

LB agar plates containing 100 µg ml^{-1} ampicillin were streaked with transformed *E. coli* from frozen stocks using 1 µl clear sterile disposable inoculation loops. Streaked plates were incubated at 37 °C overnight. A single colony was taken from the streaked plate using a clear disposable 1 µl inoculation loop and used to inoculate 100 ml of LB broth containing 100 µg ml^{-1} ampicillin. The 100 ml culture was grown overnight in an incubator shaker at 37 °C with shaking at 225 RPM. The 100 ml culture was divided between 3 large cultures flasks containing 1.2 litres of sterilised LB broth containing 100 µg ml^{-1} of ampicillin. The large cultures were grown at 37 °C with shaking at 225 RPM. When the optical density at 600 nm (OD_{600}) was between 0.8 and 1.0, the cultures were induced using 1 ml of 1M IPTG ($C_9H_{18}O_5S$, M_w = 238.3 g mol^{-1}) (Fisher Scientific, Italy) per culture. The culture was incubated for a further 4 to 5 hours. The cultures were then centrifuged at 6,000 RPM for 8

minutes in 250 ml fluorinated Nalgene HDPE bottles using a carbon fiber rotor (ThermoScientific, USA) in a ThermoScientific Heraeus Multifuge 3SR+ centrifuge to pellet the cells. The supernatant was discarded after sterilisation using a 1 % Virkon solution (1 Rely+On Virkon tablet (Antec International Limited) per litre of water). The pelleted cells were stored at -80 °C.

2.2.3 Bacterial cell lysis

The pelleted *E. coli* cells were thawed in a water bath at 30 °C. 155 mg dithiothreitol (DTT, M_w = 154.35 g mol^{-1}) (Fisher Scientific, Japan) and 1 Roche Protease Inhibitor Cocktail tablet (Roche Diagnostics, USA) were added to freshly prepared lysis buffer (10 mls) (section 2.1.3.5) and ~0.83 ml of the lysis buffer was added to each bottle containing the cell pellet. The cells and lysis buffer were mixed and vortexed to re-suspend the cell pellet. Cells were combined, vortexed briefly and incubated at room temperature for 2 hours. 160 µl of a 50 mg ml^{-1} lysozyme (Calbiochem, Canada) solution were added to the cell suspension. The cell suspension was vortexed and incubated for a further 30 minutes at room temperature. The cell suspension was then subjected to four freeze-thaw cycles using liquid nitrogen and a water bath at 30 °C. 2 ml of freshly prepared solution of 1 mg ml^{-1} deoxyribonuclease (DNase I, Bovine Pancreas, Calbiochem, South Africa) and 1 ml of 1 M $MgSO_4$ were mixed together and then added to the cell suspension following freeze-thaw. The cell suspension was incubated at room temperature for 30 minutes before adjusting the pH to ~4.5 using glacial acetic acid. The cell suspension was centrifuged at 8,500 RPM in a ThermoScientific Heraeus 3R+ Multifuge centrifuge to pellet the cell debris. The supernatant containing the protein was decanted and the cell pellet was stored at -80 °C for further extraction if required. The supernatant was filtered through a 0.45 µm Millex-HV syringe driven filter before purification.

2.2.4 Site directed mutagenesis

Site directed mutagenesis is a molecular biology method whereby the deoxyribonucleic acid (DNA) sequence at a specific site (or multiple sites) is intentionally changed *via* an amino acid insertion, deletion or substitution. The forward and reverse oligonucleotide primers were previously designed for the Arg-36 to Ser (R36S) single mutation of HGD (Pande, Pande et al. 2001) and were synthesised by Life Technologies (Ireland). These primers were used to mutate a

previously mutated mammalian plasmid that coded for a single mutant of HGD containing a Pro-23 to Thr mutation (P23T).

Forward Primer: 5′-GCA ACT CGG CGA GCG TGG ACA GCG GC-3′

Reverse Primer: 5′-GCC GCT GTC CAC GCT CGC CGA GTT GC-3′

Mutagenesis was performed using a QuikChange II site-directed mutagenesis kit (Stratagene, USA) as per manufacturer's instructions. Each polymerase chain reaction (PCR) was prepared in 0.5 ml PCR tubes (Eppendorf, Germany) and brought to a total volume of 50 µl with ultrapure water using the quantities in table 2.1:

Table 2.1: *PCR components and quantities for mutagenesis.*

Component	Quantity
10X reaction buffer	5.00 µl
dsDNA template	50 ng
Forward primer	125 ng
Reverse primer	125 ng
dNTP mix	1.00 µl
Pfu Ultra HF DNA polymerase	1.00 µl

The PCR tube was briefly vortexed and centrifuged. A ^{3}Prime thermal cycler (Bibby Scientific, UK) was used for PCR and the temperature was cycled using the parameters outlined in table 2.2.

Table 2.2: *PCR cycling parameters for site-directed mutagenesis of EmGFP-P23T plasmid for EmGFP-P23TR36S expression in HEK293T/17 cells.*

Step	Cycles	Temperature (°*C*)	Time (*minutes*)
1 (Preheat Lid)	1	105	2
2	1	95	1
3	16	95 65 68	0.5 1 9

The PCR tube was placed on ice for 2 minutes following temperature cycling. 1 µl *Dpn* I restriction enzyme was added to the amplification reaction to digest methylated DNA and mixed gently by pipetting. The reaction was centrifuged for 1 minute and incubated at 37 °C for 1 hour to digest non-mutated supercoiled dsDNA. The plasmid was then transformed in *E. coli* and amplified, purified and sequence.

2.2.5 P23TR36S transformation and amplification

Plasmid DNA was transformed into XL1-Blue supercompetent cells (Agilent, USA) as per manufacturer's instructions. Cells were thawed on ice and 50 µl thawed cells were added to a pre-chilled 15 ml polypropylene centrifuge tube. 1 µl of the *Dpn* I treated DNA (section 2.2.4) was added to the centrifuge tube. The tube was then heat pulsed at 42 °C for 45 seconds and then placed on ice for 2 minutes. 0.5 ml NZY^+ (section 2.1.3.4) pre-heated to 42 °C was added to the tube. The transformation reaction was incubated at 37 °C for 1 hour with shaking at 250 RPM. Following this, the transformation reaction was divided between 2 LB agar plates containing 100 µg ml^{-1} ampicillin and incubated overnight at 37 °C.

4 ml LB containing 100 µg ml^{-1} ampicillin was inoculated using a single colony from each of the streaked plates. The inoculated starter culture was incubated at 37 °C with shaking at 300 RPM for 8 hours. 25 ml LB was then inoculated with 50 µl of the starter culture and grown overnight at 37 °C with shaking at 300 RPM. Cells containing the mutated plasmid DNA were harvested by centrifugation at 6000 *g* at 4 °C for purification and sequencing.

2.2.6 Plasmid purification and sequencing

Plasmid DNA was purified from transformed cells using a QIAGEN Plasmid Purification Midi Kit (Qiagen GmbH, Germany) as per manufacturer's guidelines. The concentration of the DNA was determined spectroscopically using a Molecular Devices SpectraMax M2e spectrometer (Molecular Devices, USA) (equation 2.4):

$$Concentration\ (\mu g / ml) = A_{260} \times df \times 50 \mu g / ml \qquad \textbf{2.4}$$

where A_{260} is the absorbance at λ = 260 nm after baseline subtraction and *df* is the dilution factor. DNA purity was determined by calculating A_{260}/A_{280} and A_{260}/A_{230}

ratios (after baseline subtraction) where a ratio of ~1.8 for A_{260}/A_{280} and ~2.0 for A_{260}/A_{230} is considered pure.

Plasmid DNA was sequenced after purification with a T7 promoter primer using an automated capillary DNA sequencer (Medical Research Council Protein Phosphorylation and Ubiquitylation Unit, University of Dundee, Scotland).

2.3 Analytical techniques for protein purification and protein purity determination

2.3.1 Chromatography

2.3.1.1 Background

Chromatography is most commonly used as a means of purification and consists of a gas or liquid mobile phase in which an analyte is dispersed and a stationary phase with which an analyte interacts (Whitford 2005). The change in Gibbs free energy can be used to describe chromatographic processes (equation 2.5) (Hong, Koza et al. 2012).

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT\ln k \qquad \textbf{2.5}$$

where ΔG^0, ΔH^0 and ΔS^0 are the changes in free energy, enthalpy and entropy respectively, T is the absolute temperature, R is the gas constant and k is a partition coefficient. Changes in free energy are dominated by the enthalpy of adsorption for most chromatographic systems (Hong, Koza et al. 2012).

Size exclusion chromatography (SEC), also known as gel filtration chromatography, is a chromatographic technique whereby analytes in a mixture are separated based on the physical interplay between the hydrodynamic radii of each of the individual components in the mixture with the pore size of the stationary phase (Whitford 2005). Molecules that have a relatively large hydrodynamic radius elute from the column quicker than smaller molecules and elution time increases with decreasing molecule size due to the ability of the smaller molecules to penetrate the interstitial spaces within the column packing. SEC is athermal and adsorption is ideally non-existent and so ΔH^0 is zero and equation 2.5 now becomes equation 2.6 (Hong, Koza et al. 2012):

$$\frac{\Delta S^0}{R} = \ln K_D \qquad \textbf{2.6}$$

where K_D is the thermodynamic retention factor for SEC determined by the interstitial space that is accessible to the analyte, defined by equation 2.7:

$$K_D = \frac{V_R - V_0}{Vi} \qquad \textbf{2.7}$$

where V_R is the retention volume, V_0 is the interstitial volume and Vi is the intra-particle volume. K_D can range in value between 0 (where the analyte is fully excluded due to its size) to 1 where the interstitial space is fully accessible to the analyte. (Hong, Koza et al. 2012).

Ion exchange chromatography (IEX) is a chromatographic technique that separates molecules, including proteins, from a mixture based on the difference in net charge of the individual molecules (Whitford 2005). The stationary phase is a medium with either a positively (anion exchange) or negatively (cation exchange) charged surface that reversibly binds with the molecule of interest. Proteins with the same net charge as the medium will not be bound and will elute during the wash step.

There are four steps in performing IEX:

1. ***Equilibration***

   The column media is equilibrated with buffer with the aim of maximising the binding of the protein of interest to the column media.

2. ***Sample application and wash***

   The protein mixture is loaded and the column media rinsed to remove any unbound proteins.

3. ***Elution***

   The protein of interest is eluted by varying pH and ionic strength conditions. As the ionic strength increases, ions in the buffer compete with the charged side-chains of the protein for opposite charges on the IEX media and the protein of interest starts to elute.

4. ***Regeneration***

   The pH and ionic strength is increased further across a gradient to ensure that

no bound protein remains adsorbed onto the IEX media. The media can then be re-equilibrated, as in step 1.

High performance liquid chromatography (HPLC) is an analytical and preparative chromatographic technique consisting of a liquid mobile phase and a solid stationary phase. The basic setup of the essential components of a HPLC system is shown in figure 2.1.

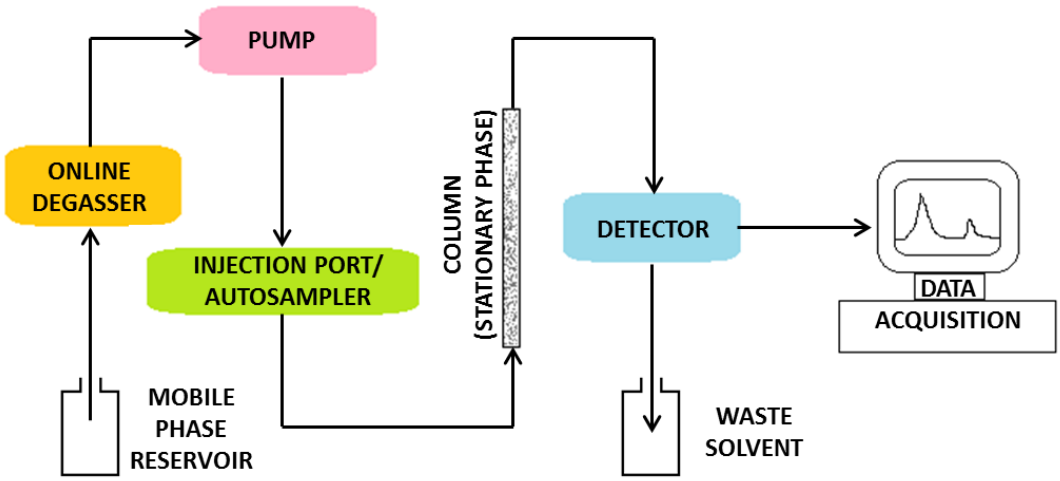

Figure 2.1: *Schematic of flow of operations and component setup in a HPLC system.*

The mobile phase is pumped through an online degasser before being pumped through the stationary phase (column) at high pressure. Samples for analysis are introduced manually *via* the injection port or *via* the autosampler. The interaction of the analyte/analytes with the stationary phase depends on the property of the stationary phase media, the mobile phase and the analyte itself. As the analyte elutes, it passes over a detector which, in turn, is connected to computer software that converts the signal to digital format for data analysis and processing before going to waste or a fraction collector.

2.3.1.2 Instrumentation

A XK 50/100 column (SEC) (GE Healthcare Bio-Sciences AB, Sweden) and a XK50/60 (IEX) column were both fitted with XK50 fast flow adapters (GE Healthcare Bio-Sciences AB, Sweden) and connected to an AKTAprime plus chromatography system (GE Healthcare Bio-Sciences AB, Sweden) comprising of a

pump, fraction collector, absorbance and conductivity monitors and valves for sample injection, gradient formation and flow diversion. Real-time data were viewed using PrimeView 5.0 software and recorded using PrimeView Evaluation software.

The SEC column was packed with Sephacryl S-200 High Resolution media (GE Healthcare Bio-Sciences AB, Sweden), a hydrophilic cross-linked allyl dextran and *N,N'*-methylene bisacrylamide co-polymer. The hydrophilic nature of the co-polymer minimises non-specific adsorption while the small cross-linking maximises rigidity and chemical stability. The narrow particle size distribution increases the resolution of the eluting peaks within the size range 5–250 kDa. The size exclusion column media was cleaned between separations of different proteins or after five separations of the same protein using 0.2 M NaOH in pure water and re-equilibrated using 5 column volumes of 0.275 M sodium acetate buffer at pH 4.5. The column was equilibrated between runs using 5 column volumes of 0.275 M sodium acetate buffer pH 4.5.

The IEX column was packed with SP Sepharose Fast Flow media (GE Healthcare Bio-Sciences AB, Sweden), a hydrophilic ion exchange media composed of 6% cross-linked agarose beads, which minimise non-specific interactions and are highly chemically and physically robust, with sulphopropyl (SP) strong cation exchange groups attached to the matrix *via* ether linkages. The column was regenerated by washing with 1 M NaOH after each use and re-equilibrating with 5 column volumes of 0.275 M sodium acetate buffer at pH 4.5.

HPLC analysis was carried out using a Shimadzu CBM-20A system controller attached to a Shidmadzu LC-20AT solvent delivery model, DGU-20A_3 on-line degasser, SIL-20AHT autosampler and SPD-M20A UV/Vis photodiode array detector. Shimadzu LC Solution software was used for real-time and post-run data analysis. A Superdex 200 10/300 GL size exclusion column composed of cross-linked agarose and dextran (GE Healthcare, Sweden) was used for SE-HPLC analysis.

2.3.1.3 Sample preparation and elution

2.3.1.3.1 Size exclusion chromatography

The filtered supernatant containing protein after bacterial cell lysis was loaded onto the Sephacryl S-200 XK 50/100 column at a flow rate of 2 ml min^{-1}. An isocratic elution using 0.275 M sodium acetate buffer pH 4.5 as the mobile phase and the parameters in table 2.3 were used. Fractions were collected in 30 ml borosilicate test tubes using an automated fraction collector.

Table 2.3: *Programmed method for size exclusion chromatographic separation of HGD and P23V protein from crude lysate.*

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

The SEC column was cleaned using 0.2 M NaOH and rinsed using 0.275 M sodium acetate buffer at pH 4.5 after 5 separations of the same protein or after each separation for different proteins.

2.3.1.3.2 Ion exchange chromatography

Fractions from the SEC step containing protein at the correct molecular weight were pooled and loaded onto the equilibrated IEX column at a flow rate of 10 ml min^{-1}. Unbound protein was eluted after loading using 0.275 M sodium acetate buffer at pH 4.5 at 10 ml min^{-1}. A gradient method that increased both the salt concentration and pH of the mobile phase (table 2.4 or table 2.5) was used for protein elution and fractions were collected in 30 ml borosilicate test tubes.

Table 2.4: *Programmed method for ion exchange chromatographic separation of HGD and P23V single mutant of HGD proteins for further purification after size exclusion chromatographic separation.*

Breakpoint	**Volume (*ml*)**	**Salt Buffer Concentration (*% v/v*)**	**Flow Rate (*ml min^{-1}*)**	**Fraction Volume (*ml*)**
1	0	15	10	0
2	1060	15	10	0
3	1061	30	10	0
4	1380	30	10	22
5	2904	43	10	22
6	3049	45	10	22
7	3470	85	10	0

Table 2.5: *Programmed method for ion exchange chromatographic separation of P23VR58H double mutant protein of HGD for further purification after size exclusion chromatographic separation.*

Breakpoint	**Volume (*ml*)**	**Salt Buffer Concentration (%)**	**Flow Rate (*ml min^{-1}*)**	**Fraction Volume (*ml*)**
1	0	30	10	0
2	600	30	10	0
3	601	30	10	22
4	2690	43	10	22

The IEX column was re-equilibrated by cleaning with 1 M NaOH at a flow rate of 10 ml min^{-1} and rinsing with 0.275 M sodium buffer at pH 4.5.

2.3.1.3.3 High performance liquid chromatography

Protein samples for HPLC analysis were prepared by dialysing protein against 0.1 M sodium phosphate at pH 7.0 using an Amicon Ultra-4 centrifugal filter with a 10,000

molecular weight cut off (MWCO) (Millipore, Ireland). Typically, a concentration of between 1–10 mg ml^{-1} was used. Protein samples were reduced using 20 mM DTT and filtered through a 0.22 μm Millex-GV syringe driven filter before adding to clear borosilicate glass HPLC autosampler vials (Chromacol, Germany) which were then sealed with a cap containing a rubber septum (ThermoScientific, USA). The parameters used for analysis are outlined in table 2.6.

Table 2.6: *Parameters for analysis using SE-HPLC using a Superdex 200 10/300 GL size exclusion column.*

Stationary Phase	Superdex 200 10/300 GL
Mobile Phase	0.1 M sodium phosphate, pH 7.0
Flow Rate	0.75 ml min^{-1}
Pressure Limit	14 bar
Analysis Time	60 min
Sample Volume	20 μl

Sample absorbance at a wavelength of 280 nm was measured and chromatograms were exported as ASCII files and analysed using OriginPro 9.1 data analysis software.

2.3.1.3.3.1 Calibration curve: Molecular weight

A calibration curve (figure 2.3) was constructed to determine the relationship between retention time (RT) and the log of the molecular weight of the protein using a Superdex 200 10/300 GL size exclusion column. The molecular weight range used was between 10 kDa and 300 kDa. Table 2.7 indicates the proteins analysed, their corresponding molecular weights, retention times and peak numbers. Gamma globulin, bovine serum albumin (BSA), chicken serum albumin and myoglobin were purchased from Sigma Aldrich (USA). Lysozyme was purchased from Calbiochem (Canada).

Protein samples were prepared in 0.1 M sodium phosphate buffer at pH 7.0 at approximately 2 mg ml^{-1} and filtered through a Millex-GV 0.22 μm syringe filter prior to analysis. A Superdex 200 10/300 GL size exclusion column and a flow rate

of 0.75 ml min^{-1} were used for analysis. The mobile phase used was 0.1 M sodium phosphate buffer at pH 7.0. The chromatogram (figure 2.2) was reconstructed and analysed and the calibration curve (figure 2.3) plotted using OriginPro 9.1 data analysis software.

Table 2.7: *Proteins used for Superdex 200 10/300 GL size exclusion column calibration and their corresponding molecular weights (M_W), peak numbers and retention times (RT).*

Peak Number	***Protein***	***M.W.*** **(*kDa*)**	***log$_{10}$(M_W)*** **(*kDa*)**	***RT*** **(*mins*)**
1	Gamma Globulin (*dimer*)	300	2.48	10.38
2	Gamma Globulin	150	2.18	13.58
3	BSA (*dimer*)	133.4	2.13	16.33
4	BSA	66.7	1.82	18.92
5	Chicken Serum Albumin	44.2	1.65	20.68
6	Myoglobin	16.7	1.22	23.17
7	Lysozyme	14.7	1.17	25.07

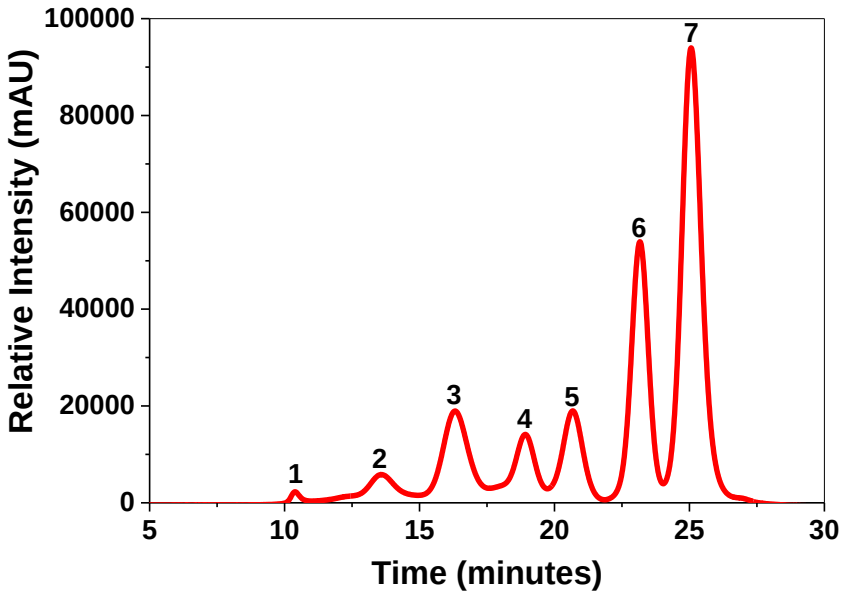

Figure 2.2: *Elution profile of protein mixture used for size exclusion column calibration.*

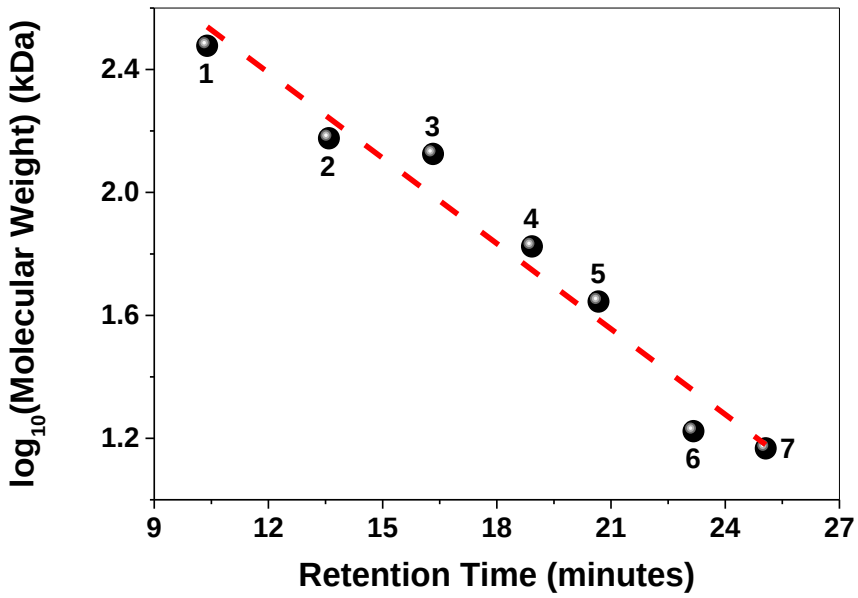

Figure 2.3: *Calibration curve for protein mixture showing log_{10} of protein molecular weight as a function of retention time.*

The calibration curve (figure 2.4) was fitted with a linear fit given in equation 2.8:

$$y = 3.50 - 0.09\,x \qquad \textbf{2.8}$$

where y is the log base 10 of the protein molecular weight in kDa and x is the retention time in minutes.

2.3.1.3.3.2 Calibration curve: Protein concentration

A calibration curve (figure 2.4) was constructed to determine the relationship between HGD concentration and the area under the peak after size exclusion HPLC analysis using a Superdex 200 10/300 GL size exclusion column and the parameters in table 2.6. HGD concentration was determined spectroscopically using the Beer-Lambert law (section 2.4.2.2). A concentration range of 1 to 10 mg ml^{-1} was used.

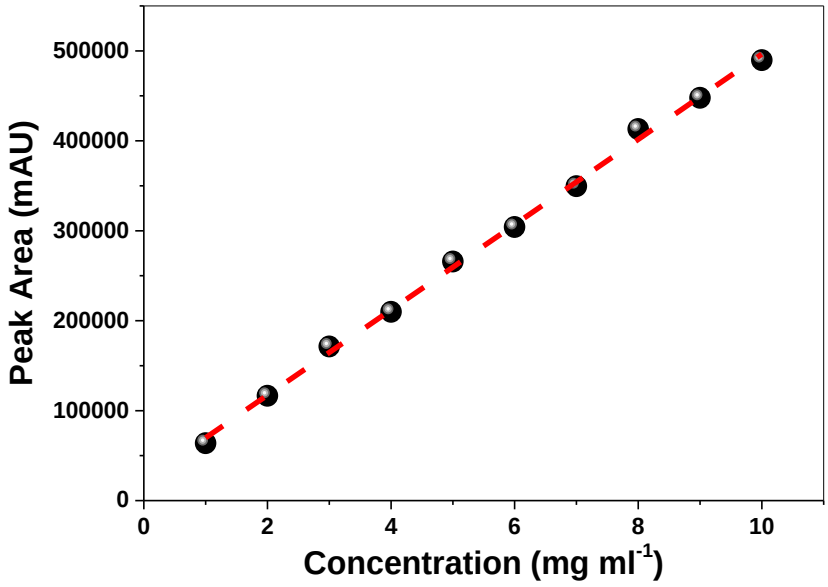

Figure 2.4: *Calibration curve of peak area as a function of HGD concentration.*

All samples were prepared using 0.1 M sodium phosphate buffer at pH 7.0 with DTT at a final concentration of 20 mM since HGD forms a dimer in the absence of a reducing agent (figure 2.5) (Pande, Pande et al. 2000).

All samples were filtered through a Millex-GV 0.22 µm syringe driven filter (Millipore, Ireland) before analysis. Samples were run using a flow rate of 0.75 ml min^{-1} and the mobile phase was 0.1 M sodium phosphate buffer at pH 7.0. Chromatograms were reproduced analysed and the calibration curve constructed using OriginPro 9.1 data analysis software. The linear relationship between peak area and protein concentration is given by equation 2.9:

$$y = 22499.98 + 47393.60\,x \qquad \textbf{2.9}$$

where y is the peak area in AU and x is the concentration of HGD in mg ml^{-1}.

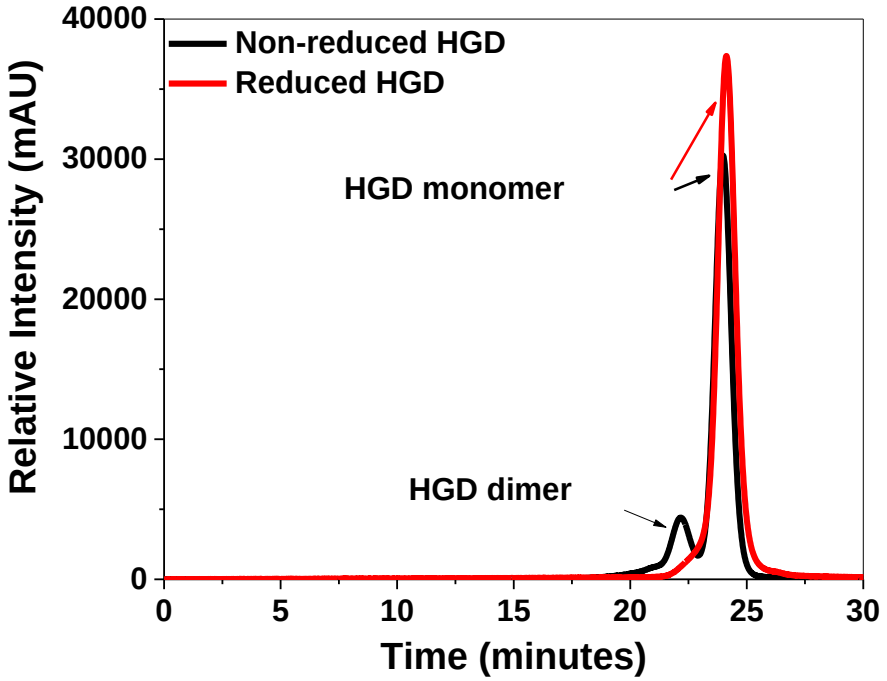

Figure 2.5: *SE-HPLC chromatogram showing non-reduced (black) and reduced HGD (red) at approximately 2 mg ml^{-1}. The minor peak (first peak) arises due to the presence of HGD dimer (~23%). When a reducing agent is used (20 mM DTT), only one peak is observed.*

2.3.2 Polyacrylamide gel electrophoresis

2.3.2.1 Background

Electrophoresis refers to the movement of particles within an electric field. Polyacrylamide gel electrophoresis (PAGE) is used for the separation and analysis of biomolecules (i.e. proteins and nucleic acids) and involves the migration of a biomolecule through a polyacrylamide gel network as a result of an electric field (Whitford 2005). Native PAGE separates biomolecules on the basis of net charge and size. Sodium dodecyl sulphate (SDS) PAGE separates proteins on the basis of their molecular weights alone (Weber, Pringle et al. 1972). After denaturation (e.g. heat denaturation), SDS will bind to the polypeptide chain of the protein (approximately 1.4 g SDS per gram of protein) to mask the inherent charge of the protein (determined by its pI) and impart a net negative charge to the protein (Whitford 2005). A reducing agent such as β-mercaptoethanol can be used to break disulphide bridges between cysteine residues (Whitford 2005).

2.3.2.2 Preparation of running and stacking gels

12.6% running gels were prepared using a recipe outlined in table 2.8. Components were mixed together and then degassed for at least 15 minutes. All buffers and reagents were prepared using ultrapure water. Sodium dodecyl sulphate (SDS, $NaC_{12}H_{25}SO_4$, M_w = 288.37 g mol^{-1}) was purchased from Riedel-de Haën Sigma Aldrich (Germany). 30% Acrylamide/Bis (37.5:1) was purchased from BioRad (USA).

Table 2.8: *Components and volumes required for 12.6% running polyacrylamide gel preparation.*

Component	Volume (*ml*)
Ultrapure water	3.15
1 M Tris-HCl pH 8.8	2.50
10% (w/v) SDS	0.10
Acrylamide/Bis (30%)	4.20

50 µl of freshly prepared 10% (w/v) ammonium persulphate ($(NH_4)_2S_2O_8$, M_W = 228.18 g mol^{-1}) (Fisher Scientific, UK) and 5 µl tetramethylethylenediamine (TEMED, $C_6H_{16}N_2$, M_W = 116.24 g mol^{-1}) were added to the degassed mixture. Gels were cast using a casting frame and stand designed for use with a short glass plate and a spacer plate with a 1 mm spacer (all purchased from BioRad, USA). Tert-amyl alcohol (~200 µl) was added to the top of the gels to extrude any air bubbles. Gels were set for at least 30 minutes and rinsed with ultrapure water before casting the stacking gel. 4% stacking gels were prepared as per table 2.9. The components were mixed together and degassed for at least 15 minutes.

Table 2.9: *Components and volumes required for 4% stacking polyacrylamide gel preparation.*

Component	Volume (*ml*)
Ultrapure water	3.05
0.5 M Tris-HCl pH 6.8	1.25
10% (w/v) SDS	0.05
Acrylamide/Bis (30%)	0.65

25 µl of freshly prepared 10% (w/v) ammonium persulphate and 5 µl TEMED were added to the degassed mixture. The stacking gel was set for at least 30 minutes before use.

2.3.2.3 Preparation of sample buffer

Stock sample buffer was prepared using the recipe outlined in table 2.10. Glycerol was purchased from Sigma Aldrich (USA). Bromophenol blue was purchased from Riedel-de Haën Sigma Aldrich (Germany).

Table 2.10: *Components and volumes required for stock sample buffer preparation.*

Component	Volume (*ml*)
Ultrapure water	4.8
0.5 M Tris-HCl pH 6.8	1.2
Glycerol	1.0
10% (w/v) SDS	2.0
0.1% (w/v) bromophenol blue	0.5

Reducing sample buffer was prepared immediately before use using 475 µl of the stock sample buffer and 25 µl β-mercaptoethanol (Sigma Aldrich, Belgium).

2.3.2.4 Preparation of molecular weight standards

SDS-PAGE Molecular Weight Standards, Low Range (figure 2.6) were purchased from BioRad (USA). 7 µl aliquots were prepared by diluting the standards 1:20 with the stock sample buffer and stored at -18 °C. Prepared standards were heated at ~95 °C for approximately 5 minutes, cooled and briefly centrifuged before loading. 5 µl of the standard were loaded per well.

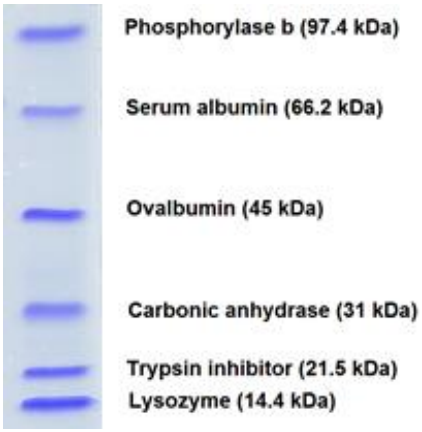

Figure 2.6: *Low range molecular weight markers and their molecular weights run on a 12.6% SDS polyacrylamide gel stained with Coomassie R-250.*

2.3.2.5 Preparation of destaining solution

Destaining solution was prepared by mixing ultrapure water, methanol (Sigma Aldrich, UK) and acetic acid respectively in a 60:30:10 (v/v) ratio.

2.3.2.6 Preparation of protein samples for SDS-PAGE analysis

Protein samples for SDS-PAGE analysis were prepared by dialysing protein against 0.1 M sodium phosphate buffer at pH 7.0 in an Amicon-4 Ultra centrifugal device and then mixing with sample buffer. Samples were prepared to yield 2–10 µg of protein per well with each well having a maximum loading capacity of 30 µl. Samples were heated at ~95 °C for approximately 5 minutes, cooled and briefly centrifuged before loading.

2.3.2.7 Running SDS-PAGE and analysis

SDS-PAGE gels were run using a Mini-PROTEAN Tetra Cell electrophoresis system (BioRad, USA) connected to an external power source. 1X Tris/glycine/SDS running buffer was prepared by diluting 10X pre-mixed Tris/glycine/SDS buffer (BioRad, USA) with ultrapure water. Protein samples were carefully pipetted into the wells. Empty wells were filled with sample buffer. Gels were run at 200 V until the dye front was ~1 cm from the bottom of the gel. Gels were stained using Coomassie Brilliant Blue R-250 (BioRad, USA) for at least 1 hour, destained using destaining solution and stored in 5% (v/v) acetic acid. Gels were imaged using an Epson Perfection V300 Photo scanner and Epson Scan software.

2.4 Protein characterisation

2.4.1 Mass spectrometry

2.4.1.1 Background

Mass spectrometry (MS) is a technique used for molecular weight determination based on the ratio of mass to charge (*m/z*) in a sample (Whitford 2005). There are several types of MS available but all include an ion source, an analyser and a detector and follow a basic operational setup whereby samples are loaded and ionised by the ion source (Whitford 2005). Ionised molecules are accelerated through an electric and/or magnetic field generated by the analyser which deflect the path based on their *m/z* ratio. The deflected ions are then detected by the detector which, in turn, is connected to computer software that converts the signal to digital format for data processing and analysis. Additional processes are sometimes used prior to or in tandem with MS for enhanced analysis.

2.4.1.2 Sample preparation

Protein samples for molecular weight confirmation were analysed using liquid chromatography electrospray ionisation quadrupole time of flight-MS (LC-ESI-Q-TOF-MS) analysis. Samples were prepared at a concentration of 0.5 mg ml^{-1} by exhaustively dialysing against 0.1 M sodium phosphate buffer at pH 7.0 in an Amicon-4 Ultra centrifugal filter and filtered through a 0.22 μm Millex-GV syringe driven filter prior to analysis. MS analysis was carried out in 'FingerPrints' Proteomics Facility, Post-Genomics and Molecular Interactions Centre, College of Life Sciences, University of Dundee, Scotland, UK.

2.4.2 Spectroscopy

2.4.2.1 Theory

Frequency and wavelength are related to the speed of light as per equation 2.10:

$$c = \lambda \nu \qquad \textbf{2.10}$$

where c is the speed of light equal to 2.998 x 10^8 m s^{-1}, λ is the wavelength in m and v is the frequency in s^{-1} or Hz. The energy associated with a photon of light is related to frequency as per equation 2.11:

$$E = h\nu = h(c/\lambda) \qquad \textbf{2.11}$$

where E is the energy of the photon in J, h is Planck's constant equal to 6.63 x 10^{-34} J s, v is the frequency in s^{-1} or Hz, c is the speed of light equal to 2.998 x 10^8 m s^{-1} and λ is the wavelength in m.

Spectroscopy is the measurement of the interaction of electromagnetic (EM) radiation with matter and can be used for analysing protein structure.

2.4.2.2 UV/Vis absorbance spectroscopy

Proteins absorb light in the UV/Vis region of the EM spectrum as a result of their aromatic side chains (mainly Tyr and Trp and, to lesser extent, Phe) and also due to the peptide bond. These amino acids absorb in the "near-UV" (i.e. 240–300 nm). Disulphide bonds also absorb light in this region at a wavelength of around 260 nm. The peptide bond between amino acid residues absorbs light in the far-UV wavelength region (i.e. 180–230 nm) (Schmid 2001).

Protein concentration in mg ml^{-1} or M can be determined spectroscopically using the Beer-Lambert Law (equation 2.12).

$$A = \varepsilon cl \qquad \textbf{2.12}$$

where A is the absorbance at λ = 280 nm, ε is the extinction coefficient (a term relating to how light is absorbed, also called molar absorptivity) in mg^{-1} ml cm^{-1} or M^{-1} cm^{-1} for a given protein, c is the concentration in mg ml^{-1} or M and l is the path-length in cm.

An extinction coefficient of ε = 2.09 mg^{-1} ml cm-1 (Wang, Lomakin et al. 2010) or ε = 42,860 M^{-1} cm^{-1} was used for HGD and its mutants when determining protein concentration.

2.4.2.3 Second derivative UV/Vis absorbance spectroscopy

2.4.2.3.1 Background

Derivative spectroscopy was used to analyse the broad absorbance peak arising between 240 and 300 nm. Since this peak is a result of light absorbed by Phe (weakly absorbs at wavelengths of between 245 to 270 nm), Tyr (265 to 285 nm) and Trp (265 to 295 nm) (Kueltzo, Ersoy et al. 2003), the second derivative of the absorbance spectrum can be used to extract data from the broad peak to provide information on each of the aromatic residues and, consequently, information about the secondary and the tertiary structure of the protein.

Absorbance (A) is expressed as a function of wavelength (λ) (equation 2.13). The first order derivative is given in equation 2.16 and the second order derivative is given in equation 2.15.

Zero Order $$A = f(\lambda) \qquad \textbf{2.13}$$

First Order $$dA/d\lambda = f'(\lambda) \qquad \textbf{2.14}$$

Second Order $$d^2A/d\lambda^2 = f''(\lambda) \qquad \textbf{2.15}$$

The signal-to-noise ratio (S/N) of the spectral data decrease with successive differentiations. To increase S/N without significantly distorting the spectral data after differentiation, the spectral data are smoothed using a Savitzky-Golay filter (polynomial order = 2, window size = 5). This method of averaging preserves the integrity of the data by performing a local polynomial regression to produce a smoothed value for each data point.

2.4.2.3.2 Sample preparation and analysis

Protein samples for second derivative UV/Vis spectroscopic analysis were prepared at a concentration of 0.1 mg ml^{-1} by exhaustively dialysing against 0.1 M sodium phosphate buffer at pH 7.0 in an Amicon-4 Ultra centrifugal filter and filtered through a 0.22 μm Millex-GV syringe driven filter prior to analysis. Samples were recorded using a Perkin Elmer Lambda 35 UV/Vis spectrometer and UV WinLab Scan Lambda 35 software. A scan rate of 240 nm min^{-1} in data intervals of 0.1 nm

for a wavelength range between 250–320 nm and was used for analysis. Spectra were plotted and analysed using OriginPro 9.1 data analysis software.

2.4.2.4 Circular dichroism spectroscopy

2.4.2.4.1 Background

Circular dichroism (CD) is a spectroscopic method for assessing both the secondary and tertiary structures of a protein. CD refers to a differential in the absorption of clockwise (right-handed, R) and anticlockwise (left-handed, L) circularly polarised light by a chiral (optically active) molecule (Kelly, Jess et al. 2005). All amino acids with the exception of glycine are chiral molecules; thus, proteins are inherently optically active.

Spectral data arising from CD are plotted as dichroism (also reported as ellipicity in degrees) as a function of wavelength. The near-UV CD spectrum ranges between 320–260 nm and is representative of protein tertiary structure since the aromatic amino acids (i.e. Trp, Tyr and Phe) absorb in this region (Kelly and Price 2000; Kelly, Jess et al. 2005). Disulphide bonds also have weak absorption bands around 260 nm (Kelly, Jess et al. 2005). The far-UV CD spectrum which ranges between 240–180 nm corresponds to peptide bond absorption and provides information on protein secondary structure features (i.e. α-helices and β-sheets) (Kelly and Price 2000; Kelly, Jess et al. 2005).

2.4.2.4.2 Sample preparation and analysis

Protein samples for near and far ultraviolet (UV) circular dichroism (CD) spectroscopic analysis were prepared at a concentration of 0.1 mg ml^{-1} by exhaustively dialysing against 0.1 M sodium phosphate buffer at pH 7.0 in an Amicon-4 Ultra centrifugal filter and filtered through a 0.22 μm Millex-GV syringe driven filter prior to analysis. Near- and far-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) using 0.1 M sodium phosphate buffer at pH 7.0 as a reference. CD spectral data were plotted using OriginPro 9.1 data analysis software.

2.5 Chemical modification: Fluorescent labelling

HGD was chemically modified at two different sites using two different dyes with similar molecular weights and photo-physical properties. HGD was also labelled using FITC which interacts with different amino acid side-groups on the protein surface.

2.5.1 Lysine modification

The ε-amino group of Lys-2, which is a primary amine, was chemically modified using a fluorescent dye (AnaTag^{TM} HiLyte Fluor^{TM} 405 Microscale Protein Labelling Kit (Eurogentec, USA)). The succinimidyl ester of the dye selectively reacts with the aliphatic amine in Lys to form a stable carboxamide bond (figure 2.7).

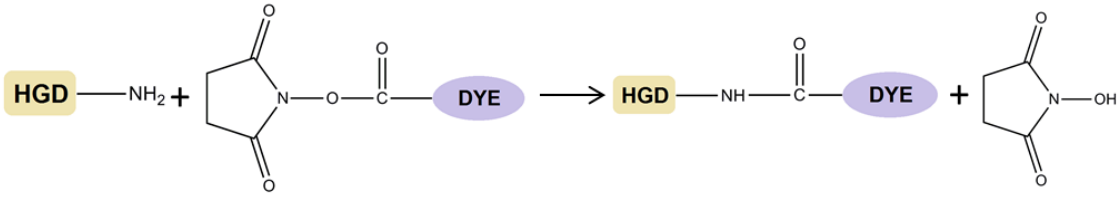

Figure 2.7: *Labelling the primary amine of Lys-2 with the succinimidyl ester of the dye.*

HGD (M_w = 20 608 Da) was dialysed against 0.1 M sodium phosphate buffer pH 7.0 without sodium azide (since the fluorescent dye would react with the azide) and used at a concentration of 11 mg ml^{-1}. The pH of the protein solution was adjusted for optimal Lys modification using supplied reaction buffer (Component B) of the kit at a 1:10 Component B:HGD ratio. 10 µl of dimethylsulphoxide (DMSO, C_2H_6SO, M_w = 78.13 g mol^{-1}) (Component D) was added to a vial of dye and contents vortexed. ~40 µl of HGD at 10 mg ml^{-1} was added to dye. The mixture was reacted for at least 1 hour with gentle mixing. The excess dye was removed using a spin column (Component C) and the dyed protein collected in a collection tube (Component G). Further purification of the dye-protein conjugate was done using an Amicon Ultra-4 centrifugal filter and exhaustively dialysing against 0.1 M sodium phosphate at pH 7.0. Exhaustive dialysis was required to ensure no unbound dye remained in solution. The fluorescence intensity of the filtrate was analysed using a Molecular Devices Spectra Max spectrofluorometer and Soft Max software by exciting at λ_{ex} = 404 nm and measuring emission at λ_{em} = 428 nm. The degree of substitution (DOS) (Equation 2.18) was calculated using the molar concentration of the dye (equation

2.16) and protein (equation 2.17). The labelled protein was stored at 4 °C. Molar concentration of dye:

$$[Dye] = \frac{(A_{\max} \times dilution\ factor)}{\varepsilon_{fluor}} \qquad \textbf{2.16.}$$

where A_{max} is the absorbance at λ = 404 nm and ε_{fluor} is the molar extinction coefficient for HiLyte Fluor™ 405 equal to 34,000 M^{-1} cm^{-1}. The molar concentration of the protein is determined by:

$$[HGD] = \frac{(A280 - (0.85 \times A_{\max}) \times dilution\ \ factor)}{\varepsilon_{HGD}} \qquad \textbf{2.17}$$

where A_{max} is the absorbance at λ = 404 nm and ε_{HGD} is the molar extinction coefficient for HGD equal to 42,860 M^{-1} cm^{-1}. 0.85 is a correction factor for the dye to correct for the contribution of the dye to absorption at λ = 280 nm. The degree of substitution is calculated from:

$$DOS = \frac{[Dye]}{[HGD]} \qquad \textbf{2.18}$$

2.5.2 Thiol modification

The thiol group of Cys-110 was chemically modified using a fluorescent dye (DyLight 405 Maleimide) (ThermoScientific, USA). The maleimide group of the dye reacts predominantly with the reduced thiol group of a cysteine to form a stable thioether bond (figure 2.8).

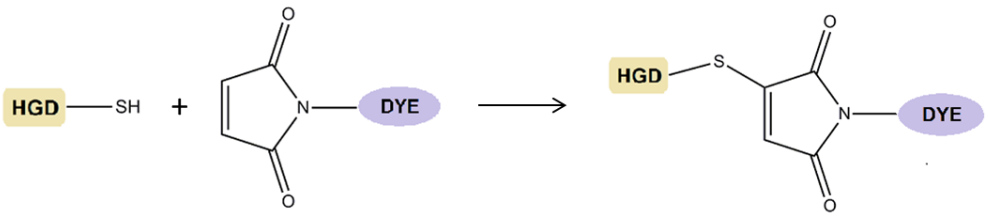

Figure 2.8: *Fluorescent labelling of the thiol group of Cys-110 via conjugation with the maleimide group of the dye to form a thioether bond.*

HGD was dialysed against reducing buffer (section 2.1.2.1) and concentrated to 10 mg ml^{-1} before adding 1 ml of this to one vial (6 mg) of 2-Mercaptoethylamine.HCl (2-MEA, Pierce Biotechnology, Rockford, IL, USA). The contents of the vial were mixed by gentle inversion and incubated at 37 °C for 90 minutes. After cooling the reduced HGD and 2-MEA solution to room temperature, 2-MEA was removed using a Dextran Desalting Column (Pierce Biotechnology, USA). The fluorescent dye was dissolved in 100 µl dimethyl fumarate (DMF) (Romil, UK). The tube was then vortexed for 30 seconds and incubated at room temperature for 5 minutes. The dye solution was then gently pipetted to dissolve contents completely. 80 µl of the dye was added to the reduced HGD solution and mixed. The dye-protein mixture was incubated overnight at room temperature with gentle shaking. Unreacted dye was removed from the solution by exhaustive dialysis against 0.1 M sodium phosphate buffer at pH 7.0 using an Amicon Ultra-4 centrifugal filter. The fluorescent activity of the filtrate was analysed using a Molecular Devices Spectra Max spectrometer and Soft Max software by exciting at λ_{ex} = 400 nm and measuring emission at λ_{em} = 420 nm. The Degree of Substitution (DOS) (equation 2.20) was determined using the molar concentration of the protein (equation 2.19) as outlined below:

Molar concentration of the protein:

$$[HGD] = \frac{(A_{280} - (A_{\max} \times 0.564)}{\varepsilon_{HGD}} \times dilution\ factor \qquad \textbf{2.19}$$

$$DOS = \frac{A_{\max}\ labelled\ HGD \times dilution\ factor}{\varepsilon_{fluor} \times [HGD]} \qquad \textbf{2.20}$$

where A_{280} is the absorbance at λ = 280 nm, A_{max} is the absorbance at λ = 404 nm, 0.564 is a correction factor for the dye, ε_{HGD} is the molar extinction coefficient for HGD equal to 42,860 M^{-1} cm^{-1} and ε_{fluor} is the extinction coefficient of the DyLight 405 Maleimide dye equal to 30,000 M^{-1} cm^{-1}.

2.5.3 Modification with FITC

Fluorescein isothiocyanate (FITC, M_W = 389.38 g mol^{-1}, ThermoScientific, USA) can form stable conjugates with both primary and secondary amines and was used to non-selectively fluorescently label HGD (figure 2.9). FITC was dissolved in DMF at

a concentration of 10 mg ml^{-1}. HGD was exhaustively dialysed against 50 mM sodium borate buffer at pH 8.5 (section 2.1.3) before conjugation. A 50 µl solution of FITC in DMF (10 mg ml^{-1}) was added to 1 ml of HGD at 3 mg ml^{-1}. Unreacted FITC was removed with by exhaustively dialysing against 0.1 M sodium phosphate buffer at pH 7.0. This was monitored by analysing the fluorescent activity of the filtrate using an excitation and emission wavelengths equal to 494 nm and 518 nm respectively.

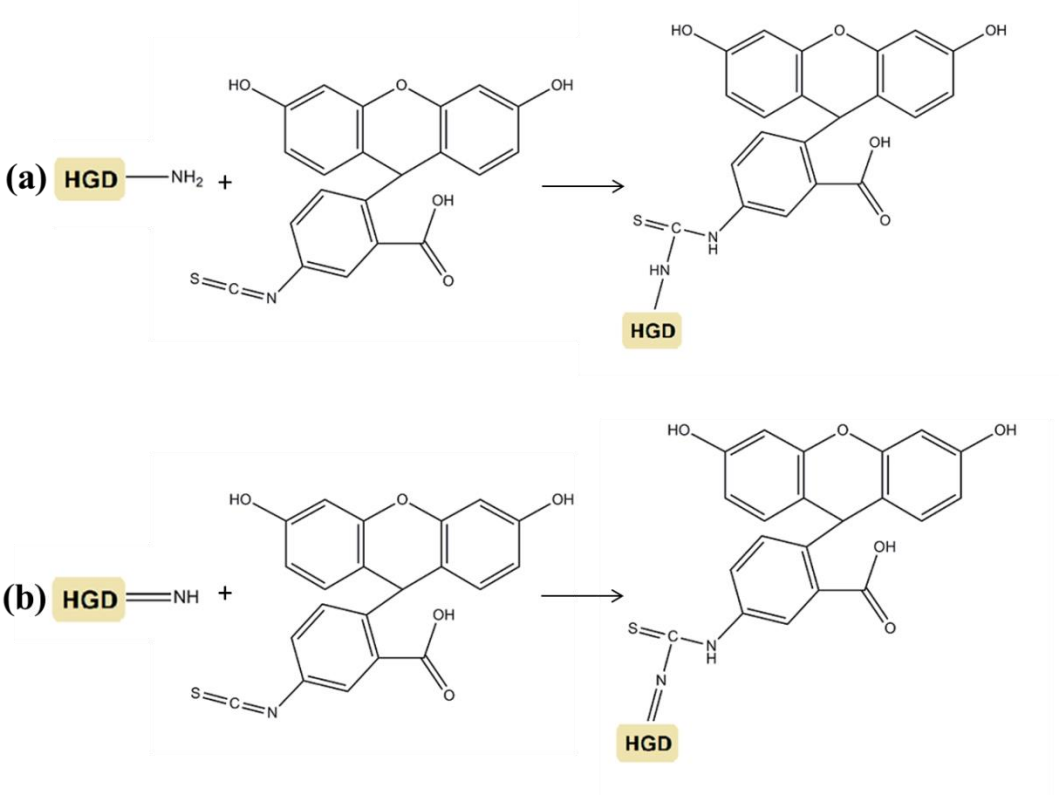

Figure 2.9: *Labelling the primary (a) and secondary (b) amines in HGD with FITC to form a thiourea bond.*

The Degree of Substitution (DOS) (equation 2.22) was determined using the molar concentration of the protein (equation 2.21) as outlined below:

Molar concentration of the protein:

$$[HGD] = \frac{(A_{280} - (A_{\max} \times 0.300)}{\varepsilon_{HGD}} \times dilution\ factor \qquad \textbf{2.21}$$

$$DOS = \frac{A_{\max}\ labelled\ HGD \times dilution\ factor}{\varepsilon_{fluor} \times [HGD]} \qquad \textbf{2.22}$$

where A_{280} is the absorbance at λ = 280 nm, A_{max} is the absorbance at λ = 494 nm, 0.300 is a correction factor for the dye, ε_{HGD} is the molar extinction coefficient for HGD equal to 42,860 M^{-1} cm^{-1} and ε_{fluor} is the extinction coefficient of FITC dye equal to 68,000 M^{-1} cm^{-1}.

2.6 Chemical modification: Biotinylation

2.6.1 Lysine modification

The primary amine of Lys-2 in HGD was chemically modified with EZ-Link NHS-PEG_4-Biotin packaged in vials containing 2 mg of the label (Pierce Biotechnology, USA). The NHS ester moiety (*N*-Hydroxysuccinimde) attached to the PEGylated biotin reacts specifically with primary amines to form a stable carboxamide bond with release of the NHS (figure 2.10).

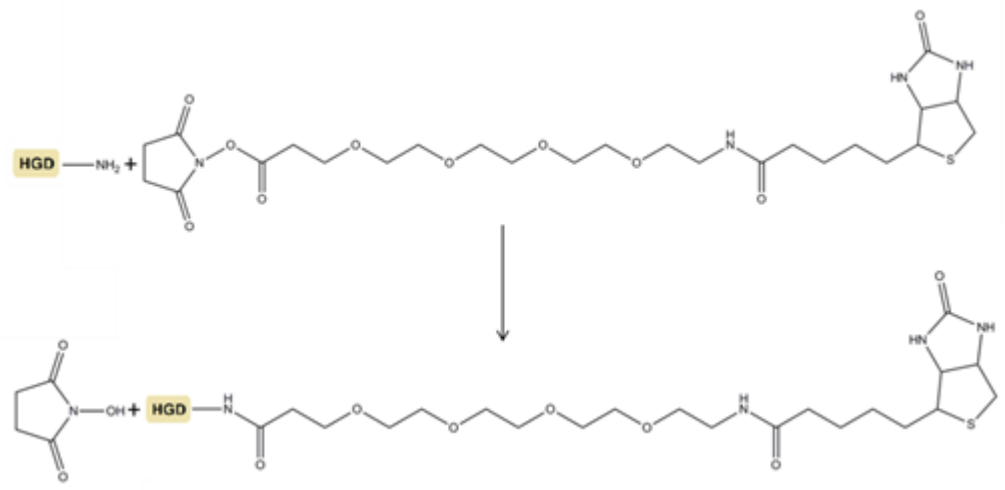

Figure 2.10: *Labelling the primary amine of Lys-2 with the succinimidyl ester of PEGylated biotin.*

HGD was dialysed against 0.1 M sodium phosphate at pH 8.0 without sodium azide. The biotin was dissolved immediately before use in 170 µl 0.1 M sodium phosphate buffer pH 7.0 (also without azide) and mixed with 1 ml of HGD at 10 mg ml^{-1}. The reaction was incubated overnight at room temperature. Excess unreacted biotin was removed by exhaustively dialysing the modified HGD against 0.1 M sodium

phosphate buffer at pH 7.0. The degree of substitution was quantified using a Fluorescence Biotin Quantification Kit (Pierce Biotechnology, USA) (section 2.6.3).

2.6.2. Thiol modification

The thiol group of Cys-110 in HGD was chemically modified with EZ-Link-Maleimide-PEG$_2$-Biotin packaged in vials containing 2 mg of the label (Pierce Biotechnology, USA). The maleimide group of the PEGylated biotin reacts specifically with reduced thiol groups at pH 6.5–7.5 to form a stable thioether bond (figure 2.11).

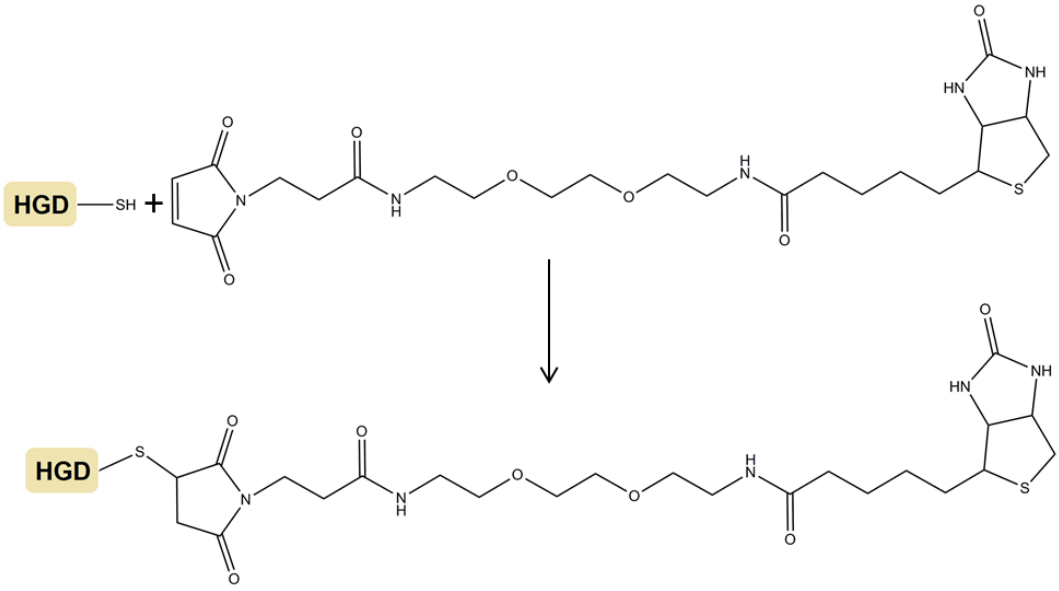

Figure 2.11: *Labelling the thiol group of Cys-110 with the succinimidyl ester of the dye.*

HGD was dialysed against reducing buffer (section 2.1.2.1) and 1 ml of a 10 mg ml^{-1} solution was added to one vial (6 mg) of 2-MEA. The contents of the vial were mixed by gentle inversion to dissolve and incubated at 37 °C for 90 minutes. The reduced HGD and 2-MEA solution was cooled to room temperature and the 2-MEA was removed using a Dextran Desalting Column. Biotin was dissolved immediately before use in 190 µl of a 0.1 M sodium phosphate buffer pH 7.0 and then added to 1 ml of HGD at 10 mg ml^{-1}. The reaction was incubated at room temperature overnight. Excess unreacted biotin was removed by exhaustively dialysing the modified HGD against 0.1 M sodium phosphate buffer at pH 7.0. The degree of substitution was quantified using a Fluorescence Biotin Quantification Kit (section 2.6.3).

2.6.3 Biotinylation efficiency determination

The biotinylated protein was quantified using a Pierce Fluorescence Quantitation Kit (Pierce Biotechnology, USA) as per manufacturer's guidelines. The standards were made up as per the following table using the 1 mM Biocytin Control (supplied) and the 1X PBS. By comparing the fluorescence intensity of the Biocytin Control with the fluorescence intensity of the biotinylated protein after assaying, it is possible to quantify the amount of biotinylated protein.

Table 2.11: *Biocytin standard preparation for biotin quantification.*

Vial	**1X PBS volume** ***(µl)***	**Volume & Biocytin Source** ***(µl)***	**Final Biocytin Conc.** **(*pmol 10µl^{-1}*)**
A	495	5 µl of Biocytin Control (***Vial A***)	100
B	20	80 µl Vial A	80
C	40	60 µl Vial A	60
D	60	40 µl Vial A	40
E	80	20 µl Vial A	20
F	90	10 µl Vial A	10
G	95	5 µl Vial A	5
H	100	0 µl Vial A	0 (Blank)

The DyLight Reporter Working Reagent (DWR) (supplied) was prepared at a 1:14 dilution by mixing 170 µl of the DWR with 2.38 ml 1X PBS. The standards and samples were analysed in triplicate by pipetting 10 µl of each standard and sample into a Nunc 96 black polypropylene MicroWell plate. 90 µl of DWR was added to each well containing the sample or the standard. The plate was covered and incubated at room temperature for 5 minutes. Fluorescence was analysed using excitation and emission wavelengths of 494 nm and 520 nm respectively. A standard curve was constructed of average relative fluorescence intensity as a function of Biocytin Control concentration within the linear range (10–60 pmol biocytin 10 µl^{-1}) using OriginPro 9.1 data analysis software. Biocytin Control concentration in pmol

10 μl^{-1} was calculated by substituting the average relative fluorescence intensity of the samples for y in the linear equation ($y = mx + c$, where y is the fluorescence intensity in AU, x is the Biocytin Control concentration in pmol 10 μl^{-1} and c is the y-intercept). HGD concentration in pmol 10 μl^{-1} was calculated spectroscopically as outlined in section 2.4.2.2. The DOS was calculated using equation 2.23:

$$\frac{pmol\ Biocytin\ 10\ \mu l^{-1}}{pmol\ HGD\ 10\ \mu l^{-1}} \qquad \textbf{2.23}$$

2.7 Liquid-liquid phase separation measurements

2.7.1 Protein sample preparation

Protein samples were prepared for liquid-liquid phase separation (coexistence curve) measurement by dialysing protein against 0.1 M sodium phosphate buffer at pH 7.0 and concentrating using an Amicon Ultra-4 centrifugal filter. Protein was then centrifuged at 15,000 RPM using a ThermoScientific Heraseus 3R+ temperature controlled centrifuge to pellet any large aggregates and the concentration of the supernatant was calculated. DTT was then added to a concentration of 20 mM.

2.7.2 Liquid-liquid phase boundary measurement

The liquid-liquid phase boundary (coexistence curve) for native HGD was measured using the same method as described previously (Broide et al., 1991, Asherie, 2004). A 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 percentage transmission (%T) of the sample was recorded at λ = 600 nm. The temperature of the system was controlled by an external thermostated water bath (ThermoScientific DC10-K10) and the temperature in the sample was monitored using an Omega HH509R thermocouple. As the protein solution was cooled, the temperature at which the intensity of transmitted light fell to half its initial value was taken as the clouding temperature (T_{cloud}). Once the transmitted intensity reached zero, the protein solution was warmed and the temperature at which the intensity of transmitted light returned to half of its initial value was defined as the clearing temperature (T_{clear}). The phase separation (T_{ph}) temperatures were calculated as the average of T_{cloud} and T_{clear}. The error value for

each point is taken as plus or minus the half the difference between T_{cloud} and T_{clear}. Values for T_{ph} (including error bars) were plotted as a function of protein concentrations. Coexistence curves were plotted using OriginPro 9.1 software.

2.7.3 Chemically modified protein partitioning analysis

A protein solution containing a mixture of modified and native protein at a known volume, concentration and modified protein fraction was centrifuged at 3,000 RPM in a ThermoScientific Heraseus 3R+ temperature controlled centrifuge at a temperature well below the phase separation temperature for 3 hours to macroscopically separate the dense and dilute protein phases. A volume of protein was pipetted from each phase and diluted appropriately to measure the protein and dye concentrations respectively. To quantify the protein modified with a fluorescent label in each phase, fluorescence intensity measurements were taken for protein in both phases; the relative proportions in each phase were then determined. For protein mixtures containing the protein modified with PEGylated biotin, the Pierce Fluorescence Biotin Quantitation kit was used as before to quantify the amount of labelled protein in each phase.

2.8 Coverslip surface conditions for determining surface effect on protein aggregation

The temperature induced assembly of a HGD double mutant (P23VR58H) was monitored in a chamber mounted on a microscope slide. A number of different surface treatments were used on the glass surface. Protein samples were prepared by dialysing against 0.1 M sodium phosphate buffer at pH 7.0 and concentrated using an Amicon Ultra-4 centrifugal filter. Care was taken when preparing the P23VR58H double mutant protein to keep the temperature at 4 °C to avoid any premature aggregate formation since P23VR58H mutants can form aggregates spontaneously at room temperature. P23VR58H was used at a concentration of 10 mg ml^{-1}. DTT was added to protein sample to a final concentration of 20 mM. All samples were filtered using a 0.22 μm Millex-GV syringe driven filter prior to analysis and the protein sample was introduced to the chambers for imaging using a pipette. Chambers for monitoring assembly in solution were constructed using silicone spacers (cut using a cork borer of known outer diameter) and borosilicate coverslips (22 mm x 22 mm) (Menzel Glaser, Germany).

2.8.1 Piranha cleaning coverslip surface

Piranha solution was prepared using equal parts 96% sulphuric acid (H_2SO_4, M_w = 98.08 g mol^{-1}) (Acros Organics, Belgium) and 30% hydrogen peroxide (H_2O_2, M_w = 34.01 g mol^{-1}) (Acros Organics, Germany) in a fumehood. 22 mm x 22 mm coverslips purchased from Fisher Scientific (UK) were placed in piranha solution for 20 minutes. The coverslips were then rinsed by placing them in ultrapure water for 20 minutes, then transferring to fresh ultrapure water for a further 20 minutes and then rinsed under a stream of ultrapure water for 30 seconds on each side. Coverslips were air dried overnight before use.

2.8.2 Poly-D-lysine coated coverslips

Dry coverslips cleaned with piranha solution were coated with a dilute (15 ng ml^{-1}) solution of poly-D-lysine (PDL) (Sigma Aldrich, USA) by pipetting 300 µl PDL onto a glass coverslip (typically 22 mm x 22 mm) and incubating at room temperature for 1 hour. Following this, the PDL solution was aspirated and the coated side was rinsed under a stream of ultrapure water for 30 seconds. Rinsed coverslips were air-dried fully before use.

2.8.3 PEGylating coverslips

Coverslips were PEGylated using a modified version of a published protocol (Chandradoss, Haagsma et al. 2014). Coverslips were cleaned with piranha solution. Following cleaning, the coverslips were PEGylated as per the protocol. Methoxypolyethylene glycol succinimidyl carbonate ester (mPEG-NHS(SC)) 5,000 and biotin PEG NHS 5,000 were purchased from Nanocs (USA). $MS(PEG)_4$ was dissolved in DMSO and purchased from ThermoScientific (USA). 3-aminopropyl trimethoxysilane (APTES) was purchased from Sigma Aldrich (Belgium). Sodium bicarbonate (NaHCO3, M_w = 84.01 g mol^{-1}) was purchased from Sigma Aldrich (Germany). The PEGylated coverslips were rinsed under a stream of ultrapure water for 30 seconds and air dried overnight before use.

2.8.4 Silanising coverslips

Dry coverslips cleaned with piranha solution were silanised using a protocol previously reported (Seed 2001). After cleaning with piranha solution, 200 µl of 99% dicholrodimethylsilane (Lancaster, England) was pipetted onto each coverslip.

Coverslips were then placed under vacuum in a dessicator for three hours. After this time, the silanised coverslips were washed thoroughly with detergent to remove residue, rinsed thoroughly with ultrapure water and air dried before use.

2.8.5 Design and construction of chamber for aggregation study

A frame was designed and drawn using SketchUp 2016 to house the chamber for monitoring aggregation. This was 3D printed using a Form 2 3D printer (Form Labs, USA) and a white photopolymer resin (which is rigid after curing).

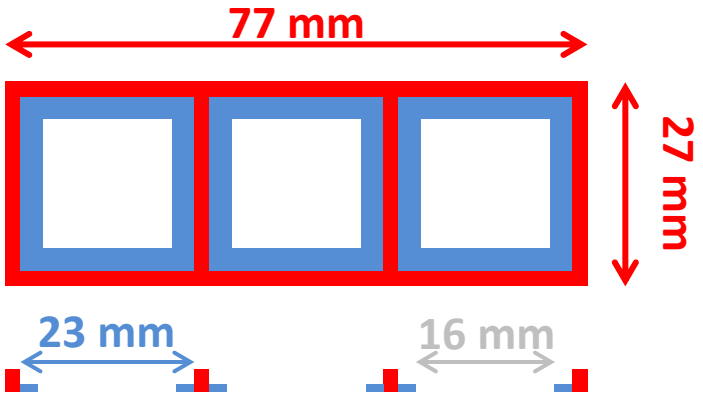

Figure 2.12: *Plan (top) and cross-section (bottom) view of 3D printed holder.*

Chambers for monitoring aggregation were assembled as follows:

1. A coverslip was placed in the frame. If a functionalised coverslip was used, the functionalised side was placed facing up.
2. A clean dry silicone spacer with a chamber of known dimensions (cut from silicone thermal pad) of desired thickness (0.5 mm or 1 mm) and chamber diameter (cut using a cork borer of known outer diameter) was placed on top and gently pressed to create a seal between the coverslip and spacer (figure 2.13A). If the frame was not used, a small piece of Blu Tack (a pressure sensitive adhesive putty material) was placed in each corner of the coverslip to ensure that there was no slippage.
3. The protein solution was added and another coverslip was placed on top (functionalised side facing down) and pressed gently to seal. A chamber to record temperature on the slide was assembled the same way but a narrow channel was cut in the silicone spacer first to accommodate a thermocouple probe (figure 2.13B).

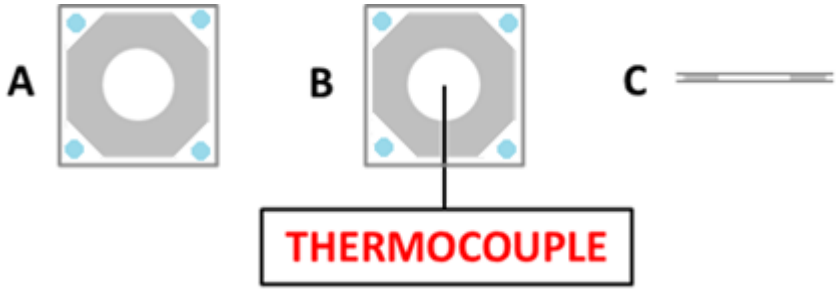

Figure 2.13: *Plan view of chamber for aggregation study (A) and temperature recording (B) and cross-section view of chamber with a top and bottom coverslip (C).*

2.9 Mammalian cell culture

Mammalian cell culture was carried out under aseptic conditions using a SafeFAST Classic 212 biological safety cabinet (Faster S.r.l., Italy) and 70% ethanol as a disinfectant. HEK293T/17 cells (LGC, UK) were grown in a Memmert INCO 153 CO_2 incubator (Germany) under 5% CO_2 and 90% relative humidity at 37 °C.

2.9.1 Initiating cell culture from frozen stocks

A cryogenic vial containing frozen cells was thawed in a water bath at 37 °C for ~2 minutes before being transferred to a biological safety cabinet. The vial contents were transferred to a sterile 15 ml centrifuge tube containing 10 ml of pre-warmed complete growth medium (CGM). The tube was centrifuged at 125 *g* for 7 minutes to pellet cells. The supernatant was removed and cells were re-suspended in 7 ml pre-warmed CGM before transferring to a T25 cell culture flask. CGM was replaced every 2–3 days with fresh CGM until cells reached ~90% confluency, after which the cells were subcultured (or passaged) (section 2.9.2).

2.9.2 Subculturing cells

Cells were subcultured (or passaged) when ~90% confluency was reached. The following protocol applies to a culture grown in a T25 cell culture flask. CGM was removed from the flask and cells were washed with 7 ml pre-warmed 1X PBS. After washing, the PBS was discarded and 3 ml 1X trypsin-EDTA solution (section 2.1.4.3) was added to the flask to detach cells. The flask was incubated at 37 °C, 5% CO_2 and 90% relative humidity for up to 10 minutes. After this time, 7 ml CGM was added to the flask. Cells were harvested by gently aspirating medium to dislodge cells from the wall of the flask. Cells were added to a sterile 15 ml centrifuge tube

and centrifuged at 125 *g* for 10 minutes. The supernatant was discarded and cells were re-suspended in CGM before counting and calculating cell viability. Cells were seeded at an appropriate concentration, typically 2 x 10^5 cells ml^{-1}.

2.9.3 Cell counting and cell viability determination

Cells were counted using a Bright-Line hemacytometer (Hausser Scientific, USA). Cell viability was assessed using Trypan Blue 0.4 % solution (Sigma Aldrich, USA). Viable cells do not absorb Trypan Blue and appear white. Dead cells appear blue as the dye can traverse the cell membrane.

2.9.4 Cryopreservation of cells

Cell stocks (1 ml aliquots) were prepared by seeding cells in T75 cell culture flasks (Corning, USA) and harvesting. Following centrifugation, cells were re-suspended in cryopreservation medium, aliquoted in Nalgene cryogenic vials (ThermoScientific, USA) and frozen at a density of 3 x 10^6 cells ml^{-1}. The vials were frozen in a stepwise manner by cooling to 4 °C for 5 minutes, then to-20 °C for 1 hour followed by -80 °C for 1 hour before placing in a liquid nitrogen cryopreservation tank for long-term storage.

2.9.5 Protein expression in HEK293T/17 cell line

Emerald Green Fluorescent Protein (EmGFP), with a blasticidin resistant gene, was used as an expression control vector in HEK293T/17 mammalian cells. A mammalian expression vector containing the sequence of a fusion protein, EmGFP-HGD, was designed and the vector was synthesized by GeneArt (ThermoFisher, GeneArt Division, Germany). The mammalian vector pc-DNA6.2_C-EmGFP-DEST was used and the HGD gene was inserted at the C-terminus. A double mutant of EmGFP-HGD (P23TR36S) was obtained by site directed mutagenesis of the HGD mammalian vector and expressed in the HEK 293T/17 cell line. Insertion of the correct mutation was confirmed by DNA sequencing.

The insertion of genetic material into a mammalian cell by non-viral methods is known as transfection. HEK293T/17 mammalian cells were transfected with EmGFP-P23TR36S. Low passage cells (< 30 passages) that had not exceeded 90% confluency at any passage were transfected. Lipofectamine 2000 transfection reagent (Invitrogen, USA) is a cationic liposome based reagent and was used for transfecting

HEK293T/17 cells with the desired plasmid DNA. Nucleic acids are negatively charged at physiological pH and complex with the positively charged Lipofectamine 2000 reagent, thus allowing uptake by the negatively charged cell membrane by overcoming electrostatic repulsion (Dalby, Cates et al. 2004).

Cells for transfection were grown on cell culture coverslips that had either 4 or 8 removable wells (Eppendorf, Germany). Coverslips were coated with PDL by adding either 0.5 ml (4 well) or 0.25 ml (8 well) PDL at 0.1 mg ml^{-1} to each well and incubating at room temperature for 2 hours. Wells were rinsed with sterile ultrapure water following coating and dried for at least 2 hours before introducing cells. Once a confluency of ~70% was reached, CGM in the wells was replaced with antibiotic and serum-free DMEM prior to transfection. Lipofectamine 2000 (5 µl Lipofectamine 2000 per 100 µl) and plasmid DNA (2.5 µg plasmid DNA per 100 µl) were separately mixed with serum-free DMEM and incubated at room temperature for 5 minutes. The DNA and Lipofectamine 2000 were then mixed to form the DNA-liposome complex and incubated at room temperature for 20 minutes. 48 µl (for 4 well) or 24 µl (for 8 well) of the DNA-liposome complex were added to the chambers containing the cells. After 4 hours, the medium containing the complex was exchanged for CGM. CGM was replenished every 2–3 days.

2.10 Microscopy

2.10.1 Background

Microscopy is an invaluable tool for visualising objects at length-scales far beyond those perceptible by the human eye. There are three main branches of microscopy providing different levels of resolution: Optical microscopy (which uses transmitted light through a lens as a means of magnification), electron microscopy (which uses a beam of electrons to visualise specimens) and scanning probe microscopy (which uses a probe to scan the surface of the object being looked at).

Optical microscopy has many distinct classes, such as bright field microscopy, phase contrast microscopy, polarization microscopy and fluorescence microscopy, all of which involve the visualisation of a specimen using transmitted light. Bright field microscopy is the simplest of these techniques and allows visualisation of a specimen simply by illumination. Phase contrast microscopy takes advantage of the difference

between the refractive indices of different visualised entities to provide higher contrast images than those generated using bright field microscopy. Fluorescence microscopy allows visualisation of specimens that are either intrinsically fluorescent or have a fluorescent molecule attached onto them. Fluorescence involves the absorption of light by an atom or molecule at one wavelength and emission of light at a longer wavelength (Herman 1998). A fluorescent microscope is designed to emit and detect photons of light for the purpose of providing spatial information to the user. Polarization microscopy is mainly used as a qualitative means of assessing the physical properties of a specimen and involves the transmission of plane polarized light through a birefringent material.

2.10.2 Protein sample preparation and imaging

The chamber containing the protein solution for monitoring aggregation (section 2.8) was heated to 30 °C using a heating stage and the temperature was monitored in a control chamber using an Omega HH509R thermocouple. The time at which the temperature of the chamber had equilibrated to the set temperature was taken as T_0. Aggregates were monitored and imaged using phase contrast and polarisation microscopy using an Olympus BX61 upright microscope using 100X, 20X and 10X objective lenses. Images were analysed using Cell^F software.

2.10.3 Preparation of transfected HEK293T/17 mammalian cells for imaging

A set of transfected cells on PDL coated cell imaging coverslips (section 2.9.5) were imaged every 24 hours following transfection for a total of 96 hours. The medium was aspirated from the wells and the well removed before imaging. A sterile microscope slide was placed on top of the coverslip. Cells were imaged on a heating stage at 37 °C using an Olympus BX61 microscope and 100X oil-immersion or 10X objective lenses. Fluorescence microscopy (using an Olympus FITC fluorescent filter) and phase contrast microscopy were used to image cells. Images were recorded and processed using Cell^F imaging software.

2.11 Model and simulation details

A model was designed specifically to numerically describe the effects of chemically modifying HGD with a small molecule fluorescent label by Nicoletta Gnan,

Emanuela Zaccarelli and colleagues at the University of Rome, La Sapienza. This work was completed as part of a collaboration.. The model and simulation details outlined below is the **exclusive** work of our collaborators and are reproduced here from a publication that arose from this collaboration, which helps to explain the experimental work described in chapter 3 (Quinn, Gnan et al. 2015).

2.11.1 Model for the unlabelled proteins

Unlabelled HGD ('U-type particles') are modelled as patchy particles complemented with a square well (SW) potential. This is based on previously published models (Kern and Frenkel 2003; Liu, Kumar et al. 2007) that produce a coexistence curve with the same width as a curve experimentally produced for bovine gamma crystallin. Unlabelled proteins are represented as hard spheres of diameter σ_m interacting *via* a SW attraction of width δ_{SW} and depth ε_{SW}:

$$V_{SW}(|\vec{r}_{ij}|) = \begin{cases} \infty \; if \; |\vec{r}_{ij}| \le \sigma_m, \\ -\varepsilon_{SW} \; if \; \sigma_m \le |\vec{r}_{ij}| \le \sigma_m + \delta_{SW}\sigma_m, \\ 0 \; otherwise. \end{cases} \qquad \textbf{2.24}$$

where $|\vec{r}_{ij}|$ is the vector between the centres of particles *i* and *j*. The surface of the patchy particle has four non-overlapping randomly placed attractive patches ('U-patch'). This differs from the model outlined in Liu, Kumar et al. 2007 where the patches were arranged in a tetrahedral geometry on the surface of the particle. The resulting interaction potential between the patches is the product of the radial contribution modulated by an angular function:

$$V_{ij}^{U} = V_{SW,U}\left(|\hat{r}_{ij}|\right) G(\hat{r}_{ij}, \hat{r}_{i\alpha}, \hat{r}_{j\beta}) \qquad \textbf{2.25}$$

where $\hat{r}_{i\alpha}$ and $\hat{r}_{j\beta}$ are the unit vectors from the centre of particle *i(j)* to the centre of the *α(β)* patch on the surface and the square well potential $V_{SW,U}$ has a width of $\delta_{U,U}$ and depth $\varepsilon_{U,U}$. The potential is modulated by function *G* and depends on the reciprocal orientation of the two particles:

$$G(\hat{r}_{ij},\hat{r}_{i\alpha},\hat{r}_{j\beta}) = \begin{cases} 1\, if \begin{cases} \hat{r}_{ij}\cdot r_{i\alpha} > \cos(\theta_{\max})_U, \\ -\hat{r}_{ij}\cdot r_{i\beta} > \cos(\theta_{\max})_U, \end{cases} \\ 0\, otherwise. \end{cases} \quad \textbf{2.26}$$

Length and energy are given units of σ_m and ε_{SW} respectively. Temperature (T) is measured in units of ε_{SW}. Parameters for U-type particles are given in table 2.12 and are based on a previous model (Liu, Kumar et al. 2007). This model produces a curve of the correct width and critical volume fraction ($\phi_c = 0.21$) for HGD.

Table 2.12: *Parameters of the interaction potential for unlabelled HGD, as given by Liu, Kumar et al. 2007.*

$\delta_{U,U}$	$\varepsilon_{U,U}$	$\cos(\theta_{max})$	δ_{SW}	ε_{SW}
0.05	5	0.95	0.5	1.0

2.11.2 Model for the labelled proteins

The small molecule fluorescent dye is modelled (L-particle) as having a fifth larger and more attractive patch (L-patch) to the U-type particle. The fifth patch is modelled as being attached to the U-type particle, as opposed to being a composite part of its surface, and thus can extend beyond the surface and, therefore, form bonds with more than one other particle. The fifth patch is forbidden from forming bonds with another L-patch.

2.11.3 Evaluation of the critical points

Umbrella sampling grand canonical Monte Carlo simulations (US-GCMC) and histogram reweighting were performed for accurate critical point and liquid-liquid coexistence curve evaluation of U-type particles (Ferrenberg and Swendsen 1988; Virnau and Müller 2004). At the critical point, the order parameter M has a probability distribution $P(M)$ that follows the one typical of the Ising universality class (Nicolaides and Bruce 1988; Wilding 1997; Tsypin and Blöte 2000; Romano, Tartaglia et al. 2007). The parameter $M = \rho + su$ is the linear combination of the number density ρ and the energy density u through the field mixing parameter s (Wilding 1997; Romano, Tartaglia et al. 2007; Rovigatti, Heras et al. 2013).

The critical point is located by tuning the temperature T and the chemical potential μ until the numerical joint distribution $P(N, E: T, \mu)$ projected over N shows the double-peaked shape that is typical of the Ising probability distribution $P(M)$ which is indicative of the presence of large density fluctuations in the vicinity of the critical point within the system. A fitting procedure based on the histogram reweighting technique was implemented to transform the joint distribution $P(N, E; T, \mu)$, evaluated at T and μ, into $P(N, E; T', \mu')$, evaluated at T' and μ' to extract the best values for T', μ' and s in order to find the $P(M)$ that was closest to the Ising distribution (Ferrenberg and Swendsen 1988; Wilding 1997). This reweighting technique is also used to evaluate liquid-liquid coexistence curve. For temperatures lower than T_c, liquid-liquid coexistence densities are calculated by imposing the equality of the areas below the two peaks for the joint distribution $P(N, E; T, \mu)$ projected over N and centred on $N_c = \rho_c L^3$ where ρ_c is the critical density and L is the length of the simulation box.

The procedure for critical point evaluation in the binary mixture is similar to that for the pure system except a different order parameter is required. In this case, $M = \rho_U + \alpha\rho_L$ where $\rho_{U(L)}$ is the number density of the species U(L) and α is a mixing parameter (Rovigatti, Heras et al. 2013). As a result, the numerical joint distribution $P(N_U, N_L, \mu_U, \mu_L)$ depends solely on the number of particles of the two species and their associated chemical potentials. Since there is no energy dependence, T cannot be varied when locating the critical points and so the critical temperature T_c is searched for for each μ_U and μ_L for a given concentration $x_{L(U)}$ to locate T_c. Thus, the specific T_c corresponding to the desired concentration of L-type particles is selected for comparison with the experimental amine labelled protein compositions.

Chapter 3:
How fluorescent labelling alters the solution behaviour of proteins

3.1 Introduction

3.1.1 Background

Chemical modification of proteins is a useful mechanism to label a protein for detection or to modify their behaviour and is currently used in a range of applications such as the modification of therapeutic biomolecules for improved stability and/or function, or in the production of novel materials in synthetic biology (Veronese 2001; Walsh and Jefferis 2006; Sahoo 2012; Spicer and Davis 2014; Boutureira and Bernardes 2015). Fluorescent labelling of proteins is one of the most useful forms of protein chemical modification as it has allowed for the direct *in situ* visualisation of proteins. Advances in the development of specialised imaging techniques have led to a marked increase in resolution with microscopy and single molecule detection being used for studies involving protein distribution, translocation and interactions both *in vivo* and *in vitro* in addition to its use as a means to screen for the formation of protein crystals (Kasten 1999; Hillger, Nettels et al. 2007; Huang, Bates et al. 2009; Pusey, Barcena et al. 2015). Fluorescent labelling of proteins, coupled with either spectroscopy or microscopy can be used to understand the mechanisms underlying many biological processes such as inter/intra-cellular communication, genomics, protein condensation diseases and protein self-association during biotherapeutic production (Zhang, Campbell et al. 2002; Hillger, Nettels et al. 2007; Hawe, Sutter et al. 2008; Sahoo 2012).

There are three main established strategies used to label fluorescent proteins: small molecule organic fluorescent labels that covalently attach to the protein of interest, fluorescent fusion proteins (such as GFP) that are co-expressed or fused to the protein of interest and quantum dots which are functionalised in order to form a biotin-avidin pair with the protein of interest (Tsien 1998; Zheng, Juette et al. 2014). Of these three established methods, fluorescent labelling *via* covalent attachment of a small molecule fluorescent dye is considered to be the most versatile (Zheng, Juette et al. 2014) given the relatively small sizes of the molecules used and well established labelling chemistries.

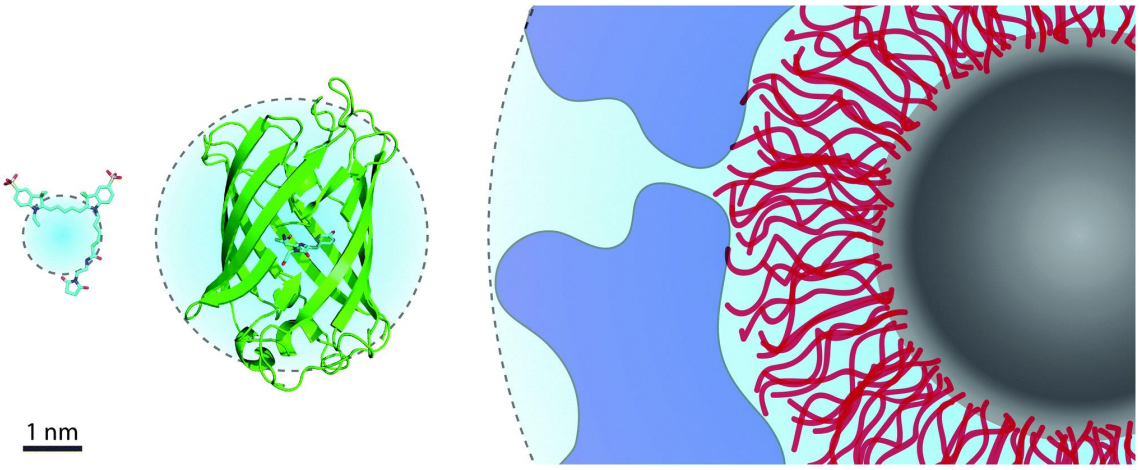

Figure 3.1: *Relative size of a small molecule organic fluorescent dye (left), GFP (centre) and quantum dot (right) (taken from Zheng, Juette et al. 2014).*

Despite the advances being made with fluorescent imaging, little consideration has been given to the impact a fluorescent label has on protein solution behaviour. Yet, this is an important consideration when interpreting data obtained using labelled protein. Protein behaviour in solution is sensitive to changes in its environment, such as pH, and can also be considerably altered by changes to the surface of the protein, such as mutagenesis and specific binding of ions (Asherie 2004; Roosen-Runge, Zhang et al. 2014; James, Quinn et al. 2015; Li, Persson et al. 2015). It is therefore reasonable to expect some impact on protein solution behaviour due to a surface modification using a fluorescent label. Some studies have reported impacts that fluorescent labels can have on protein characteristics; FITC labelled BSA has been shown to possess an increased diffusion coefficient in addition to increased levels of irreversibly adsorbed proteins at the oil-water interface (Gajraj and Ofoli 2000). Chemically modified GFP exhibited a more acidic pI relative to its unlabelled counterpart (Richards, Stathakis et al. 1999). The varying effects of using different small molecule fluorophores on cellular function and adherence of leucocytes to endothelial cells has been investigated (De Clerck, Bridts et al. 1994). A study using monomerically labelled and mixtures of labelled and unlabelled group specific antigen (or 'Gag', a viral protein) showed the production of morphologically indistinguishable clusters at a lipid membrane (Gunzenhäuser, Wyss et al. 2014).

Macromolecular interactions, including colloid and protein interactions, can be characterised using phase diagrams (Asherie, Lomakin et al. 1996; Sear 1999;

Asherie 2004; Möller, Grobelny et al. 2014). The net inter-protein interaction potential arises from both attractive and repulsive contributions from surface-exposed amino acid side-chains (Lomakin, Asherie et al. 1999; James and McManus 2012). Proteins are inherently anisotropic due to the variation in the side-chain chemistries of each surface-exposed amino acids and this anisotropy has been shown to have a direct impact on protein phase behaviour (Asherie, Lomakin et al. 1996; Bonneté, Finet et al. 1999; McManus, Lomakin et al. 2007; James, Quinn et al. 2015). Liquid-liquid phase separation (LLPS) of a protein in a solution occurs for proteins interacting *via* short-ranged attractive interactions (Asherie, Lomakin et al. 1996; Asherie 2004; Möller, Grobelny et al. 2014). The temperature at which LLPS occurs is sensitive to the strength of the net inter-protein attraction and can be used as a means of probing relative changes in inter-protein interactions as a result of altered protein chemistry, the introduction of a second protein type or changes in solution conditions (Broide, Tominc et al. 1996; Wang, Lomakin et al. 2010). As a result, LLPS is a phase transition that can be readily used as a means of assessing how fluorescent labelling affects the phase behaviour of a labelled protein relative to its unlabelled form.

3.1.2 Aim of study

The first aim of this study was to fluorescently label HGD protein using commercially available small molecule fluorescent labels to quantify the changes in phase behaviour of HGD after the addition of small amounts (≤ 1% total protein) of labelled protein. This was done by fluorescently labelling at two distinct sites (Lys-2 and Cys-110, figure 3.2) on the protein surface (i.e. only one position at a time) in addition to using a dye with less labelling specificity. The changes in phase behaviour due to labelling were quantified by measuring the change in liquid-liquid phase separation temperature (T_{ph}) after the addition of labelled protein. The second aim of this study was to devise a model to numerically explain the observed experimental results.

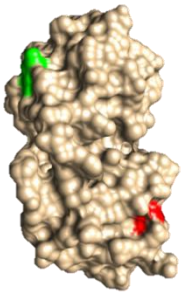

Figure 3.2: *Topological rendering of HGD showing Lys-2 (green) and Cys-110 (red).*

3.2 Results

HGD protein was produced recombinantly in *E. coli* and extracted and purified using SEC (figure 3.3) and IEX (figure 3.4). The purified protein was then characterised using SE-HPLC, SDS PAGE, in-tact mass spectrometry and circular dichroism spectroscopy.

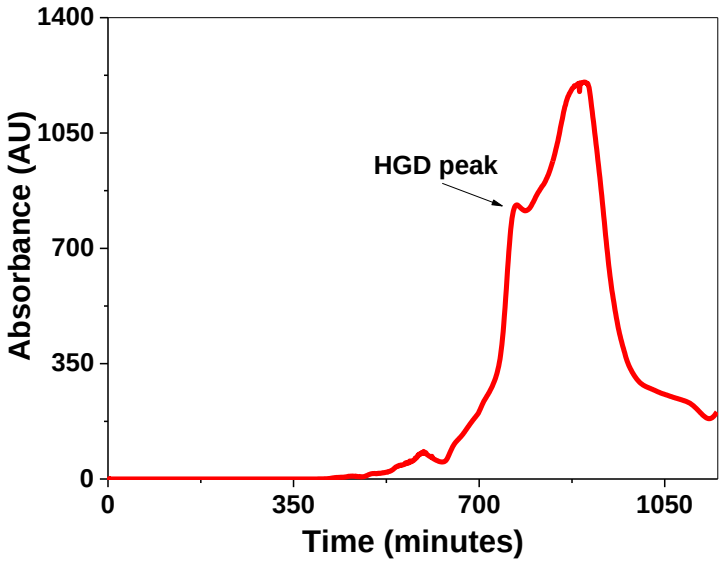

Figure 3.3: *Size exclusion chromatogram indicating the peak associated with HGD.*

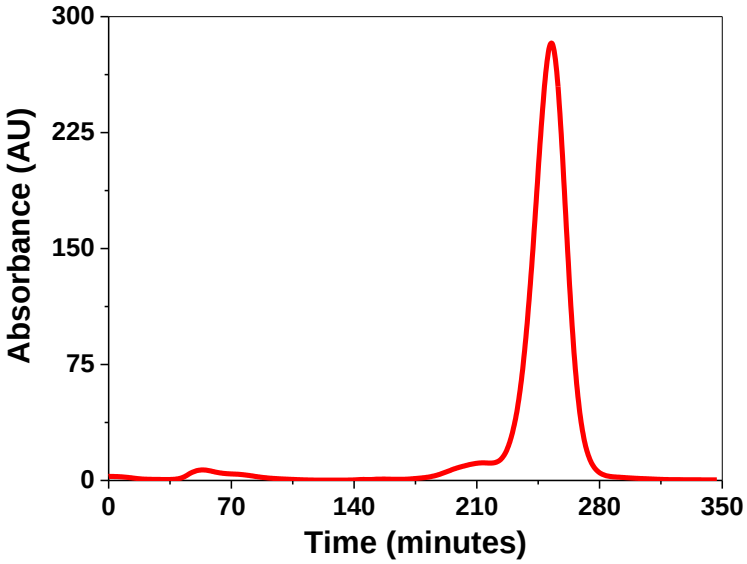

Figure 3.4: *Ion exchange chromatogram after purifying fractions from the size exclusion step. The major peak contains HGD protein. Following purification, only fractions from the peak tail were pooled to maximise protein purity.*

SE-HPLC (figure 3.5) and reducing SDS-PAGE (figure 3.6) were used to determine protein purity. SE-HPLC produces one peak at 24.138 minutes. Using the calibration curve for this column under the same buffer conditions (section 2.3.1.3.3.1) indicates a molecular weight of ~21 kDa, a value consistent with the molecular weight of a HGD monomer found using intact ESI-MS (20 608 ±1 Da).

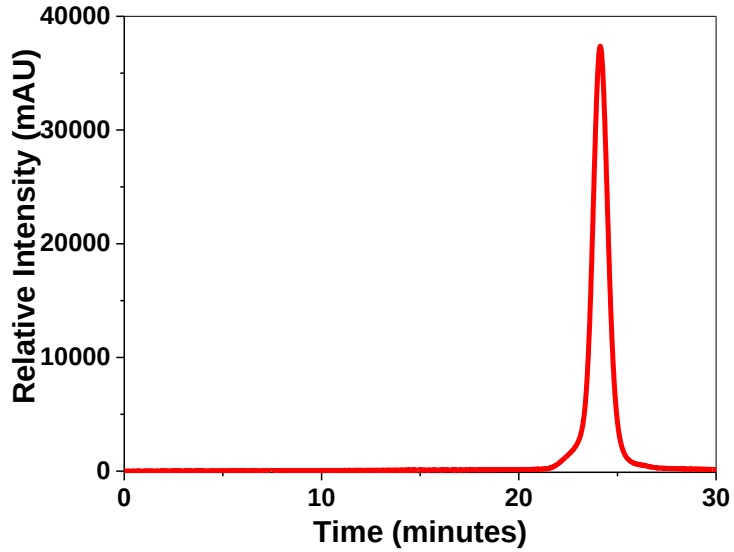

Figure3.5: *SE-HPLC chromatogram for HGD prepared in 0.1 M sodium phosphate at pH 7.0with 20 mM DTT.*

SDS-PAGE was also used to assess protein purity. A single clean band just below the 21.5 kDa molecular weight marker indicated pure monomer protein (figure 3.6).

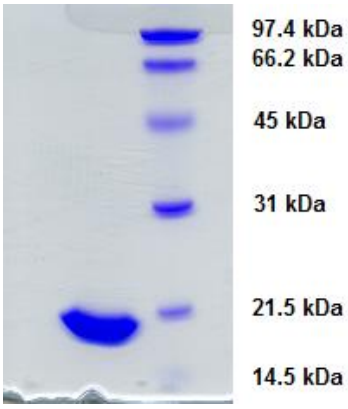

Figure 3.6: *12.6% reducing SDS-PAGE gel showing a single band indicating the purity of HGD protein.*

3.2.1 Lysine modification

HGD was labelled at the Lys-2 position using AnaTag HiLyte 405, a fluorescent dye with excitation and emission wavelengths of 404 and 428 nm respectively and a succinimidyl ester reactive moiety for primary amine specificity (section 2.5.1). In addition to the primary amine of the N-terminus group, HGD has only one other primary amine: a Lys in position 2 of the amino acid sequence. Careful pH control during labelling ensured that the ε-amino group (a primary amine) of Lys-2 was preferentially labelled since the pK_a value of the α-amino group is different from that of the ε-amino group (Sélo, Négroni et al. 1996). This allowed for unambiguous labelling efficiency determination since a maximum of one dye molecule per protein was possible. After labelling, the modified HGD was used in mixtures with unmodified HGD at varying labelled protein fractions and the liquid-liquid coexistence curve for each composition was measured. Two spectroscopic techniques, circular dichroism (CD) and second derivative UV absorbance spectroscopy, were used to analyse if a structural change to the protein occurred as a result of fluorescent labelling.

3.2.1.1 Phase diagram measurement

Liquid-liquid phase boundaries were determined for mixtures of unlabelled HGD and amine modified HGD at varying labelled protein compositions ranging from x_L =

0.002 – 0.01, where x_L is the fraction of labelled protein (figure 3.7). The error bars shown indicate the temperatures recorded for T_{cloud} and T_{clear}. Fluorescently labelled proteins are generally used at low concentrations given the highly sensitive nature of fluorescence detection. Thus, the range of labelled protein compositions used here are comparable to those used in applications where fluorescent labelling is required.

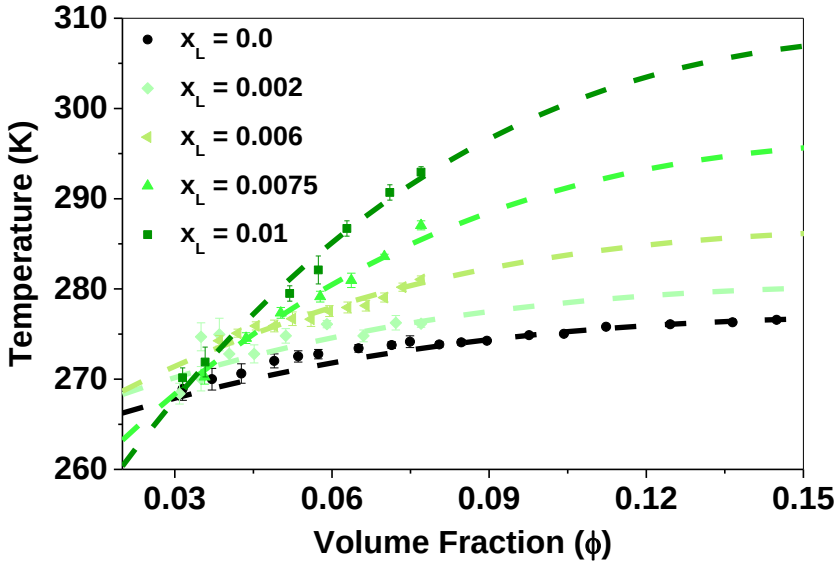

Figure 3.7: *Liquid-liquid coexistence curves for unlabelled HGD (black) and HGD fluorescently labelled at Lys-2 with labelled protein fractions of x_L = 0.002 – 0.01 (green). Values for data points shown given in tables A1 and A2 in Appendix.*

Each liquid-liquid coexistence curve was fitted using the below equation (3.1):

$$\left(\frac{\phi-\phi_c}{\phi_c}\right) = A\left(1-\frac{T}{T_c}\right)^{\beta} \qquad \textbf{3.1}$$

where ϕ and ϕ_c represent the volume fraction and the critical volume fraction respectively, A is a parameter that determines the width of the coexistence curve, T and T_c are the temperature and critical temperature respectively (measured in Kelvin) and β is an exponent term for the three dimensional Ising model equal to 0.325 (Broide, Berland et al. 1991). While ϕ_c was fixed in the estimation of T_c using equation 3.1, parameter A was varied. Values for T_c, ΔT_c and A for each labelled protein composition are outlined in table 3.1.

Table 3.1: *Estimated critical temperature (T_c), change in critical temperature (ΔT_c) and width parameter (A) for each composition of HGD amine modified with HiLyte Fluor 405 at ϕ_c = 0.21.*

Fluorescently Labelled Protein Fraction	T_c (K)	ΔT_c (K)	A
x_L = 0.00	277	-	2.6
x_L = 0.002	280	3	2.5
x_L = 0.006	287	10	2.2
x_L = 0.0075	297	20	1.8
x_L = 0.01	308	31	1.6

3.2.1.2 Structural analysis

Two complementary spectroscopic analytical methods, circular dichroism (CD) spectroscopy and second derivative UV absorbance spectroscopy, were used to assess if there was a structural change in the protein as a result of fluorescent labelling. Since unfolding can lead to exposure of hydrophobic amino acids in the protein's interior and, consequently, an increase in the net attraction of the protein, it was essential to determine if a structural change was the cause of the increase in T_{ph}.

3.2.1.2.1 Circular dichroism spectroscopy

Near- and far-UV CD spectroscopy were used to compare the tertiary and secondary structures of unlabelled HGD and fluorescently labelled HGD at the Lys-2 position.

The spectroscopic data for near-UV CD (λ = 260 – 320 nm) arise as result of the aromatic amino acids, with Phe having a wavelength range of ~ 260 – 265 nm, Tyr having a wavelength range of ~275 – 285 nm and Tyr having a wavelength range of 285 – 305 nm, although overlap between Tyr and Trp signals can occur (Kelly, Jess et al. 2005). Figure 3.8 shows the near-UV spectroscopic data for unlabelled (black) and amine modified (green) HGD. While the spectroscopic data for the wavelength ranges corresponding to Tyr and Trp are shown to be largely similar, there is a perturbation in the data for the wavelength region corresponding to Phe.

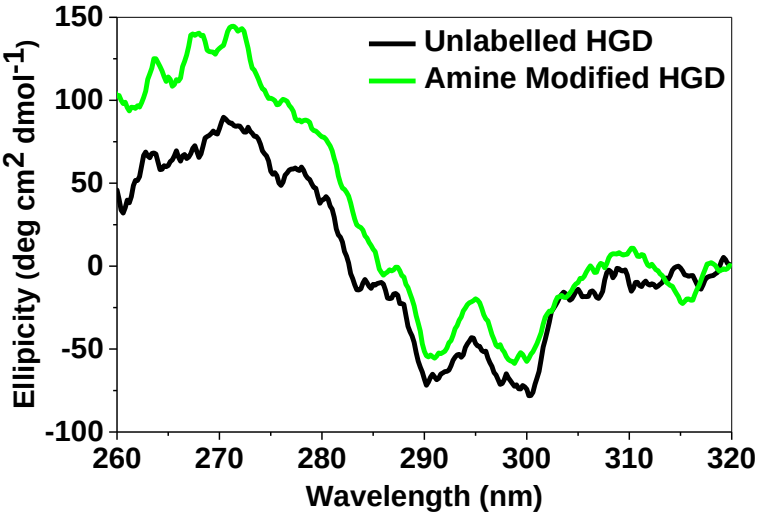

Figure 3.8: *Near-UV CD spectra for unlabelled HGD (black) and amine modified HGD (green). There is a slight perturbation in the spectroscopic data for the wavelength range corresponding to Phe, indicating a localised change to the tertiary structure.*

Using Chimera software, it is possible to look at the location of amino acid residues found on the protein surface. On the surface of HGD, there is one Phe residue in proximity to Lys-2 in position 11 (figure. 3.9). An interaction between the fluorescent dye attached to Lys-2 and Phe at position 11 could lead to the largely localised change in the near-UV CD spectroscopic data of the amine modified HGD relative to the unlabelled HGD.

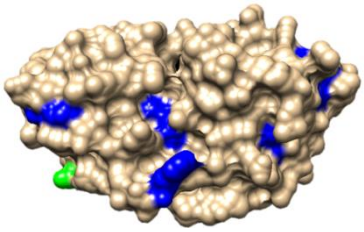

Figure 3.9: *Topological rendering of HGD showing Lys-2 in green and relative positions of Phe residues (blue).*

The far-UV CD spectroscopic data (for wavelengths of λ = 240 nm and below) arise due to the absorption of the peptide bond between proteins and is used to provide secondary structure information. From the data in figure 3.10, the secondary structures of unlabelled HGD and amine modified HGD are very similar.

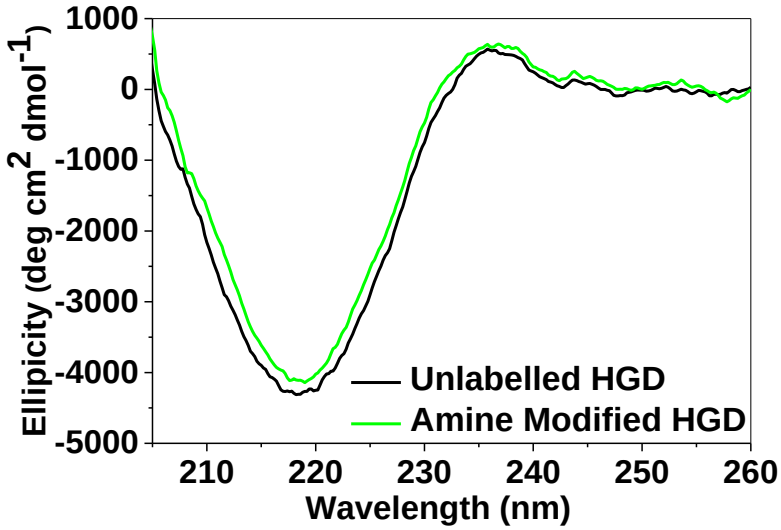

Figure 3.10: *Far UV CD spectra for unlabelled HGD (black) and amine modified HGD (green) indicating very similar secondary structures.*

3.2.1.2.2 Second derivative UV absorbance spectroscopy

As a complement to near-UV CD spectroscopic analysis of the unlabelled and amine modified HGD proteins, the second derivative of the UV-Vis absorption spectrum was analysed (figure. 3.11). The data indicate that a small change in the second derivate spectrum in the wavelength region corresponding to Phe (λ = 245 – 270 nm) can be seen. This observation is very consistent with the near UV CD data in the same wavelength region. The data for the wavelength regions corresponding to Trp (λ = 265 –295 nm) and Tyr (λ = 265 – 285 nm) remains largely unchanged.

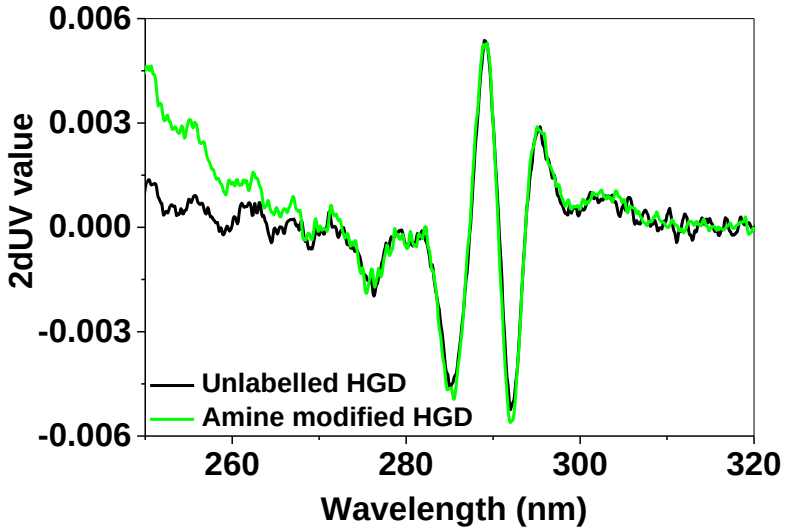

Figure 3.11: *Second derivative UV analysis of unlabelled HGD (black) and amine modified HGD (green). A perturbation exists in the wavelength region corresponding to Phe. This is a localised change given that the spectral data for higher wavelengths are very similar.*

3.2.2 Thiol modification

HGD was also labelled at the Cys-110 position (section 2.5.2) using DyLight 405 Maleimide, a fluorescent dye with excitation and emission wavelengths of 400 and 420 nm respectively. These values are comparable to the excitation and emission wavelengths of the fluorescent dye used for amine modification of HGD at Lys-2. The maleimide reactive moiety ensures reduced thiol labelling specificity and reacts with a free thiol to form a stable thioether bond. Although HGD has six unpaired Cys residues, only one (in position 110) is surface exposed (Pande, Pande et al. 2000). Again, the sole availability of a single amino acid residue for labelling ensured unambiguous labelling efficiency determination since a maximum of one dye molecule per protein was possible. The labelled HGD was used in mixtures with unlabelled HGD at varying labelled protein compositions and the liquid-liquid coexistence curve for each composition was measured. Two spectroscopic techniques, CD spectroscopy and second derivative UV absorbance spectroscopy, were again used to analyse if a structural change to the protein occurred after fluorescent labelling.

3.2.2.1 Phase diagram measurement

Liquid-liquid phase boundaries were determined for mixtures of unlabelled HGD and thiol modified HGD at varying labelled protein compositions ranging from x_L = 0.0001 – 0.0002, where x_L is the fraction of labelled protein (figure 3.12). T_{ph} was measured as outlined in section 2.7.2 In this case, significantly lower volume fractions of labelled protein were used, since attempts to add higher concentrations led to precipitation of the protein.

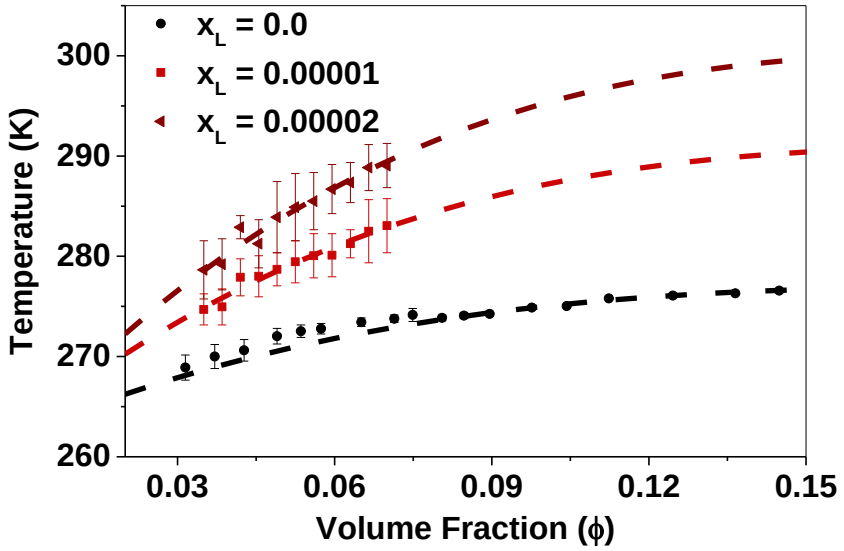

Figure 3.12 *Liquid-liquid coexistence curves for unlabelled HGD (black) and HGD fluorescently labelled at Cys-110 with labelled protein fractions of x_L = 0.0001 and 0.0002 (red). The dashed line is a fit of the data when the critical volume fraction is equal to 0.21. Values for data points shown given in tables A1 and A3 in Appendix.*

The liquid-liquid coexistence curves were fitted with equation 3.1 to estimate T_c for each labelled protein composition (table 3.3). Again, the value for ϕ_c (0.21) was fixed in the estimation of T_c for labelled proteins and parameter A was varied.

Table 3.3: *Estimated critical temperature (T_c), change in critical temperature (ΔT_c) and width parameter (A) for each composition of HGD thiol modified with DyLight 405Maleimide at ϕ_c = 0.21.*

Fluorescently Labelled Protein Fraction	T_c (K)	ΔT_c (K)	A
x_L = 0.0000	277	-	2.6
x_L = 0.0001	291	14	2.1
x_L = 0.0002	301	24	1.9

3.2.2.2 Structural analysis

Near- and far-UV CD spectroscopy and second derivative UV absorbance spectroscopy were used to determine if a structural change as a result of fluorescently labelling Cys-110 contributed to the increased net attraction measured experimentally.

3.2.2.2.1 Circular dichroism spectroscopy

Near- and far-UV CD spectroscopy were used for tertiary and secondary structure comparative analyses respectively. Figure 3.13 (near-UV) and figure 3.14 (far-UV) indicate no significant change in either the tertiary or secondary structures of the thiol labelled protein relative to the unlabelled protein and, hence, indicate that protein unfolding did not contribute to the increase in T_{ph}.

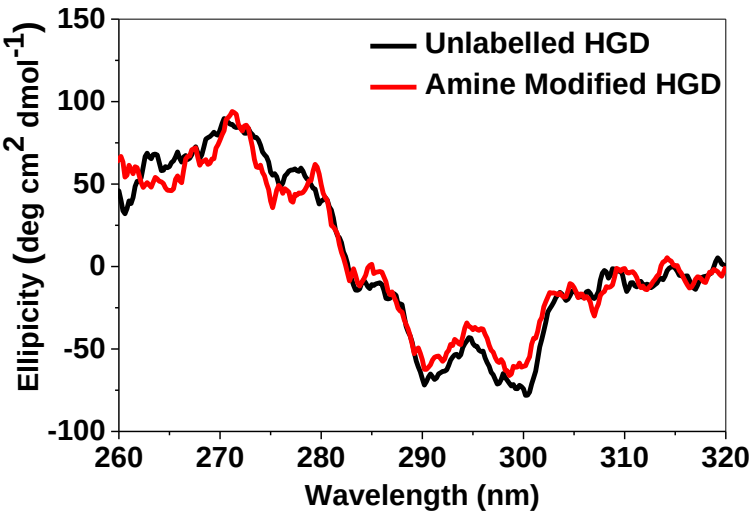

Figure 3.13: *Near-UV CD spectra for unlabelled HGD (black) and thiol modified HGD (red) indicating very similar tertiary structures.*

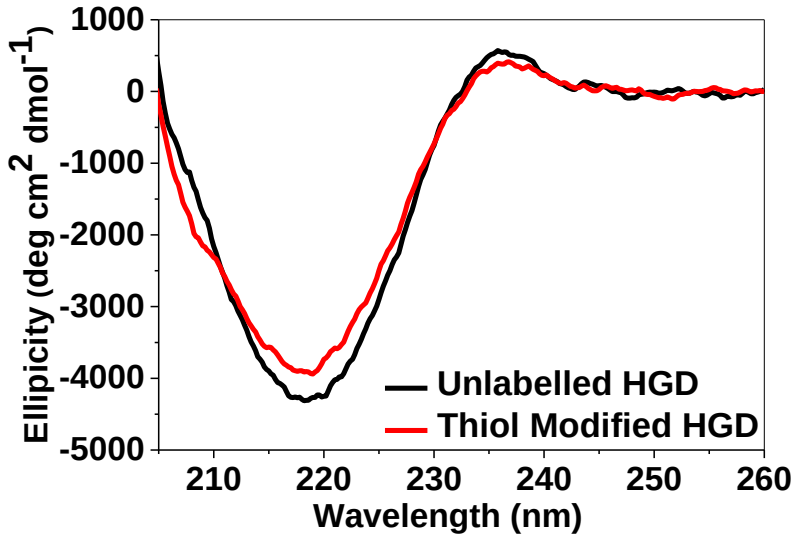

Figure 3.14: *Far-UV CD spectra for unlabelled HGD (black) and thiol modified HGD (red) indicating very similar secondary structures.*

3.2.2.2.2 Second derivative UV spectroscopic analysis

Second derivative UV spectroscopic analysis was again used as a complement to near-UV CD spectroscopy. The spectroscopic data (figure 3.15) also indicate no significant change in the data for thiol labelled protein relative to the unlabelled protein and, hence, indicate no significant change in the tertiary structure. These data are consistent with the data for near-UV CD which also indicated that there was no change in the tertiary structure of the protein.

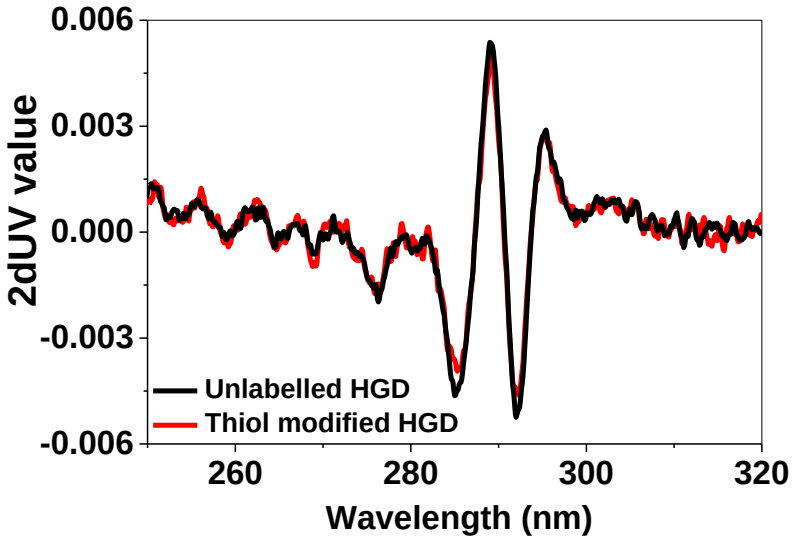

Figure 3.15: *Second derivative UV analysis of unlabelled HGD (black) and thiol modified HGD (red) indicating that the tertiary structures of the labelled and unlabelled protein are very similar.*

3.2.3 Size exclusion HPLC analysis

SE-HPLC was carried out on a mixture of unlabelled and amine modified HGD (x_L=0.01) under reducing conditions to determine if higher molecular weight oligomers were present since this can contribute an increase in T_{ph} and, thus, the net attraction of the protein (Pande, Lomakin et al. 1995; Asherie, Pande et al. 1998; Pande, Pande et al. 2000). Asherie, Pande et al. (1998) showed that T_{ph} at the critical point, increased from ~278 K for a monomer only solution of γ_{IIIb}-crystallin (bovine equivalent of HGD) to 305 K for a dimer only solution of γ_{IIIb}-crystallin. Pande, Pande et al. 2000 showed that a mixture monomer, dimer and higher molecular weight oligomers of the R14C single mutant of HGD also led to an increase in T_{ph} of ~20 K.

Figure 3.16 indicates that there are no detectable peaks at lower retention times thus indicating an absence of higher molecular weight oligomers in the solution containing the labelled protein. Unreduced HGD has a naturally occurring dimer, due to the formation of a disulphide bridge between two monomers that can easily be detected by HPLC at low concentration. Where a dimer between fluorescently labelled monomers occurs, we would expect that to be formed with a relatively large binding energy due to the hydrophobic nature of the label itself and the potential for $\pi - \pi$ interactions between the dye molecules to occur upon dimer formation. Hence, it should also be possible to detect protein dimers mediated by the fluorescent dye molecules attached to the protein by HPLC if present. A single peak for the fluorescently labelled protein is detected by HPLC, and when compared with the un-reduced un-labelled HGD, suggests that no detectable concentration of higher molecular weight oligomers are present in solution and the presence of low molecular weight oligomers are unlikely to be the source of the increase in T_{ph}.

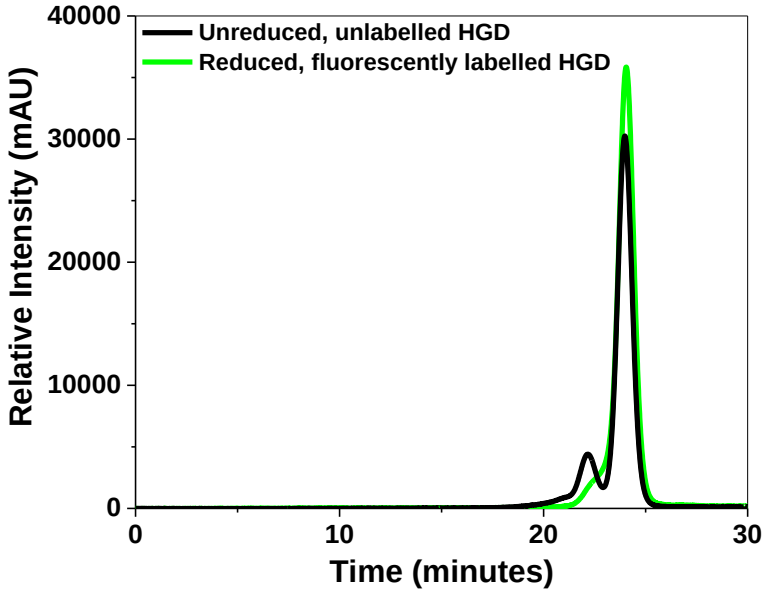

Figure 3.16: *SE-HPLC chromatogram for unreduced, unlabelled HGD (black) and a mixture of amine modified HGD and unlabelled HGD (x_L=0.01) (green).*

3.2.4 Fluorescently labelled protein partitioning

To determine if the fluorescently labelled protein preferentially partitioned into one phase over another, a solution of known volume (90 µl) containing a mixture of unlabelled HGD and HGD fluorescently labelled at the Lys-2 position at a known labelled protein fraction (x_L = 0.002) was cooled to and centrifuged at 0 °C to fully separate the dilute and dense phases (section 2.7.3). A fixed volume was carefully taken from each phase and diluted to measure the volume fraction and fluorescence intensity in each phase. Using these measurements, the dilute and dense phase protein volume fractions were calculated to be ϕ = 0.074 and ϕ = 0.35 respectively; these values were plotted on the liquid-liquid coexistence curve for HGD with a labelled protein fraction of x_L = 0.002. The labelled protein fraction after complete phase separation was calculated to be x_L = 0.0012 in the dilute phase and x_L = 0.0025 in the dense phase. This is consistent with previously published data where in a mixture of two proteins, the protein with the stronger net attraction preferentially partitions into the dense phase (Wang, Lomakin et al. 2010).

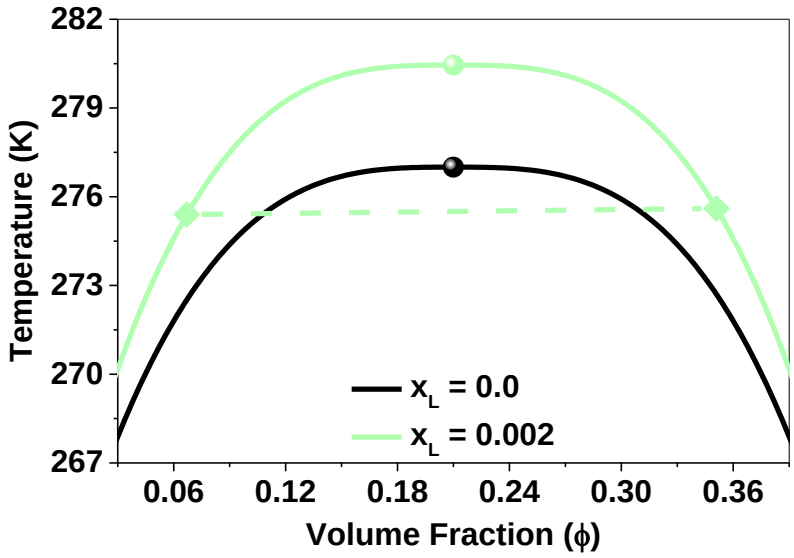

Figure 3.17: *Liquid-liquid coexistence curves for unlabelled HGD (black) and HGD fluorescently labelled at Lys-2 with labelled protein fraction of x_L = 0.002 (green). The diamonds indicate the volume fractions of the dilute and dense phases following separation and the tie-line connects these two points. The circles indicate the critical volume fraction for each protein composition.*

3.2.5 FITC labelling

3.2.5.1 Background

FITC is a widely used derivative of fluorescein with an isothiocyanate group that has excitation and emission peak maxima at 495 nm and 514 nm respectively. FITC does not possess the same degree of specificity as fluorophores with either a maleimide or NHS reactive moiety and will interact with various different side-chain groups. However, in general, only interactions with primary and secondary amines yield stable products. FITC labelling was used as a means of determining if the observations using site (primary amine and thiol) specific dyes were related to the specific chemistries of those dyes or if the increase in attraction is a more general phenomenon for widely used dyes.

3.2.5.2 Phase diagram measurement

A mixture of HGD and HGD fluorescently labelled with FITC in an aqueous solution (0.1 M sodium phosphate buffer at pH 7.0) was prepared as outlined in section 2.5.3 to a concentration of 100 mg ml^{-1} (ϕ = 0.071) to measure part of the ascending limb of its liquid-liquid coexistence curve (figure 3.18). The composition of HGD labelled

with FITC (x_L) used as a fraction of total protein was x_L = 0.0002. T_{ph} was measured and the coexistence curve fitted with equation 3.1. Estimated values for T_c, ΔT_c and A are given in table 3.3. Once again, a value of ϕ_c=0.21 was fixed.

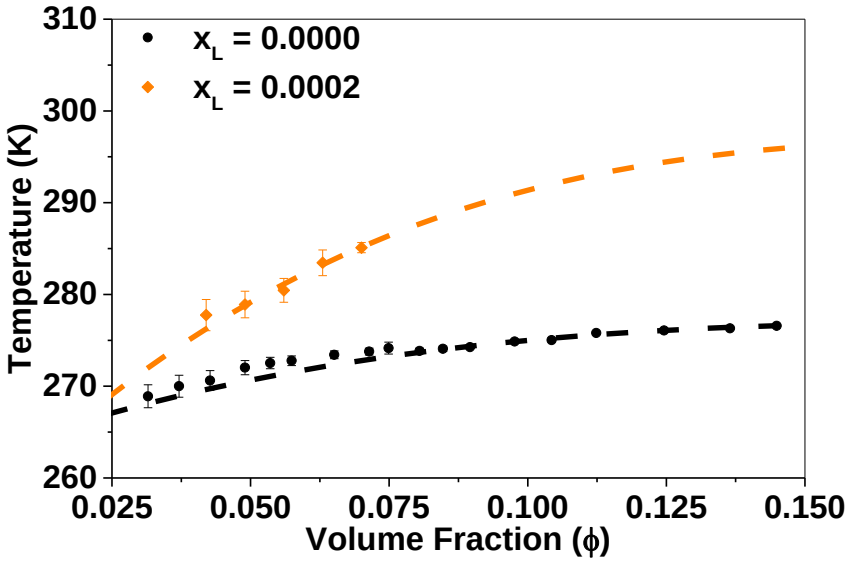

Figure 3.18: *Liquid-liquid coexistence curves for unlabelled HGD (circles) and HGD fluorescently labelled with FITC at a labelled protein fraction of x_L = 0.0002 (orange). Values for data points shown given in tables A1 and A4 in Appendix.*

A substantial increase in T_{ph} (~20 K) was observed under these conditions in addition to a narrowing of the liquid-liquid coexistence curve (as indicated by a decrease in parameter A). This is to be expected given the diminished labelling specificity associated with FITC and is consistent with the data for both Lys and Cys modifications of HGD whereby fluorescently labelling HGD led to an increase in the phase separation temperature of the protein mixture relative to the unlabelled as well as a decrease in the width of the coexistence curve.

Table 3.3: *Estimated critical temperature (T_c), change in critical temperature (ΔT_c) and width parameter (A) for each composition of HGD modified with FITC at ϕ_c = 0.21.*

Fluorescently Labelled Protein Fraction	T_c (K)	ΔT_c (K)	A
x_L = 0.0000	277	-	2.6
x_L = 0.0002	297	20	1.9

3.2.6 Modelling the phase behaviour of a fluorescently labelled protein in a mixture

A patchy particle model was designed specifically to numerically investigate if the experimental results, could be captured using a patchy particle model (as described in section 2.11) (Quinn, Gnan et al. 2015). This work was done by Nicoletta Gnan and Emanuela Zaccarelli and colleagues at the University of Rome, La Sapienza as part of a collaboration. The unlabelled protein (designated 'U particle') was based on a previous model used to describe the coexistence curve of bovine γ-crystallin using Kern-Frenkel patchy particles exhibiting a square-well (SW) attraction (Kern and Frenkel 2003; Liu, Kumar et al. 2007). The surface of the U particle has four randomly placed non-overlapping attractive sites ('U patch'); In using randomly placed attractive patches, it was possible to correctly reproduce the coexistence curve (i.e. correct width and ϕ_c) for HGD (Quinn, Gnan et al. 2015).

The labelled protein (designated 'L particle') was modelled as for the U particle with four attractive sites but had a fifth wider and more attractive site ('L patch') on its surface to mimic the fluorescent label. This was based on the assumption that a small molecule fluorescent label (e.g. FITC, figure 3.19 (left)) has a molecular weight that is larger than any of the aromatic amino acids (i.e. Phe, Trp and Tyr, figure 3.19 (right)) and is significantly more hydrophobic as a result of having several aromatic groups in its structure.

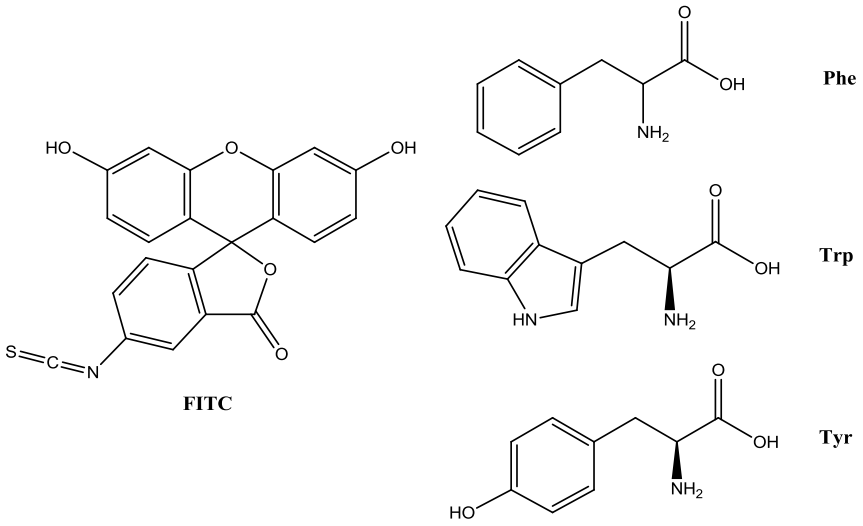

Figure 3.19: *Structure of FITC (left) and structures of the aromatic amino acid residues Phe, Trp and Tyr (right).*

Since the L patch is attached to the surface as opposed to being a composite part of the L particle structure, it extends beyond the surface and can readily interact with multiple U patches of both U and L particles (figure 3.20). The possibility of an interaction with another L patch was prohibited since this would lead to the formation of L particle clusters, an eventuality that was excluded experimentally (section 3.2.1.3) – at least within the composition range investigated.

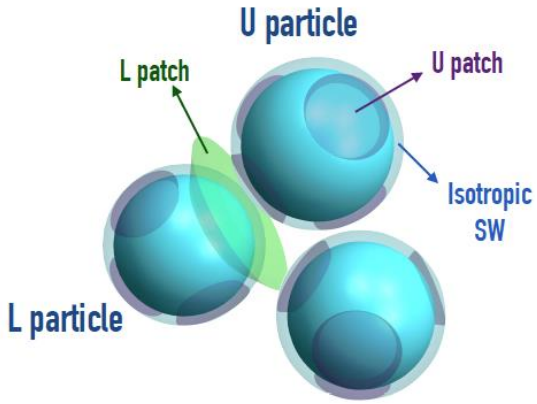

Figure 3.20: *Patchy particles used for modelling unlabelled HGD ('U particle'). Unlabelled HGD has four patches ('U patch'). Fluorescently labelled HGD ('L particle') has four patches as for the U particle and a fifth larger patch ('L patch' shown in green) representing the fluorescent label (Quinn, Gnan et al. 2015).*

The model devised in support of the experimental data (equations 2.25 and 2.26 and table 2.12 in 2.11) was numerically compared to the data for unlabelled HGD to ensure that the width of the curve and the critical volume fraction was correctly captured. Figure 3.21 shows good agreement between the experimental data (purple) and the patchy particle model (blue) in terms of coexistence curve width and critical volume fraction.

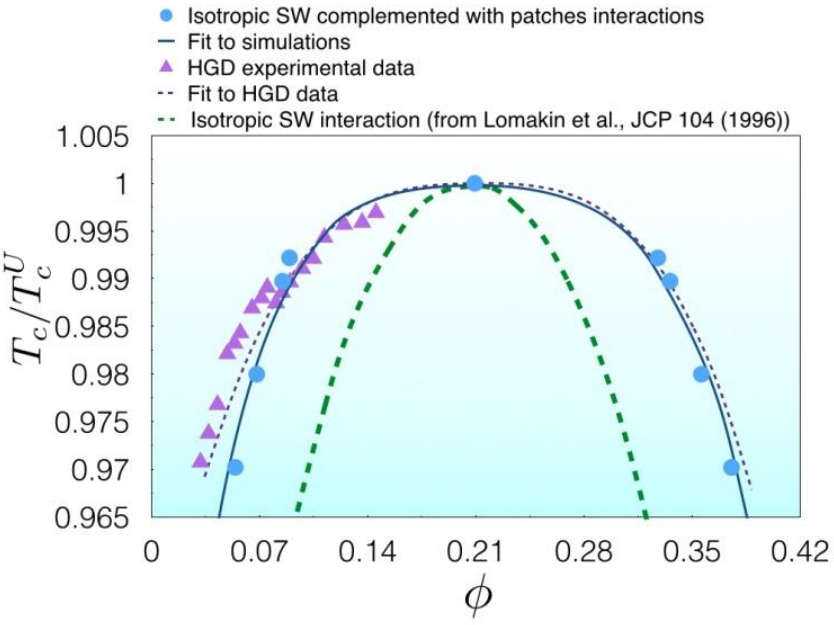

Figure 3.21: *Comparison between the experimental liquid-liquid coexistence curve of unlabelled HGD protein (purple) and two different numerical models (blue and green). Blue circles are data points for the U particle (a Kern-Frenkel patchy particle model with four randomly placed attractive patches on the surface complemented with a SW attraction representing HGD). The dark blue line is a fit of these data. The thick green dashed line is a liquid-liquid coexistence curve for γ-crystallins from a previously published simple SW model (Lomakin, Asherie et al. 1996) in which the isotropic attraction correctly describes the critical volume fraction and temperature but fails in describing the correct width of the liquid-liquid coexistence curve. Taken from Quinn, Gnan, et al. 2015.*

Unlabelled protein-labelled protein interactions were modelled by varying the parameters (attraction strength ($\varepsilon_{L,U}$), angular width in radians ((θ_{max})$_L$) and patch width ($\delta_{L,U}$)) that characterised the L patch. The combinations of these values, as well as their corresponding T_c and ϕ_c (with respect to the critical temperature of the unlabelled system T_c/T^U_c) values are summarised in table 3.4.

Table 3.4: *L patch parameters and corresponding T_c and ϕ_c values for different labelled protein compositions (x_L). Reproduced from Quinn, Gnan et al. 2015.*

$\varepsilon_{L,U}/\ \varepsilon_{U,U}$	$(\theta_{max})_L$	$\delta_{L,U}$	x_L	$T_c/T^U{}_c$	ϕ_c
1.0	0.723	0.175	0.0112	1.0056	0.214
	0.723	0.175	0.0157	1.0081	0.215
	0.723	0.175	0.0214	1.0099	0.214
	0.723	0.175	0.0290	1.0130	0.214
1.1	0.723	0.175	0.0113	1.0062	0.215
	0.723	0.175	0.0148	1.0081	0.214
	0.723	0.175	0.0216	1.0105	0.214
	0.723	0.175	0.0310	1.0142	0.214
1.2	0.723	0.175	0.0197	1.0105	0.217
	0.723	0.175	0.0195	1.0105	0.217
	0.723	0.175	0.0261	1.0105	0.213
	0.723	0.175	0.0162	1.0105	0.217
	0.723	0.175	0.0205	1.0111	0.217
	0.723	0.175	0.0260	1.0142	0.216
	0.723	0.175	0.0416	1.0203	0.217
1.5	0.723	0.175	0.0166	1.0105	0.217
	0.723	0.175	0.0223	1.0126	0.219
	0.723	0.175	0.0251	1.0139	0.219

Figures 3.22 and 3.23 show the resulting fits to the Ising probability distribution for the estimation of the critical points for unlabelled particles and binary mixtures containing labelled particles at a composition of $x_L \sim 0.02$.

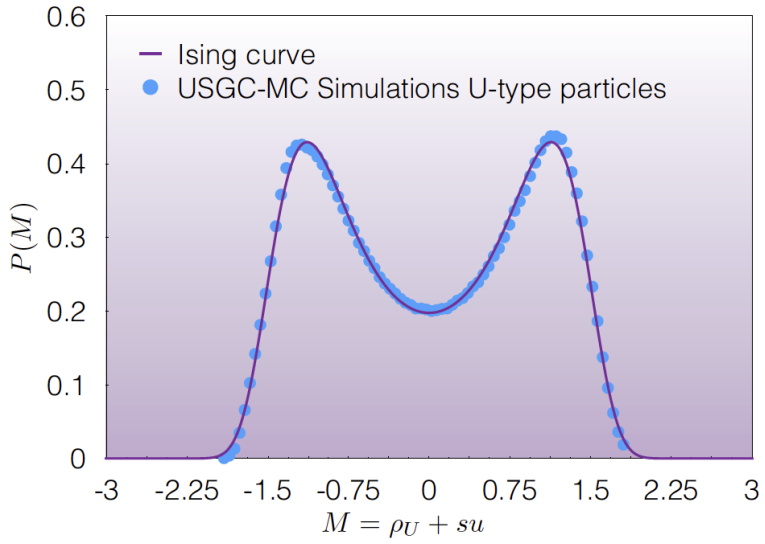

Figure 3.22: *Best fit of numerical data (circles) to the Ising order parameter distribution P ($M = \rho_U + su$) (solid line) for a monodisperse system of U particles where s = 0.028 and L = 5 for simulation box. Taken from Quinn, Gnan et al. 2015.*

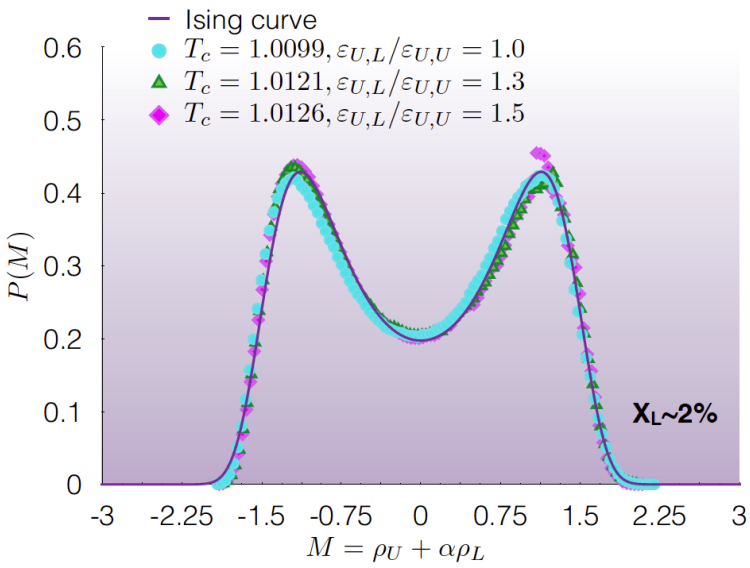

Figure 3.23: *Best fit of numerical data (circles, triangles and diamonds) for a binary mixture of U type and L type particles with a labelled particle composition of x_L = 0.02 with varying L patch interaction strength ($\varepsilon_{L,U}/\varepsilon_{U,U}$). The order parameter is given by $M = \rho_U + \alpha\rho_L$ where α = 0.001. Simulation box L =5. Angular width ($(\theta_{max})_L$) = 0.723 and width ($\delta_{L,U}$) = 0.175. Taken from Quinn, Gnan et al. 2015.*

The majority of simulations for the binary mixture were carried out with a wide angular width relative to the angular width of the U particles to allow for multiple bonding. The volume available in which to form bonds with other particles ("bonding volume") depends on both the angular width and the interaction width. The patch can be involved in more than one bond whenever the condition in equation 3.2 is satisfied.

$$\sin(\theta_{\max})_L > \frac{1}{2(1+\delta_{U,L})} \qquad \textbf{3.2}$$

Since the bonding angle is already considerably wide, an increase in the interaction width does not lead to a significant increase in critical temperature and a further increase in bond angle would create overlaps with other patches on the particle's surface. In addition, if the L patch is overly attractive, bonds with other particles become irreversible thus rendering estimation of the critical point impossible. This occurs at $\varepsilon_{L,U}/\varepsilon_{U,U}$ = 1.5 (where bond angle equals 0.723 and patch width equals 0.175). Consequently, the range of bonding energies that are directly relevant to the

experimental results (where the energetic contribution arising from the fluorescent molecule is estimated to be 3–4 times larger than a typical protein-protein interaction) cannot be directly probed. Thus, extrapolation of the critical temperature for higher bonding energies is required for comparison with experimental results. The values of the extrapolated bond energies using a labelled protein composition of $x_L \sim 0.02$ is shown in figure 3.24.

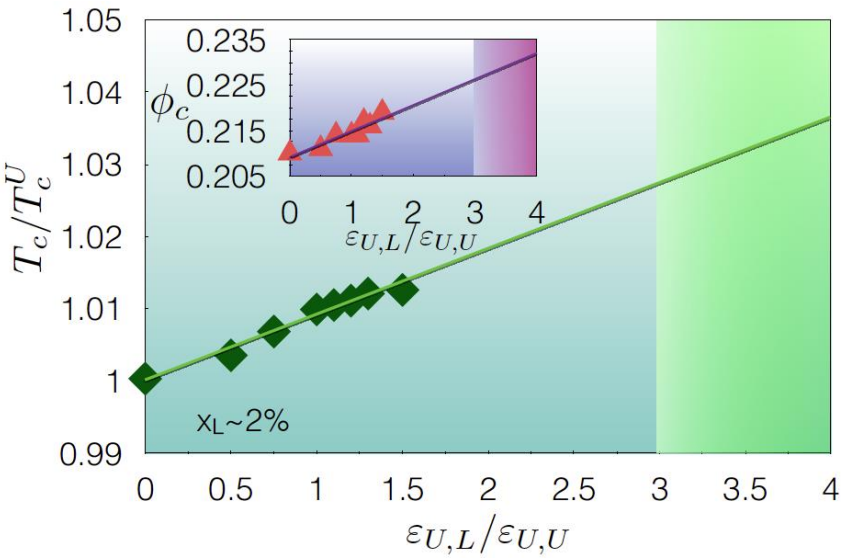

Figure 3.24: *Evolution of the critical temperature of the mixture of labelled and unlabelled particles (normalised to the critical temperature of the unlabelled system T^U_c) as a function of the bond energy between a U-patch and a L-patch $\varepsilon_{L,U}$. The fraction of labelled particles is $x_L \sim 0.02$. The darker area in the plot corresponds to the range of interaction energies comparable to experiments. Inset: critical volume fraction ϕ_c as a function of $\varepsilon_{L,U}$. Taken from Quinn, Gnan et al. 2015.*

The significant increase in critical temperature of the mixture as a result of adding a very small amount of L particles (i.e. $x_L = 0.02$) qualitatively agrees with the experimental data which showed that the addition of a small amount of fluorescently labelled protein resulted in an increase in the critical temperature. A shift in the critical volume fraction relative to the unlabelled system is also observed. However, the extrapolated data show that for a bonding energy of ~4, the critical volume fraction only shifts to 0.23. In the experimental data, a point on the descending arm of liquid-liquid coexistence curve for a mixture of unlabelled protein and amine labelled protein at $x_L = 0.002$ has been determined (section 3.2.1.4). This value is consistent with a critical volume fraction equal to 0.21 (the critical volume fraction for the unlabelled protein).

The relative concentrations of L particles in each phase (dilute and dense phases) have also been examined. L particles were found to be present in the dense phase alone. This observation is also consistent with the experimental data where the majority (~2:1) of labelled protein partitioned into the dense phase following total liquid-liquid phase separation.

3.3 Conclusion

HGD was labelled using three different small molecule fluorescent dyes: Ana Tag HiLyte 405 fluorescent dye was used for specifically labelling HGD in the Lys-2 position; DyLight 405 Maleimide fluorescent dye was used for specifically labelling HGD at the Cys-110 position; FITC was used to non-specifically fluorescently label HGD. Using fluorescent dyes for labelling specific amino acids (i.e. Lys and Cys) and carefully controlling reaction conditions, it was possible to label HGD with a maximum of 1 dye molecule per protein ensuring that the labelling efficiency was calculated accurately. Each labelled protein was used in mixtures of various labelled protein fractions and the liquid-liquid phase boundaries were measured for each. Labelled protein fractions as low as x_L = 0.0001 were sufficient to increase liquid-liquid phase separation temperature (T_{ph}) significantly (ΔT_c ~ 14 K) (and indicated an increase in the net attraction of the solution relative to a solution containing HGD alone). T_{ph} was affected by not only the varying chemistries of the fluorescent label but also by the position labelled.

A specifically designed patchy particle model in which unlabelled HGD is represented as a particle containing four attractive sites and labelled HGD is represented as for the unlabelled particle but with the addition of a fifth larger and more attractive site was used to numerically assess the experimental observations. This model proved to be qualitatively consistent with the experimental data for amine labelled HGD and approximately conserved the critical points estimated in the experimental system.

The data outlined here present several important observations. Anecdotally, it is known that fluorescent labelling can diminish protein solubility. This has led to the derivatization and development of small molecule fluorescent labels with hydrophilic moieties to impart increased solubility. Here, it has been shown that the origin of the

decreased solubility as a result of fluorescent labelling is not directly due to the formation of aggregates (however, a reduced solubility would lead to aggregate formation downstream) or a significant structural change but is due to an increased net inter-protein attraction as a result of an increased number of attractive sites on the surface of the protein. It has also been shown that the effect of labelling is also highly dependent on the site used for labelling due the inherent anisotropic nature of protein. While this result is not unsurprising, the low labelled protein compositions required to induce such a significant change in the liquid-liquid phase separation temperature are somewhat unexpected.

Fluorescence methods are widely used as a means of understanding protein behaviour. This study quantitatively assessed the impact different labelling methodologies have on the solution behaviour and, therefore, stability of fluorescently labelled proteins. From this study, it is clear that fluorescently labelled proteins are best used in small quantities. Consequently, studies requiring the use of fluorescently labelled proteins must take into account these important effects so as to avoid a misinterpretation of data as a result of an increase in the net inter-protein attraction.

In a broader sense, the results presented here serve to provide an insight into the role anisotropic protein-protein interactions play in determining protein solution behaviour.

Chapter 4:
Chemical modification of human γD-crystallin as a general route to probe anisotropic protein-protein interactions

4.1 Introduction

4.1.1 Background

Using a combination of experiments and coarse grained simulations, we have demonstrated that fluorescently labelling HGD using small molecule fluorophores increased the temperature at which LLPS occurs (T_{ph}) due to an increase in net-attraction between all proteins in the system. Small molecule fluorophores are significantly larger and inherently more hydrophobic in nature (given their aromatic structures) than any naturally occurring amino acid. It was, therefore, unsurprising that chemical modification using fluorophores led to a significant change in T_{ph} for solutions of varying composition of fluorescently labelled protein.

In this chapter, T_{ph} has been experimentally measured across a range of volume fractions for different types of modified HGD in mixtures with unmodified HGD. First, two separate amino acids on the surface of HGD, Lys-2 and Cys-110, have been chemically modified using PEGylated biotin. Second, a single mutant of HGD that replaces a proline with a valine at position 23 (P23V) has been used in a mixture with HGD.

PEGylated biotin was used to chemically modify sites 2 and 110 of HGD. Biotin is a vitamin (also known as Vitamin H, Vitamin B7, or Coenzyme R) with a valeric acid side-chain that can be derivatized to enhance its functionality (figure 4.1). It is a relatively small molecule with a molecular weight of 244.3 g mol^{-1} compared to the molecular weight of HGD which is approximately 21 kDa. The interaction of biotin with the larger biotin-binding protein avidin (67 – 68 kDa) and other biotin-binding proteins streptavidin and NeutrAvidin is the strongest known non-covalent interaction ($K_d = 10^{-15}$ M for biotin-avidin) and this interaction is often exploited for protein purification and detection.

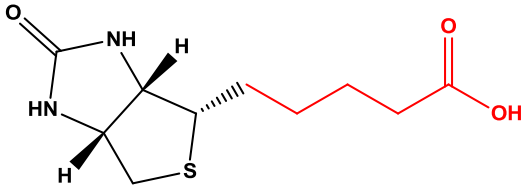

Figure 4.1: *Structure of biotin showing valeric acid (highlighted in red).*

While biotin itself is water-soluble, its derivitization with a PEG linker (figure 4.2) is intended to enhance the hydrophilicity of the biotin molecule by increasing the solvation shell around the molecule.

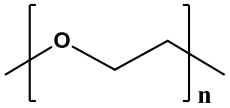

Figure 4.2: *Structure of ethylene glycol.*

The P23T single mutant of HGD is associated with a number of cataract phenotypes and forms amorphous aggregates at low concentrations (1–2 mg ml^{-1} at room temperature) with an inverted temperature dependence at physiological pH (Nandrot, Slingsby et al. 2003; Shentu, Yao et al. 2003; Mackay, Andley et al. 2004; Pande, Annunziata et al. 2005). Two other mutations at site 23 (P23S and P23V) have also demonstrated a decreased solubility (to different extents) with an inverted temperature dependence (Pande, Annunziata et al. 2005). The dramatically altered solubility associated with P23T, P23S and P23V mutants is due to an increase in the surface hydrophobicity of the mutant protein relative to the native HGD as a result of the addition of an attractive "sticky" patch to the protein's surface (Pande, Ghosh et al. 2010). Previously, the effect of the P23V mutation on the liquid-liquid phase boundary was investigated (Pande, Annunziata et al. 2005; McManus, Lomakin et al. 2007). The P23V single mutant is the only mutant of the three mutants discussed that can be concentrated to a concentration range where LLPS can be measured since its (retrograde) solubility is higher than both P23S and P23T (McManus, Lomakin et al. 2007). These studies found that the liquid-liquid phase boundary for P23V was indistinguishable from the liquid-liquid phase boundary for HGD despite the remarkable change in the mutant's solubility relative to the native protein. It was determined that this observation was directly related to the highly anisotropic nature of the protein where its net attraction was spatially averaged in the liquid phase but fully engaged in the solid phase (McManus, Lomakin et al. 2007). The mutation of HGD of a Pro to a Val at site 23 is analogous to the chemical modification of HGD using small molecule fluorescent labels in the sense that both give rise to the addition of an attractive patch on the surface of the protein (albeit at different positions on the protein surface). Banerjee, Pande et al. (2010) showed that there was no change in T_{ph} for the E107A single mutant of HGD relative to native HGD. Like P23V, E107A

has very similar tertiary and secondary structures to HGD. However, when E107A was used in a mixture with α-crystallin, there was a notable increase in T_{ph}, the value of which was dependent on the fraction of α-crystallin present in the mixture. It is, therefore, reasonable to ask if the net protein-protein attraction for the P23V mutant (which adds an attractive, hydrophobic patch to its surface) that was averaged out in the liquid phase for a single protein system (i.e. P23V alone) remains averaged out in a binary protein system (i.e. in mixtures with native protein).

4.1.2 Aim of study

The aim of this study was to experimentally measure T_{ph} across a range of volume fractions for different types of modified HGD. First, two separate amino acids on the surface of HGD, Lys-2 and Cys-110, were chemically modified using PEGylated biotin and used in mixtures with HGD. Second, the P23V single mutant of HGD was used in a mixture with HGD.

In modifying HGD at different sites, it was possible to demonstrate that the effect of surface modification, be it by chemical modification or mutagenesis, depends not only on the physicochemical properties of the modifier but also on the site of the modification.

4.2 Results

Two forms of modified HGD were used to investigate the phase behaviour of mixtures of modified and unmodified protein: HGD chemically modified with PEGylated biotin and the P23V mutant of HGD. HGD for chemical modification was produced recombinantly in *E. coli* and chemically modified after purification using PEGylated biotin with reactive moieties for residue specific modification. The chemically modified sites (shown in figure 4.3) used in this study were Lys-2 and Cys-110. A single mutant form of HGD containing a mutation of Proline (Pro, P) at position 23 to a Valine (Val, V) (P23V) (achieved by site-directed mutagenesis) was also produced using recombinant methods (also shown in figure 4.3).

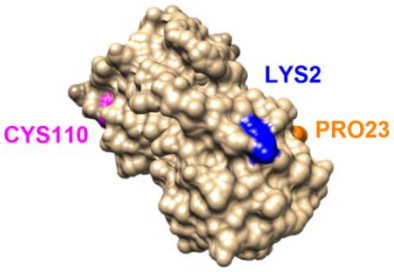

Figure 4.3: *Topological rendering of HGD showing the locations of modified sites, Lys-2 (blue), Cys-110, and P23V (orange).*

The liquid-liquid phase separation (LLPS) temperature (T_{ph}) was used to quantify the change in net-interaction potential between proteins for mixtures of modified and unmodified proteins.

4.2.1 Production and purification of P23Vsingle mutant protein

HGD was expressed, extracted, purified and characterised as detailed in chapter 3. The purity of each batch produced was confirmed by SE-HPLC and SDS PAGE. The P23V single mutant protein of HGD was produced recombinantly in *E. coli* and extracted and purified using SEC (figure 4.4) and IEX (figure 4.5) Characterisation of the purified product included SE-HPLC, SDS PAGE, mass spectrometry and circular dichroism. SE-HPLC (figure 4.6) and SDS-PAGE (figure 4.7) confirmed the product purity.

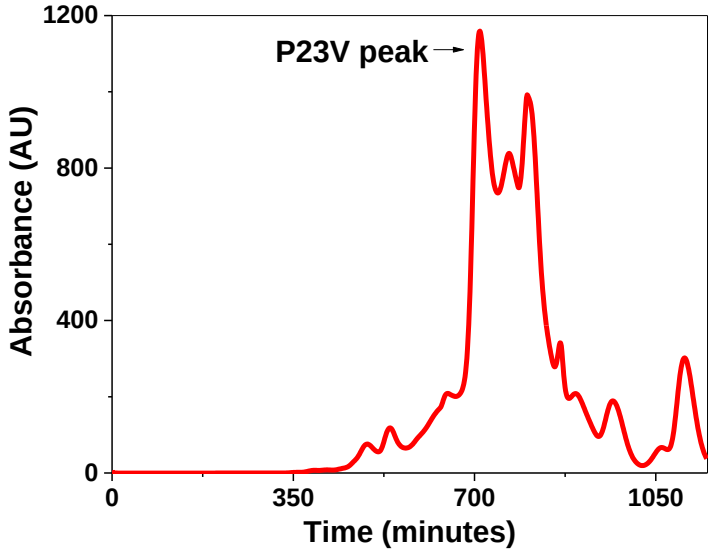

Figure 4.4: *Size exclusion chromatogram indicating the peak associated with P23V.*

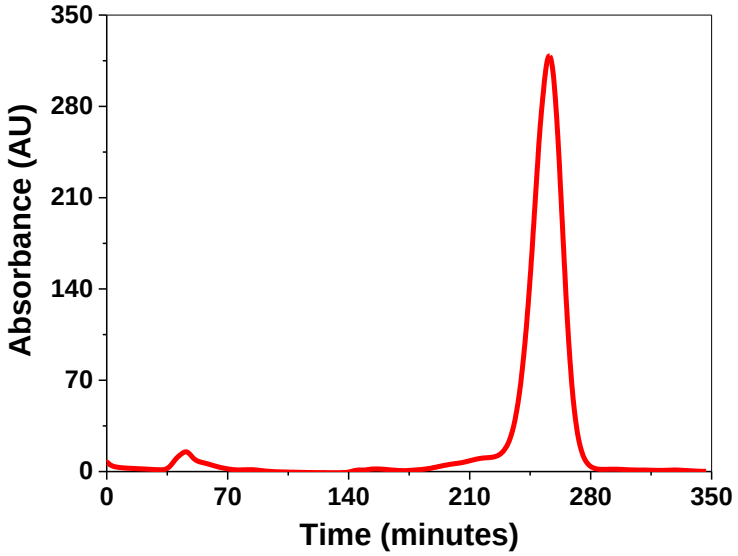

Figure 4.5: *Ion exchange chromatogram after purifying fractions from the size exclusion step. The major peak contains the P23V mutant of HGD. Following purification, only fractions from the peak tail were pooled to ensure protein purity.*

The SE-HPLC chromatogram for P23V shows one peak at 24 minutes (figure 4.6). Using the calibration curve for this column under similar buffer conditions, this indicates a molecular weight of ~21 kDa, a value consistent with the previously published molecular weight of a P23V monomer (20 608 ±1 Da) (Pande, Annunziata et al. 2005).

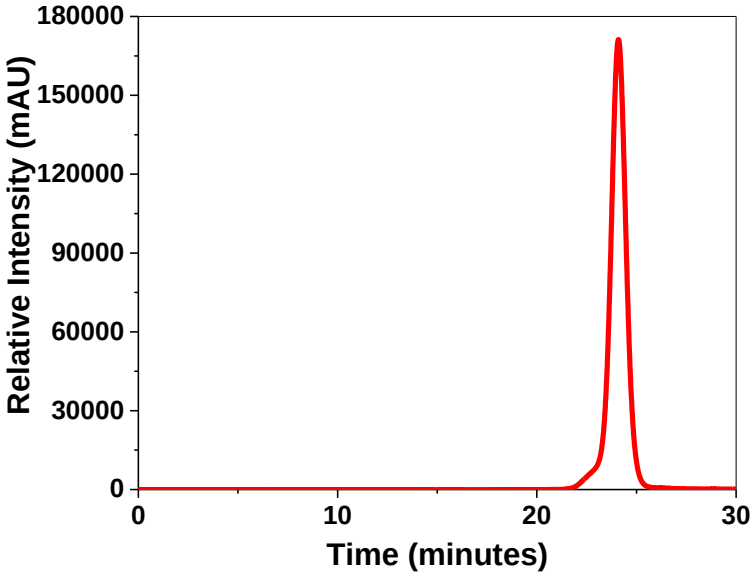

Figure 4.6: *SE-HPLC chromatogram for P23V single mutant of HGD prepared in 0.1 M sodium phosphate at pH 7.0 with 20 mM DTT.*

SDS-PAGE was also used to assess protein purity. A clean band just below the 21.5 kDa molecular weight marker indicates pure monomer protein (figure 4.7).

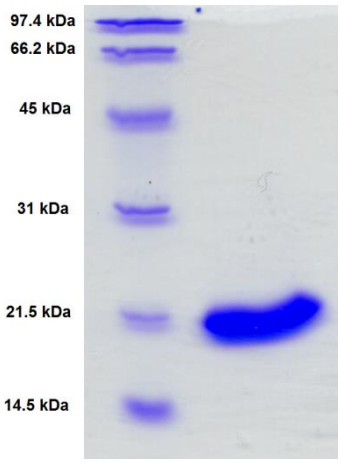

Figure 4.7: *12.6% reducing SDS-PAGE gel showing a single band indicating the purity of P23V single mutant protein.*

4.2.2 Biotinylated HGD

4.2.2.1. Phase diagram measurement

HGD was chemically modified using two forms of PEGylated biotin for separate modification of a primary amine at Lys-2 and a thiol at Cys-110. The ascending limbs of the liquid-liquid coexistence curves using two modified protein compositions (x_m) were measured for both Lys-2 (figure 4.8) and Cys-110 modifications (figure 4.9) up to total protein volume fractions of ~0.14 and ~0.07 respectively.

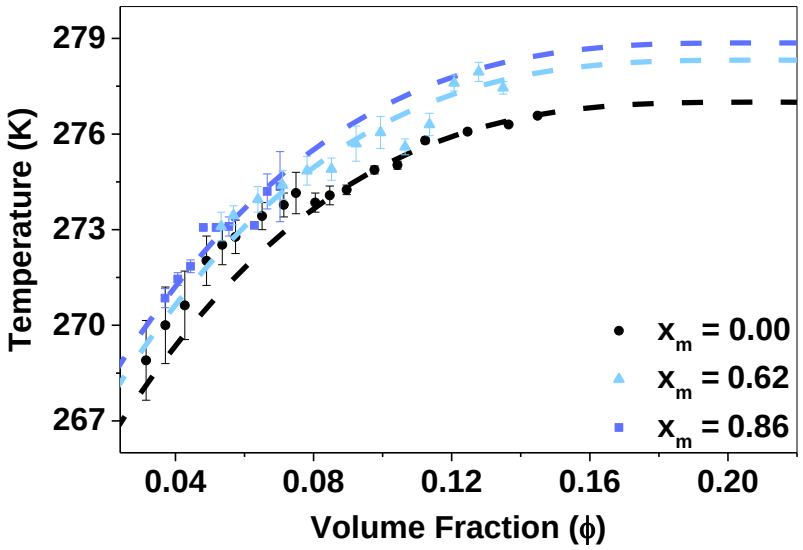

Figure 4.8: *Liquid-liquid coexistence curves for unlabelled HGD (black circles) and HGD chemically modified with PEGylated biotin at Lys-2 with modified protein fractions of x_m = 0.62 (light blue triangles) and x_m = 0.86 (cornflour blue squares). Values for data points shown given in tables A1 and A5 in Appendix.*

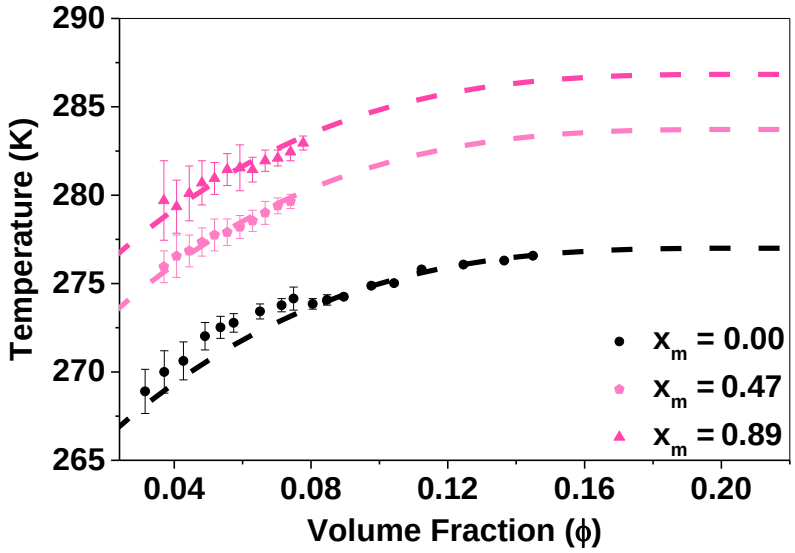

Figure 4.9: *Liquid-liquid coexistence curves for unlabelled HGD (black circles) and HGD chemically modified with PEGylated biotin at Cys-110 with modified protein fractions of x_m = 0.47 (light pink pentagons) and x_m = 0.89 (dark pink triangles). Values for data points shown given in tables A1 and A6 in Appendix.*

The data for unmodified HGD and each composition of modified protein was fitted using equation 5.1 (Broide, Berland et al. 1991):

$$\left(\frac{\phi-\phi_c}{\phi_c}\right)=A\left(1-\frac{T}{T_c}\right)^{\beta} \qquad \textbf{5.1}$$

where ϕ and ϕ_c are the volume fraction and critical volume fraction respectively (where the partial specific volume, *v*, of HGD and its modified forms is taken to be 7.1 ± 0.1 x 10^{-4} ml mg^{-1}), *A* is a parameter related to the width of the coexistence curve, *T* and T_c are the phase separation temperature and critical temperature respectively and β is an exponent term for the three-dimensional Ising model equal to 0.325. The critical volume fraction, ϕ_c, was taken to equal 0.21, a value consistent with ϕ_c for all gamma crystallins (Liu, Asherie et al. 1996). The estimated values for T_c, ΔT_c and *A* are given in table 5.1 for each composition of chemically modified protein.

Table 4.1: *Values for estimated changes in critical temperature to the nearest K (ΔT_c) with respect to unmodified HGD using equation 1 for each binary protein mixture in aqueous solution*

Modified Protein	T_c **(K)**	ΔT_c **(K)**	***A***
Unmodified HGD (x_m = 0.00)	277	-	2.6
HGD amine modified with PEG_4-Biotin (x_m = 0.*62*)	278	1	2.6
HGD amine modified with PEG_4-Biotin (x_m = 0.*86*)	279	2	2.6
HGD thiol modified with PEG_2-Biotin (x_m = 0.*47*)	283	6	2.6
HGD thiol modified with PEG_2-Biotin (x_m = 0.*89*)	287	10	2.6

The data clearly indicate that there is a greater change in T_{ph} after modifying the thiol at site 110 than modifying the primary amine at site 2. This observation is consistent with the published data that are presented in Chapter 3 of this thesis where HGD was chemically modified at Lys-2 and Cys-110 using small molecule fluorescent dyes (Quinn, Gnan et al. 2015). When fluorescently labelled protein was in mixtures with unlabelled HGD at protein compositions between 0.0001 and 0.01 (partial data shown in figure 4.10), an increase in T_{ph} with increases in the proportion of modified protein were observed. The proportion of labelled protein that was required to

increase T_{ph} was substantially lower for protein modified at the Cys position (by a factor of ~50).

The range of the protein-protein interaction is related to the width of the liquid-liquid coexistence curve where a decrease in coexistence curve width is indicative of an increase in the interaction range of the protein (Lomakin, Asherie et al. 1996; Chen, Vekilov et al. 2004). While there is very little change in T_{ph} for HGD modified with PEGylated biotin at the amine position, there is a substantial increase in T_{ph} with increasing proportions of HGD modified with PEGylated biotin at the thiol position. However, there is no change in the width of the coexistence curve for HGD modified with PEGylated biotin (parameter A in equation 4.1, see table 4.1). This would suggest that the increased net attraction as indicated by the increase in T_{ph} is not accompanied by an increase in the range of the attraction. Figure 4.10 shows partial coexistence curves for HGD in mixtures with fluorescently labelled HGD modified at the amine position (NH) and thiol position (SH) where an increase in T_{ph} was accompanied by a decrease in curve width. This is in contrast to the coexistence curves for HGD modified with PEGylated biotin. There was also a substantial difference in the fraction of modified protein required to induce a change in T_{ph} of similar magnitude (x_m=0.0001 for fluorescently labelled HGD versus x_m=0.89 for HGD chemically modified with PEGylated biotin). Hence there are notable differences between this and the case of fluorescently labelled proteins.

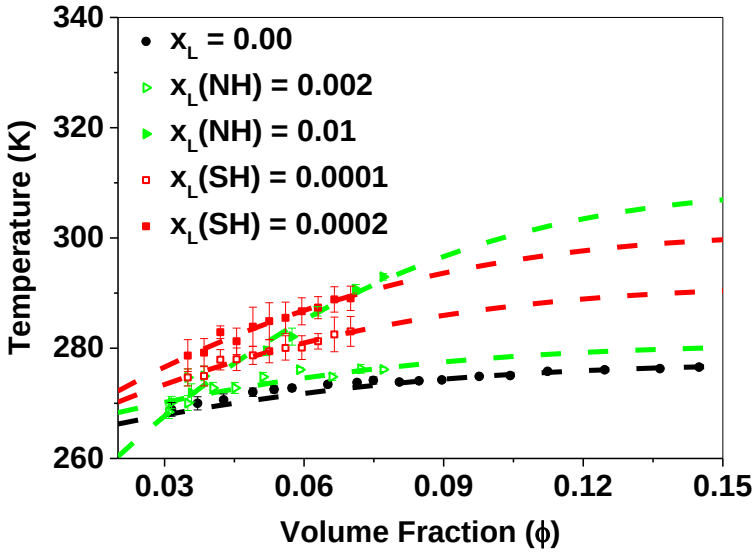

Figure 4.10: *Liquid-liquid coexistence curves for HGD (black) and HGD chemically modified via fluorescent labelling of a primary amine (green) (A), and a thiol (red) (B).*

Electrostatic inter-protein interactions between charged groups on the protein's surface contribute to its solubility (Cooper 2011). The pI of HGD (pI = 7.2 ± 0.1 (Banerjee, Pande et al. 2011)) is almost at physiological pH and is very close to the working pH used in the data presented here. Lysine (along with histidine and arginine) is one of three positively charged amino acid residues at neutral pH (pH 7.0). The modification of Lys-2 by covalent attachment of the PEGylated biotin thus removes one positive charge from the protein surface and results in a protein with a net negative charge. This change in charge may be a contributing factor as to why the increase in net attraction (as indicated by the lower change in T_{ph}) is not as substantial as the increase in T_{ph} for HGD modified at Cys-110 with PEGylated biotin. Since the structures of the small molecule fluorescent dyes previously used for specific labelling of HGD are proprietary, it is not possible to comment on how this affected the electrostatic contribution of the net interaction potential. FITC, which has a partial negative charge at neutral pH, can modify both primary and secondary amines. As a result, HGD labelled with FITC will carry a net negative charge which should contribute to its solution stability. A mixture of HGD and HGD fluorescently labelled with FITC (x_L = 0.0002) led to an increase in T_{ph} of 20 K. This would suggest that the net negative charge of FITC labelled HGD did not contribute

as significantly to the net protein interaction potential as the increase in surface hydrophobicity as a result of FITC labelling.

4.2.2.2 Modified protein partitioning

A binary protein solution of known volume fraction containing HGD and HGD chemically modified at Cys-110 with a modified protein composition of x_m = 0.89 was cooled (-4 °C) to induce liquid-liquid phase separation and the liquid phases were macroscopically separated by centrifugation at -4 °C into the dilute and dense protein phases (figure 4.12). The volume fraction and modified protein composition of each phase was determined. The dilute phase was determined to have a modified protein composition x_m equal to 0.80, while in the dense phase x_m equalled 0.87 indicating that the protein fraction of HGD in the dilute and dense phases was 0.20 and 0.13 respectively. The propensity of the less attractive unmodified HGD to preferentially partition into the dilute phase is consistent with previously published data (Wang, Lomakin et al. 2010) and is directly related to the relative strengths of the inter-protein attractions, i.e. HGD-HGD, HGD-modified-HGD and modified HGD-modified HGD interactions, with the more attractive protein (i.e. modified HGD) preferentially partitioning into the dense liquid phase. This observation is also consistent with our previous data using fluorescently labelled HGD where the majority of the labelled protein (i.e. the more attractive protein) partitioned into the dense protein phase (Quinn, Gnan et al. 2015).

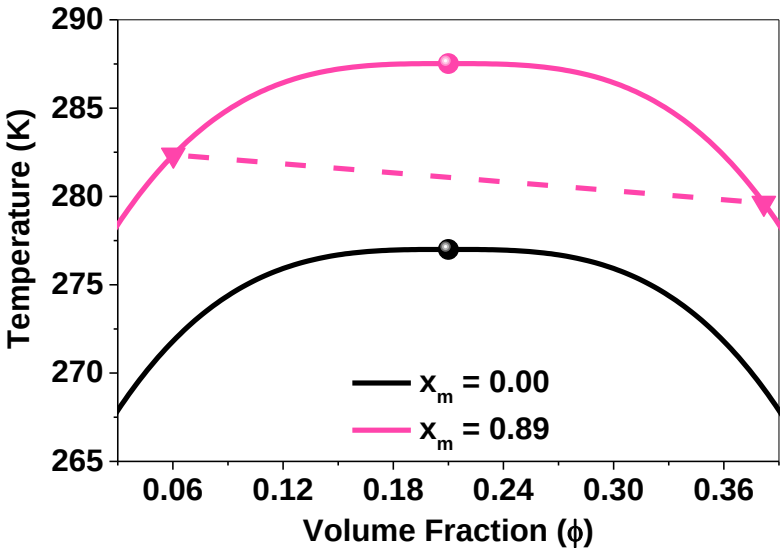

Figure 4.12: *Liquid-liquid coexistence curves for unmodified HGD (black) and a binary protein mixture in aqueous solution containing 89% (of total protein) PEG_2-Biotin modified HGD at Cys-110 . The triangles indicate the volume fractions for the dilute phase (ϕ = 0.06) and concentrated phase (ϕ = 0.38) after liquid-liquid phase separation. The dashed line is a tie-line and connects the pair of volume fractions after phase separation. The percentage of biotinylated protein was determined to be 80% in the dilute phase and 87% in the concentrated phase.*

4.2.2.3 Structural analysis

A structural or conformational change as a result of unfolding can cause an increase in the net attraction of a protein due to the exposure of hydrophobic residues. To determine if this was the case for both the amine and thiol modified proteins, near and far ultraviolet (UV) circular dichroism (CD) spectroscopy and second derivative UV spectroscopy were used to assess the impact of chemically modifying HGD with PEGylated biotin on its respective tertiary and secondary structures.

4.2.2.3.1 Circular dichroism spectroscopy

Far-UV CD spectral data for amine modified (4.13) and thiol modified (figure 4.14) HGD indicates that there was no significant change in the secondary structure of the biotin modified protein relative to the unmodified protein.

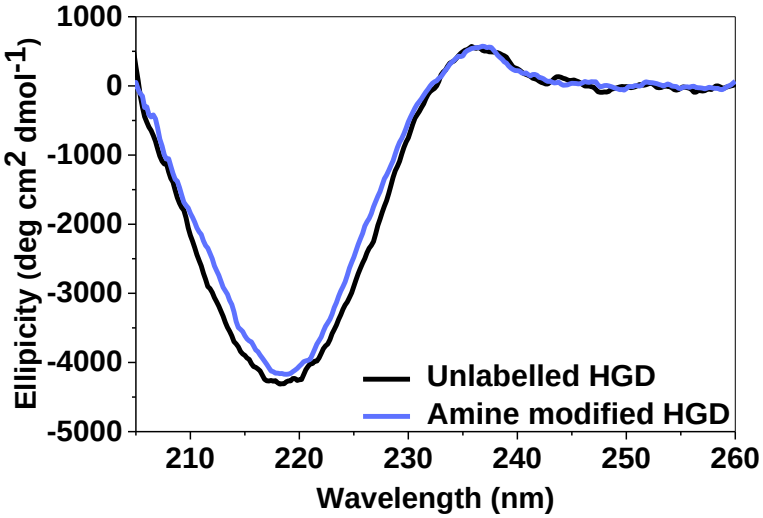

Figure 4.13: *Far UV CD spectra for unlabelled HGD (black) and amine modified HGD (blue) indicating very similar secondary structures.*

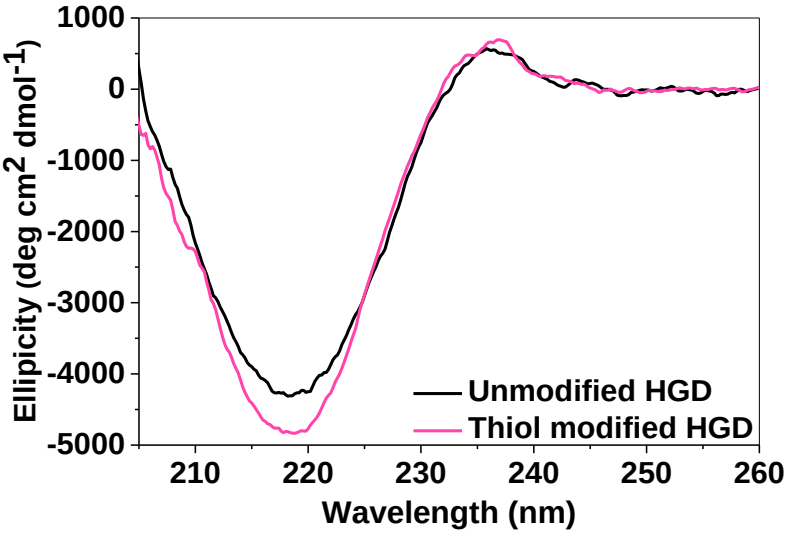

Figure 4.14: *Far-UV CD spectra for unlabelled HGD (black) and thiol modified HGD (pink) indicating very similar secondary structures.*

Near-UV CD spectral data for amine modified (4.15) and thiol modified (figure 4.16) HGD indicates that there was no significant change in the tertiary structure of the modified proteins relative to the unmodified protein.

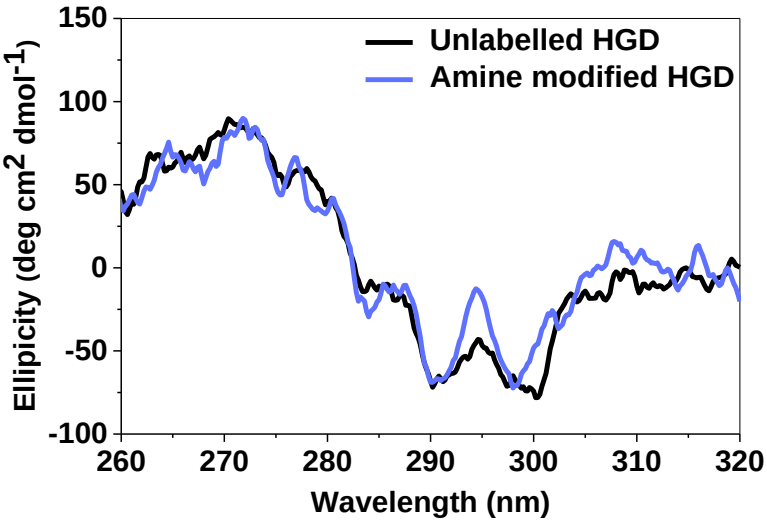

Figure 4.14: *Near-UV CD spectra for unlabelled HGD (black) and amine modified HGD (green). There is a slight perturbation in the spectral data for the wavelength range corresponding to Phe indicating a localised change to the tertiary structure.*

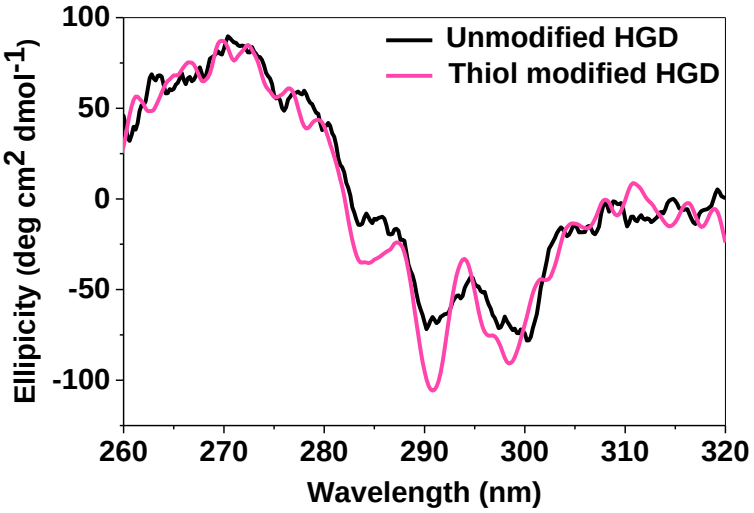

Figure 4.16: *Near-UV CD spectra for unlabelled HGD (black) and thiol modified HGD (pink) indicating very similar tertiary structures.*

Near and far UV CD spectral data has already been published for the P23V single mutant of HGD (Pande, Annunziata et al. 2005). These data indicated very similar spectra for both near and far UV wavelengths. Again, this is indicative of a minimal impact on tertiary and secondary structures as a result of a modification to the surface of HGD and thus that a structural or conformational change is not responsible for the net increase in attraction of the modified protein, as indicated by the increase in T_{ph}.

4.2.2.3.2 Second derivative UV spectroscopy

Second derivative UV spectroscopic analysis was used as a complement to near-UV CD spectroscopy for tertiary structure analysis. The spectroscopic data for amine modified (figure 4.17) and thiol modified (figure 4.18) HGD indicate no significant change in the modified protein's tertiary structure relative to unmodified HGD. This is consistent with the spectral data obtained using near-UV CD.

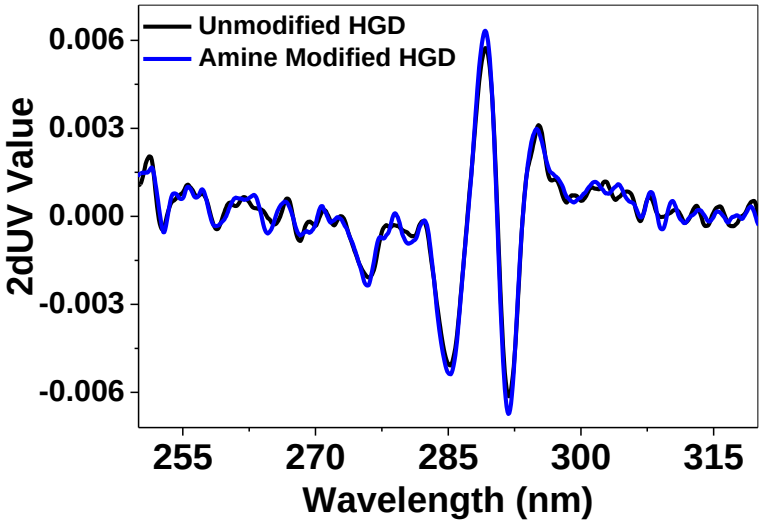

Figure 4.17: *Second derivative UV analysis of unlabelled HGD (black) and amine modified HGD (blue).*

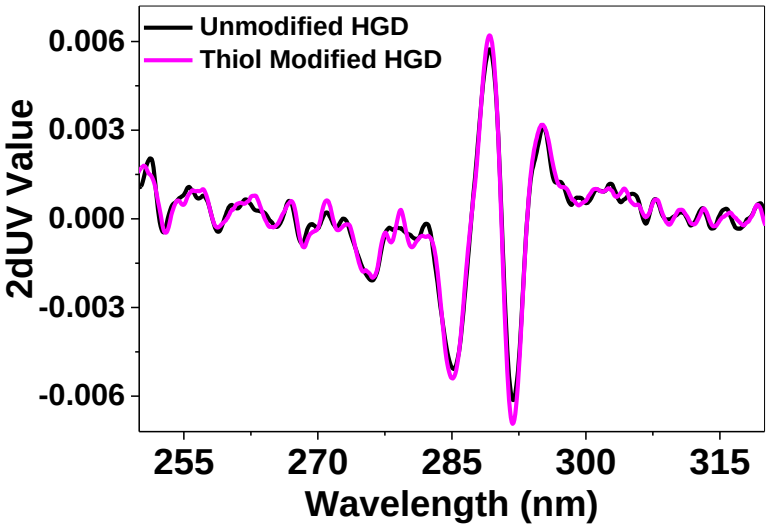

Figure 4.18: *Second derivative UV analysis of unlabelled HGD (black) and thiol modified HGD (pink).*

4.2.3 HGDP23V mixture; Effect on liquid phase diagram

A partial liquid-liquid phase boundary was measured for a 50:50 mixture of HGD and P23V to determine the effect of the mutation in a mixture with the native protein in terms of its net interaction potential (figure 4.19).

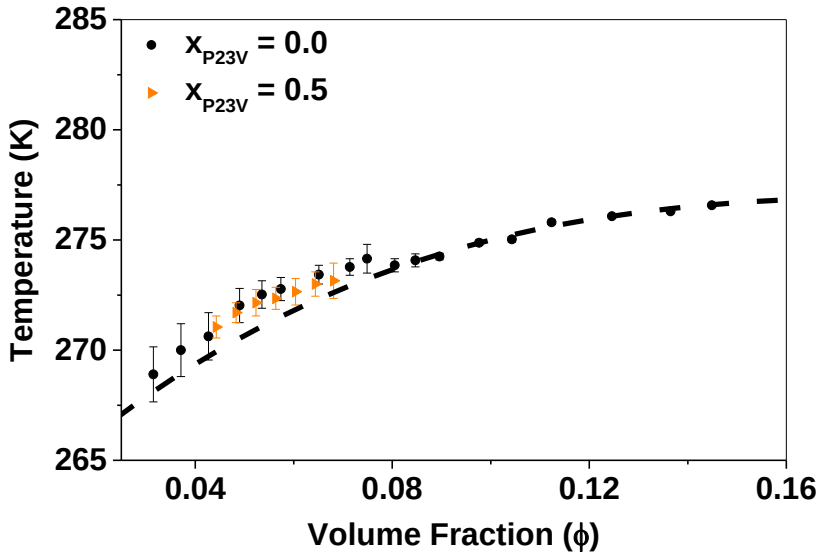

Figure 4.19: *Liquid-liquid coexistence curve for unmodified HGD (black) and a binary protein mixture in aqueous solution containing 50% (of total protein) P23V mutant of HGD (orange) showing the phase separation temperature across a range of volume fractions for HGD in a 50% mixture with P23V. Values for data points shown given in tables A1 and A7 in Appendix.*

There is no apparent change in the LLPS temperature for a mixture of P23V and HGD relative to phase separation temperature of HGD (or, indeed, P23V) alone (figure 4.19). These data would suggest that, even when used in a mixture, the effect of the mutation at site 23 remains averaged in the liquid phase.

4.3 Conclusion

HGD was modified in two ways: by chemical modification of two different amino acid residues (Lys-2 and Cys-110) using PEGylated biotin and by mutagenesis to replace a Pro at site 23 with a Val. These modified proteins were then used in mixtures with unmodified HGD to assess what impact, if any, the presence of the modified protein had on solution behaviour relative to solutions containing unmodified protein alone. Both mixtures containing a chemically modified protein (i.e. Lys modified HGD and Cys modified HGD) showed an increase in the net

attraction of the system (as indicated by an increase in T_{ph}) relative to the unmodified protein, albeit to different extents; Mixtures containing HGD amine modified at Lys-2 with modified protein compositions of x_m = 0.62 and 0.86 both showed a very slight increase in T_{ph} (≤~2 K) while mixtures containing HGD thiol modified at Cys-110 with modified protein compositions of x_m = 0.47 and 0.89 showed a more significant increase (≤~10 K). These results are qualitatively consistent with the results presented in the previous chapter using HGD fluorescently labelled with a small molecule fluorescent label which showed that the protein modified at the Cys-110 position resulted in a much greater increase in T_{ph} than the protein modified at Lys-2. In this study, the width of the liquid-liquid coexistence curve measured for each protein mixture containing chemically modified and unmodified protein was unchanged. This is in contrast with the data presented in the previous chapter using small molecule fluorescent dyes, where a decrease in curve width was observed - indicative of an increase in the range of the attraction. Complete phase separation of a sample containing thiol modified HGD indicated that, as expected, the majority of the chemically modified protein (that had a higher net attraction than the unmodified protein) preferentially partitioned into the dense protein phase. This observation is consistent with previous studies using HGD and other proteins with a higher net attraction. Near and far UV CD data indicated that the increase in T_{ph} of solutions containing the chemically modified proteins is not as a result of a significant structural change in either the secondary or tertiary structure of the modified protein. These results are also consistent with the data presented in the previous chapter using small molecule fluorescent dyes.

In this chapter, it is also shown that a single mutant protein (P23V) that contains a more hydrophobic amino acid than its native form does not change the LLPS phase transition temperature when used in a 50:50 mixture with the native protein. This further demonstrates how both the modification and its position on the protein surface can impact the way in which changes to a phase boundary can manifest.

The results presented here show that the effect of modifying the surface of a protein on its solution behaviour is not only highly dependent on the physicochemical properties of the modifier used but also on the site of the modification. This study has direct industrial applicability given the widespread use of surface modifications,

especially the use of Lys and Cys modifications, in biotherapeutic innovation and development.

Chapter 5:

The self-assembly of double mutants of human γD-crystallin

5.1 Introduction

5.1.1 Background

Protein self-assembly is dependent on both the intrinsic characteristics of the protein and on its environment (Whitesides and Boncheva 2002). Using a combination of chemical modification and mutagenesis, we have shown that changes to the surface of a protein affect its phase behaviour (James, Quinn et al. 2015; Quinn, Gnan et al. 2015). For HGD, a number of single and double mutations are known to influence the phase behaviour of the protein. The McManus group have identified several different double mutant proteins for HGD with interesting phase behaviour (James, Quinn et al. 2015). In this chapter, some aspects of the phase behaviour for two of these double mutants, P23VR58H and P23TR36S were investigated in more detail.

Protein condensation is always characterised by the creation of an interface between a new dense phase and the solution phase (Buell 2017). A discontinuity in the concentration of solutes in a bulk solution makes the formation of this interface energetically unfavourable and nucleation occurs when fluctuations in energy can overcome this energy barrier (Vekilov 2010; Karthika, Radhakrishnan et al. 2016). This process is known as homogeneous nucleation. The formation of a nucleus can also occur *via* the introduction of another interface into the system, a process known as heterogeneous nucleation (Buell 2017). As a result of the different routes to nucleation, the formation of condensed phases is divided into two processes: nucleation and growth (Buell 2017).

Because nucleation rarely occurs in the bulk solution alone (and is, therefore, rarely homogeneous), the interaction of the protein solution with interfaces must be taken into account. Heterogeneous nucleation can influence phase behaviour in different ways such as by introducing variations in local protein concentration as a result of adsorption and by introducing a decrease in interfacial energy as a result of adsorption which in turn drives growth (Buell 2017).

When proteins in solution are exposed to a solid surface, they tend to accumulate at the solid/liquid interface (Haynes and Norde 1994). This is due to a lower free energy associated with the interface than the bulk material which, in turn, drives thermodynamic stabilization *via* adsorption (Nakanishi, Sakiyama et al. 2001). This

interfacial adsorption of proteins has implications for industries such as biotechnology, food and medical devices and can be both favourable (e.g. in chromatographic processes) and unfavourable (e.g. adsorption to medical devices) depending on the application (Nakanishi, Sakiyama et al. 2001; Anand, Sharma et al. 2010). Solid/liquid interfacial adsorption of proteins is influenced by the chemical properties of the protein itself (e.g. pI and surface patchiness), the solution conditions (e.g. ionic strength and pH) and the characteristics of the surface with which the protein solution is in contact (e.g. roughness and polarity) (Anand, Sharma et al. 2010; Ouberai, Xu et al. 2014). Surface passivation is commonly used to diminish non-specific protein binding to glass surfaces and involves creating a hydrophilic surface *via* PEGylation or non-covalent protein blocking (Hua, Young Han et al. 2014; Hai, Yifan et al. 2015). Surface modifications that impart a net charge or change the polarity of the surface have been developed to alter the extent of protein adsorption (Thevenot, Hu et al. 2008). Generally, hydrophilic surfaces are considered to be more resistant to protein adsorption than hydrophobic surfaces (Ostuni, Chapman et al. 2001; Wu 2009).

The Pro-23 to Tyr (P23T) single mutant of HGD is associated with congenital cataract and leads to the formation of reversible amorphous aggregates with low solubility and an inverse temperature dependence under physiological pH and temperature (Nandrot, Slingsby et al. 2003; Shentu, Yao et al. 2003; Mackay, Andley et al. 2004; Pande, Annunziata et al. 2005). Pro-23 to Val (P23V), which is not a naturally occurring mutant of HGD, is far more soluble than P23T yet maintains an inverse temperature dependence (Pande, Annunziata et al. 2005; McManus, Lomakin et al. 2007). The Arg-58 to His mutation of HGD is also associated with congenital cataract and forms crystals with a normal temperature dependence under physiological conditions. Previously, a double mutant (P23VR58H) of HGD was created that incorporated both mutations (James, Quinn et al. 2015). For the double mutant, the solubility line lies between the solubility lines for P23V and R58H and the aggregates it forms are consistently spherical in shape (James, Quinn et al. 2015). One particularly interesting feature of the growing spherical assemblies for this protein is a very consistent aggregate size as the particles grow. Work done by Susan James demonstrated that the rate of growth was concentration dependent (figure 5.1)

(James 2015). Furthermore, for assemblies grown in bulk solution, at some point, protein crystals begin to grow on the surface.

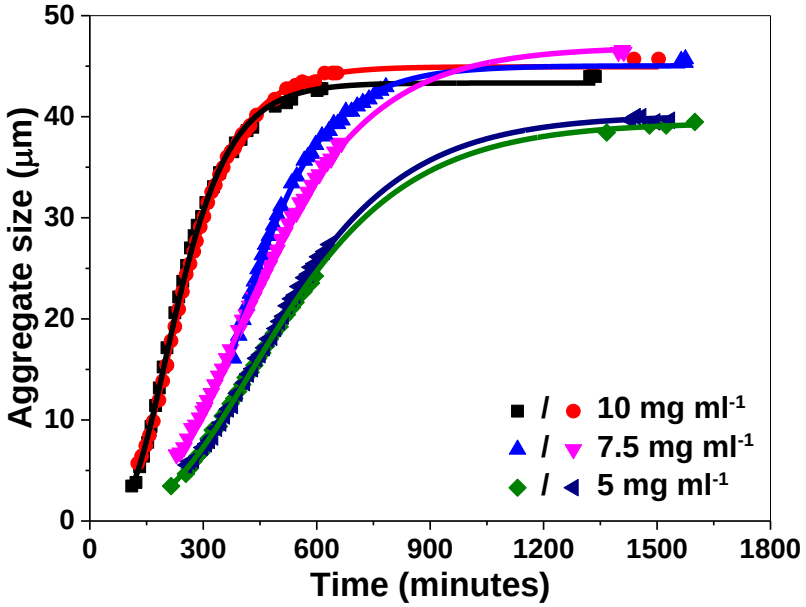

Figure5.1: *Growth of P23VR58H double mutant aggregates for different concentrations. Reproduced from James 2015.*

When these experiments were replicated in significantly reduced volumes (~80 µl versus ~360 µl in the original experiments), it became apparent that surface effects were dominating protein phase behaviour. In a bid to minimise this interfacial interaction, the nucleation and growth of the P23VR58H double mutant was investigated for a number of different surfaces.

Intracellular self-assembly of proteins is a feature of normal biological function as well as being implicated in the pathogenesis of protein condensation diseases such as sickle cell anaemia and cataract (Eaton and Hofrichter 1990; Brangwynne 2011; Shiels and Hejtmancik 2013). While protein studies carried out *in vitro* in simple aqueous solution are of huge importance to furthering our understanding of the mechanisms driving protein aggregation, they do not account for the highly complex and crowded intracellular environment and the effects of this environment on protein behaviour (Kuznetsova, Turoverov et al. 2014; Danielsson, Mu et al. 2015). The intracellular environment is an extremely crowded milieu consisting of proteins, nucleic acids, polysaccharides, etc. (Kuznetsova, Turoverov et al. 2014). Given the heterogeneity of the intracellular environment and the influence of heterogeneity on

protein nucleation and growth of the P23VR58H double mutant, we wanted to investigate the nucleation and growth of a protein expressed in a mammalian cell line. Protein concentration is an important variable when considering intracellular protein phase transitions (Brangwynne 2013). As a result, a double mutant of HGD with a low solubility at physiological pH and temperature was selected: P23TR36S. Like R58H, the Arg-36 to serine (R36S) single mutant of HGD is also responsible for congenital cataract and also has a significantly lower solubility than wild-type HGD that leads to the formation of crystals with normal temperature dependence (Kmoch, Brynda et al. 2000; Pande, Pande et al. 2001). Though position 36 is not a crystal contact in the native protein, it becomes one in the mutant protein (Kmoch, Brynda et al. 2000; Basak, Bateman et al. 2003). The equilibrium solubility lines for R36S and R58H are similar, but crystallization of the R36S mutant occurs more rapidly than the R58H mutation (Pande, Pande et al. 2001). The phase behaviour of P23TR36S was previously investigated in 0.1 M sodium phosphate buffer at pH 7.0 with 20 mM DTT (James, Quinn et al. 2015). This double mutant was shown to produce rhombic crystals with an inverted temperature dependence at physiological temperature. Because of this interesting phase behaviour at low concentrations and physiological temperature, it was selected for expression in a mammalian cell line.

5.1.2 Aim of this study

The aim of this chapter was to study the influence of environment on nucleation and growth of protein aggregates. Previous studies on the P23VR58H double mutant of HGD were carried out in large volumes. The work presented here was carried out at smaller volumes which served to amplify the surface effects on nucleation. To minimise the interaction of the protein with the surface, a number of different surface modifications were explored and their influence on the nucleation and growth of this double mutant was investigated. In addition, the effects of heterogeneity on the nucleation and growth of a different double mutant of HGD (P23TR36S) were investigated by expressing this protein in a mammalian cell line.

5.2 Results

5.2.1 The growth of P23VR58H aggregates in bulk and in small volumes

5.2.1.1 Production and characterisation of P23VR58H double mutant protein

The primers for the P23VR58H double mutant protein of HGD were designed previously and transformed in *E. coli* (James, Quinn et al. 2015). The P23VR58H double mutant protein of HGD was produced recombinantly in *E. coli* and extracted and purified using SEC (figure 5.2) and IEX (figure 5.3) SE-HPLC (figure 5.4) and SDS-PAGE (figure 5.5) confirmed the product purity.

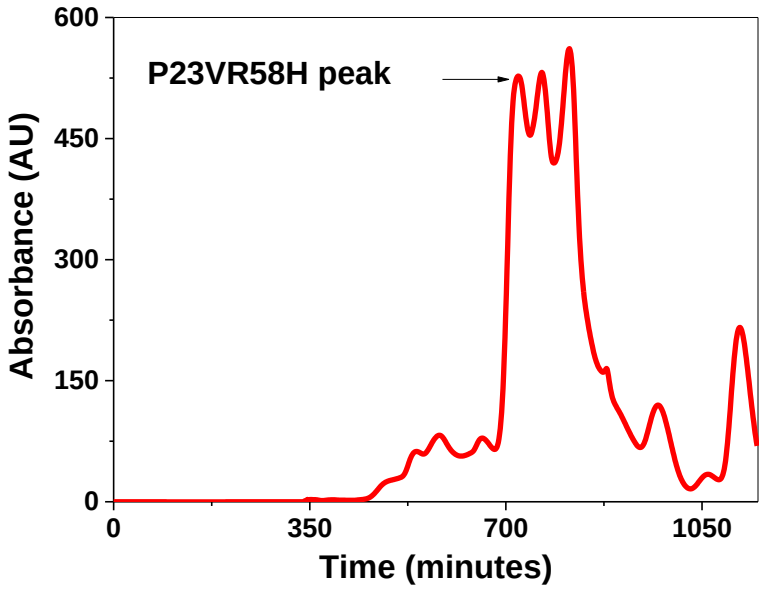

Figure 5.2: *Size exclusion chromatogram indicating the peak associated with P23VR58H double mutant protein.*

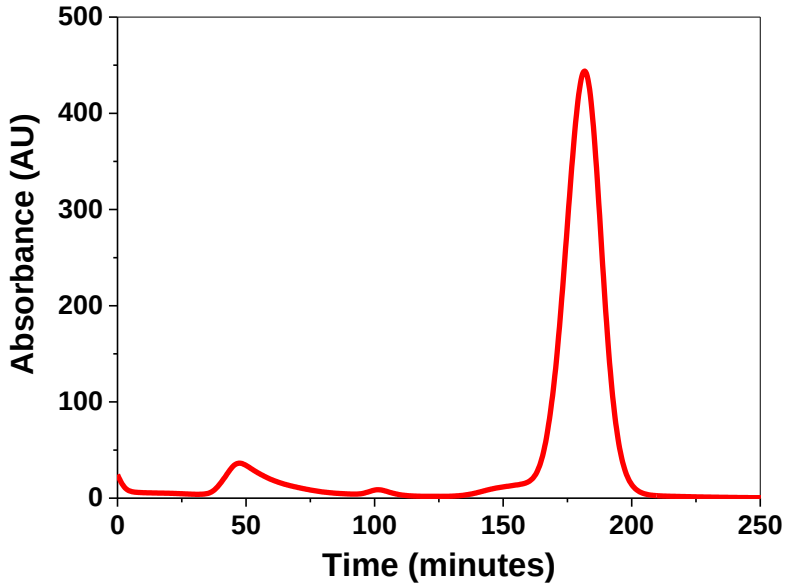

Figure 5.3: *Ion exchange chromatogram after purifying fractions from the size exclusion step. The major peak contains P23VR58H double mutant of HGD protein. Following purification, only fractions from the peak tail were pooled.*

SDS-PAGE (figure 5.3) and SE-HPLC (figure 5.4) and were used to confirm protein purity. The chromatogram for P23VR58H produced using SE-HPLC shows one peak at 24 minutes (figure 5.3). Using the calibration curve for this column under similar buffer conditions indicates a molecular weight of ~21 kDa, a value consistent with the published molecular weight of a P23VR58H monomer (20,589 ±1 Da) (James, Quinn et al. 2015).

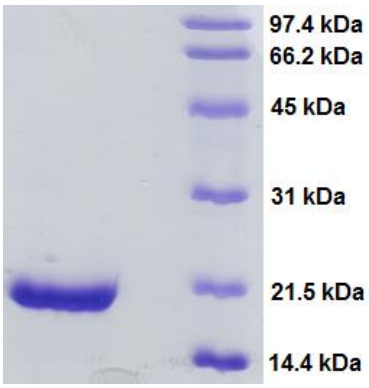

Figure 5.4: *12.6% reducing SDS-PAGE gel showing a single band indicating the purity of P23VR58H double mutant protein.*

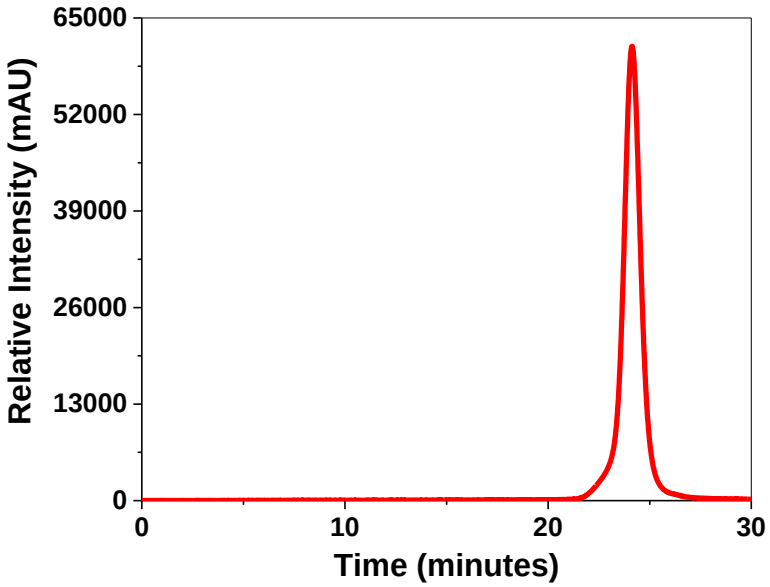

Figure 5.5: *SE-HPLC chromatogram for P23VR58H double mutant of HGD prepared in 0.1 M sodium phosphate at pH 7.0 with 20 mM DTT.*

5.2.1.2 Nucleation and growth of P23VR58H

The previous studies on P23VR58H protein aggregation were carried out using large volumes (~360 µl). When smaller volumes were used to conduct the same measurements (≤ ~80 µl), it became apparent that surface effects influenced the nucleation and growth of P23VR58H aggregates to a greater extent. We attempted to minimise this interaction by altering the surface in contact with the protein solution. To maintain a degree of consistency with the earlier work done, each study was carried out at 30 °C using a protein concentration of 10 mg ml^{-1} in 0.1 M sodium phosphate buffer, pH 7.0 with 20 mM DTT. The surface treatments employed for each chamber used in this study are described in table 5.1.

Table 5.1: *Chamber variables for monitoring growth of P23VR58H double mutant protein aggregates.*

Surface	Height (*mm*)	Diameter (*mm*)	Volume (*μl*)
Unclean, uncoated	1.00	10.04	80
Piranha cleaned, uncoated	1.00	10.04	80
Piranha cleaned, poly-D-lysine coated	1.00	10.04	80
Piranha cleaned, PEGylated	1.00	10.04	80
Piranha cleaned, silanised	1.00	10.04	80
Piranha cleaned, uncoated	0.50	10.04	40

Initially, uncleaned, uncoated coverslips were used to monitor the formation of P23VR58H aggregates in smaller volumes. The aggregates formed using these coverslips were found to form inconsistently and exhibited a polydisperse size distribution (figure 5.6). The polydispersity observed under these chamber conditions is not unexpected given that manufacturing artefacts (e.g. grease and particulates) were not removed and would hence provide multiple interfaces for heterogeneous nucleation to proceed which would influence the growth of aggregates to different extents.

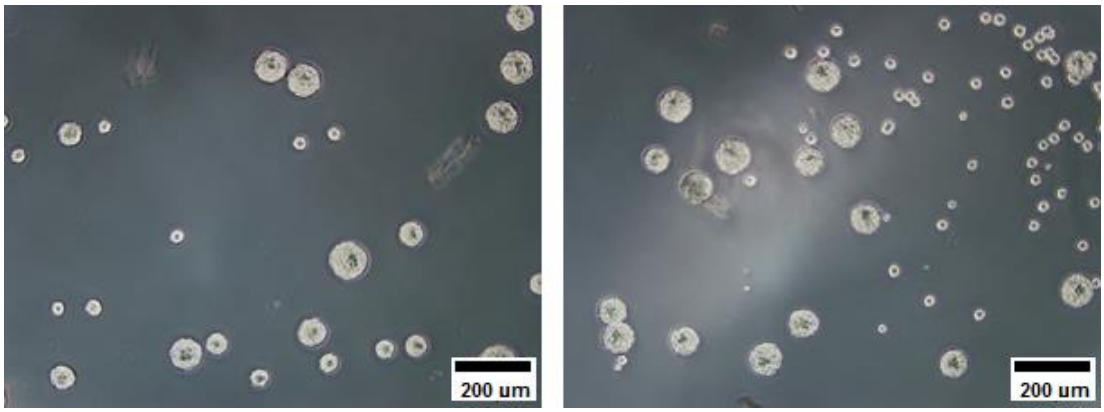

Figure 5.6: *Phase contrast microscopy images taken after 29 hours of P23VR58H aggregates formed in a chamber constructed using uncleaned, uncoated coverslips (d x h = 10.04 mm x 1 mm).*

We then attempted to minimise the interaction of the protein with the glass surface of the coverslip using several different surface treatments and hence promote homogeneous nucleation in the bulk. Piranha solution is a highly corrosive mixture of sulphuric acid and hydrogen peroxide used for the removal of organic matter from glass surfaces. The purpose of cleaning the coverslips with piranha solution was to create a negatively charged, hydrophilic surface which would limit the adsorption and subsequent nucleation of P23VR58H at the surface.

Aggregates formed in this chamber were more uniform in size than in the chamber formed using uncleaned, uncoated coverslips (figure 5.7). Silica and silicate glass surfaces acquire a negative charge in an aqueous environment due to the dissociation of the terminal silanol groups (Israelachvili 1992; Behrens and Grier 2001). This acquired charge could have also contributed to minimising interfacial adsorption *via* electrostatic interactions. As a result, this uniformity in the nucleation and growth of P23VR58H aggregates is not unexpected.

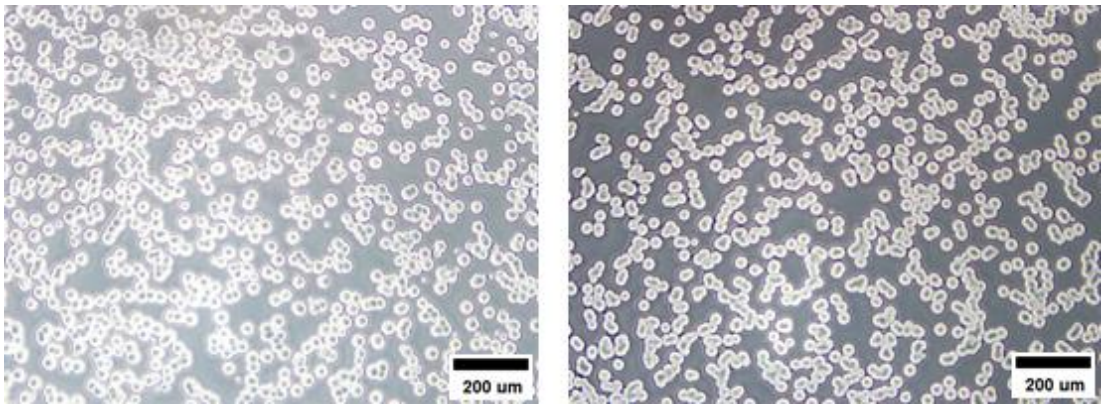

Figure 5.7: *Phase contrast microscopy images taken after 29 hours of P23VR58H aggregates formed in a chamber constructed using piranha cleaned, uncoated coverslips (d x h = 10.04 mm x 1 mm).*

Surface passivation using natural or synthetic proteins has been used to limit the adsorption of protein to a surface. Poly-lysine, which has two commonly sold enantiomers, namely poly-D-lysine (PDL) and poly-L-lysine (PLL), is a synthetic positively charged polymer often used in mammalian cell culture to promote cell adherence *via* electrostatic interactions (Mazia, Schatten et al. 1975) and in immunoassays to promote antibody binding (Stearns, Zhou et al. 2016). It has also been used in very low concentrations to coat plates used for fluorescence correlation spectroscopy measurements to minimise peptide interactions with the glass surface (Mittag, Milani et al. 2014). In a next step, coverslips were coated with PDL after

cleaning with piranha solution. The aggregates formed in this chamber formed at different rates and exhibited a polydisperse size distribution (figure 5.8). This is most likely due to a heterogeneity introduced by the quality of the surface coating which may have resulted in variations in the adsorption of PDL and, consequently, variations in the interaction of P23VR58H with the PDL coating which led to heterogeneous nucleation.

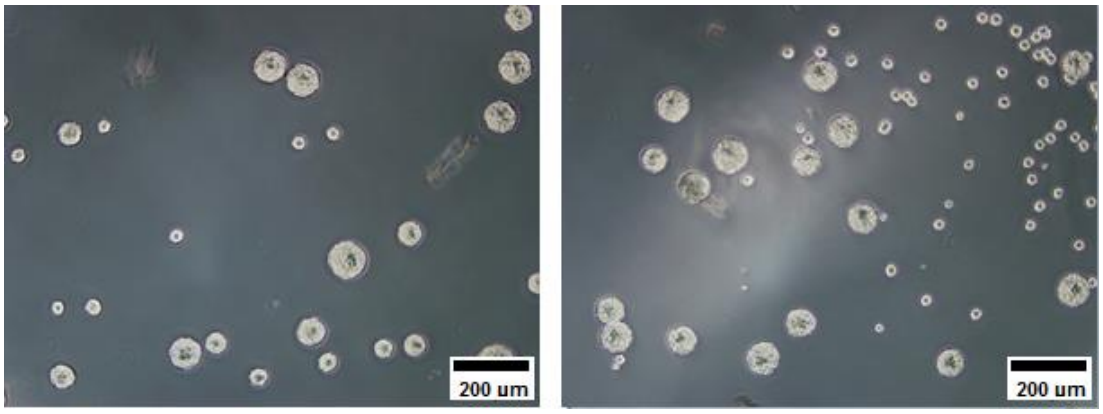

Figure 5.8: *Phase contrast microscopy images taken after 29 hours of P23VR58H aggregates formed in a chamber constructed using poly-D-lysine coated coverslips (d x h = 10.04 mm x 1 mm).*

Coating with polyethylene glycol (PEG), which is a hydrophilic polymer, is considered the most effective way of passivating glass surfaces (Elbert and Hubbell 1996; Kingshott and Griesser 1999). Coverslips were PEGylated after piranha cleaning. Figure 5.9 shows the aggregates formed in chambers constructed using PEGylated coverslips. While significantly better than for PDL coated surfaces, aggregates formed in PEGylated chambers showed variations in the size distribution, again suggesting that some surface induced heterogeneous nucleation was occurring. Again, this is most likely indicative of variations in the quality of the surface coating where patches of the PEGylated surface would have interacted differently than on un-PEGylated patches which would result in variations in the nucleation and growth of the protein.

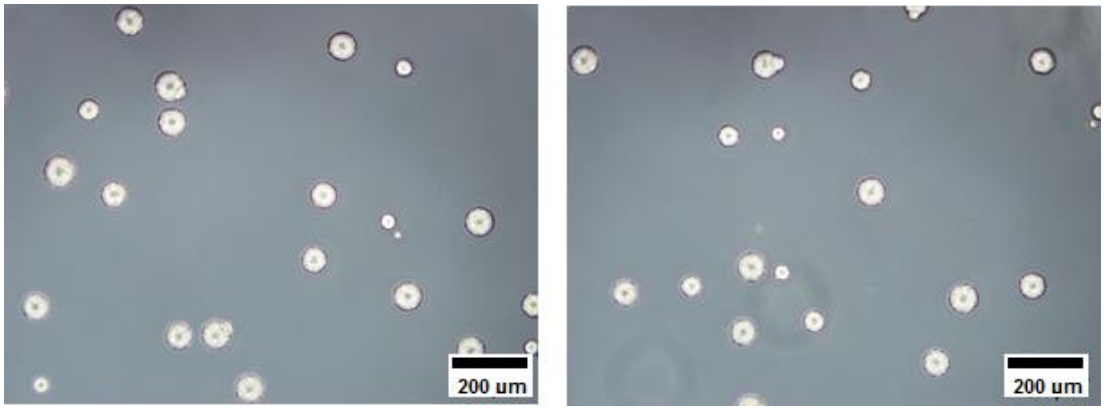

Figure 5.9: *Phase contrast microscopy images taken after 29 hours of P23VR58H aggregates formed in a chamber constructed using PEGylated coverslips (d x h = 10.04 mm x 1 mm).*

We also investigated how the creation of a hydrophobic surface would influence the interaction of the protein solution with the solid interface. Glassware can be silanised to increase its hydrophobicity and is achieved by functionalising the silanol groups on the glass surface with reactive silane (figure 5.10) (Seed 2001).

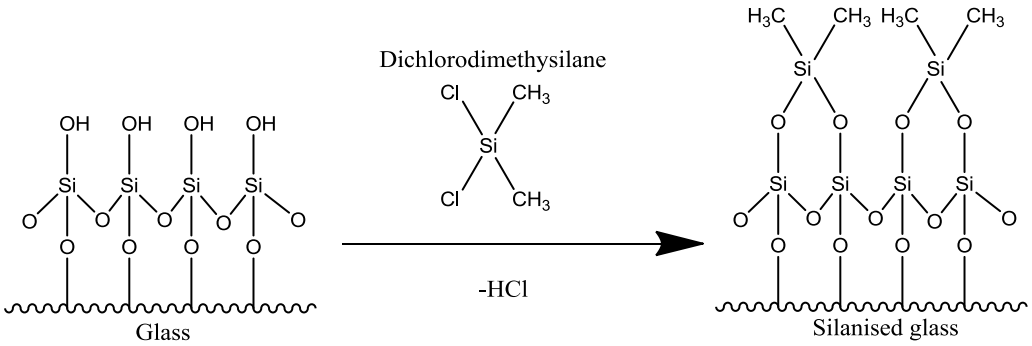

Figure 5.10: *Silanisation of glassware using dichlorodimethylsilane*

Figure 5.11 shows the aggregates formed in chambers constructed with silanised coverslips. The uniformity in size of the aggregates formed in this chamber suggests that the hydrophobic surface used served to minimise the surface interaction of the protein and hence ensured that more nucleation occurred in bulk solution, rather than at the surface interface.

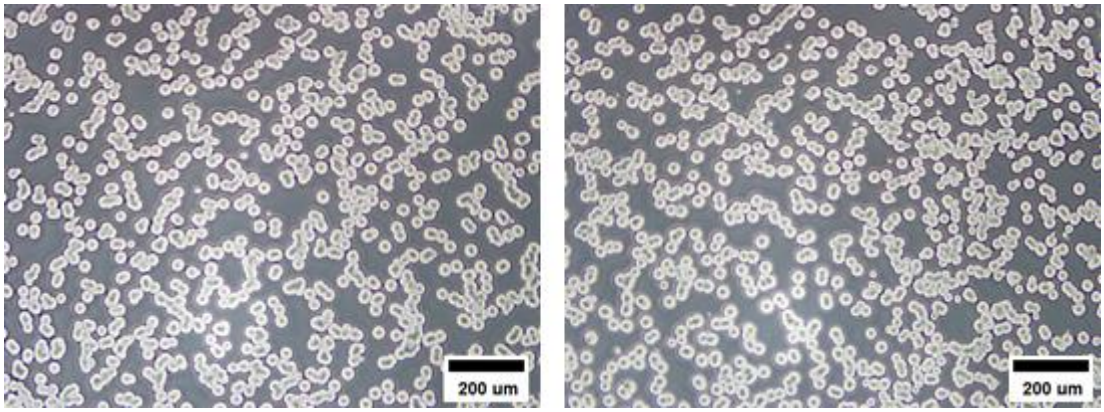

Figure 5.11: *Phase contrast microscopy images taken after 29 hours of P23VR58H aggregates formed in a chamber constructed using piranha cleaned, silanised coverslips (d x h = 10.04 mm x 1 mm).*

Thus far, piranha cleaned chambers produced the most monodisperse protein aggregates. The chambers used up to this point were ~80 µl in volume. To see if surface effects could be minimised by introducing a charged hydrophilic surface with even smaller volumes, we decreased the volume further by decreasing the chamber height. Figure 5.12 shows aggregates formed uniformly in this chamber suggesting that the interaction with the charged, hydrophilic surface was minimised using piranha cleaning (i.e. no further heterogeneity was introduced by increasing the surface to volume ratio).

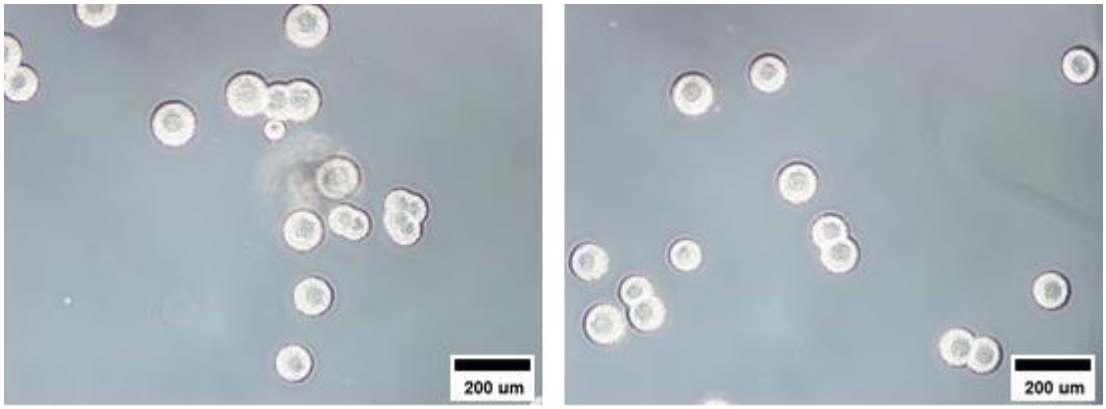

Figure 5.12 *Phase contrast microscopy images taken after 29 hours of P23VR58H aggregates formed in a chamber constructed using piranha cleaned, silanised coverslips (d x h = 10.04 mm x 0.5 mm).*

5.2.1.3 Kinetics of aggregate growth

Surface conditions that minimised interfacial protein interactions and thus gave rise to aggregates with similar rates of nucleation were analysed by measuring the early stage growth. The kinetics of P23VR58H growth were monitored for the following conditions where surfaces appeared to promote more homogeneous nucleation in the

bulk: piranha cleaned, uncoated chambers (d x h = 10.04 mm x 1 mm), piranha cleaned, uncoated chambers (d x h = 10.04 mm x 0.5 mm) and piranha cleaned, silanised chambers (d x h = 10.04 mm x 1 mm). Average aggregate diameter was plotted as a function of time (T) for each condition (figure 5.13). T_0 was taken as the time the chamber was equilibrated to 30 °C.

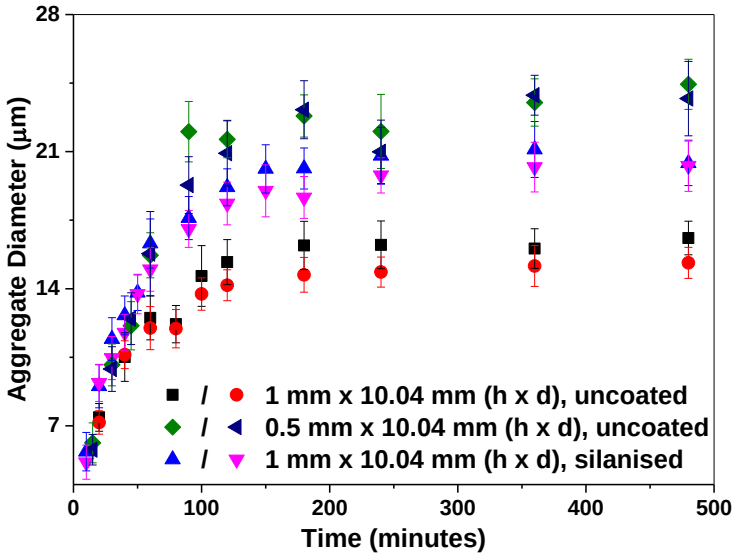

Figure 5.13: *Growth of aggregates between T_0 and T = 500 minutes.*

While the initial growth rate was apparently unaffected by the surface treatment used (figure 5.14; table 5.2), the final size of the aggregates formed was rather different for different surface treatments. Aggregates stop growing once equilibrium between the aggregate and the monomer solution is established (James, Quinn et. al 2015). Since the monomer in the bulk is depleted as aggregate growth progresses, the final size will be determined by the number of nuclei initially formed. However, the total volume occupied by the aggregates formed at a particular timepoint for each surface treatment is very similar, even for different particle sizes.

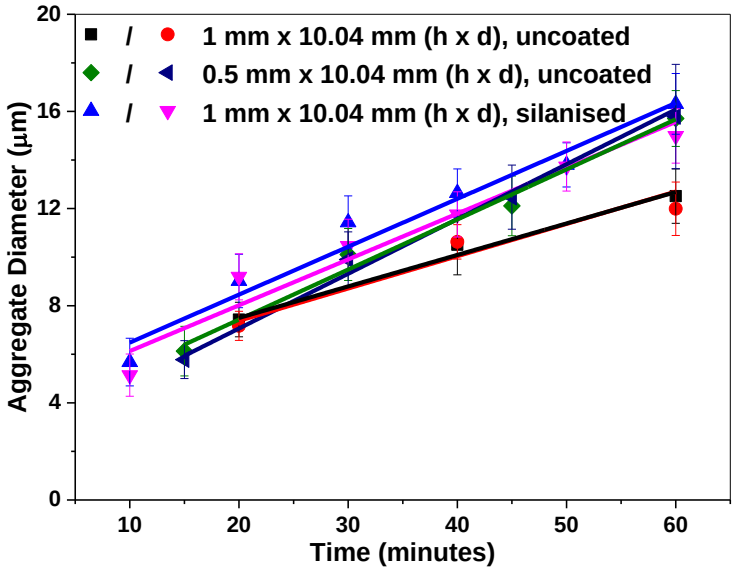

Figure 5.14: *Growth of aggregates between T_0 and T = 60 minutes.*

Table 5.2: *Growth rates of P23VR58H aggregates for each repeat under different experimental conditions.*

Condition	Slope ($\mu m\ min^{-1}$)	Intercept
1 mm x 10.04 mm (h x d), uncoated	(1) 0.130 (2) 0.133	(1) 0.012 (2) 0.032
0.5 mm x 10.04 mm (h x d), uncoated	(1) 0.206 (2) 0.225	(1) 0.017 (2) 0.016
1 mm x 10.04 mm (h x d), silanised	(1) 0.197 (2) 0.189	(1) 0.019 (2) 0.021

The growth rates of the aggregates, as indicated by the slopes, are very similar for the conditions monitored. We wanted to compare the growth rates for different conditions listed above with growth rates for the concentrations of P23VR58H determined from the earlier measurements in larger volumes. To do this, the particle growth for each concentration of P23VR58H used in the earlier work (James 2015) was analysed in the same way (figure 5.15). The growth rates determined from this linear fit are given in table 5.3.

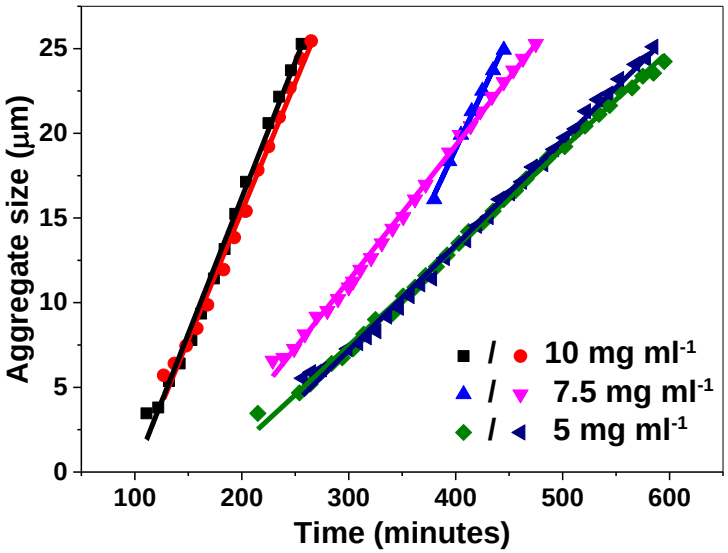

Figure 5.15: *Growth of P23VR58H aggregates for different concentrations of P23VR58H.*

Table 5.3: *Growth rates of P23VR58H aggregates for each repeat under different concentrations.*

Condition	Slope ($\mu m\ min^{-1}$)	Intercept
10 mg ml^{-1}	(1) 0.160 (2) 0.152	(1) -15.901 (2) -14.988
7.5 mg ml^{-1}	(1) 0.135 (2) 0.078	(1) -34.749 (2) -12.620
5 mg ml^{-1}	(1) 0.058 (2) 0.062	(1) -9.970 (2) -11.415

The data presented in table 5.2 and 5.3 suggest that the growth rate of P23VR58H at 10 mg ml^{-1} is similar regardless of the surface to volume ratio when the coverslips are treated by either cleaning in piranha solution or by silanising the surface – indicating that these methods are sufficient to allow nucleation and growth to be measured in smaller volumes. The negative intercept shown in table 5.3 is an artefact of an induction period.

The next question is to consider the mechanism by which particle growth proceeds. Ostwald ripening occurs in the late stages of a first order phase transition in systems where there is competitive growth between clusters which drives the larger clusters to "consume" smaller clusters to minimise the interfacial free energy (Streets and Quake 2010). The rate of growth of the diameter associated with Ostwald ripening shows a power-law time dependence equal to $t^{1/3}$. For our data, the power law time

dependence is approximately $t^{1/3}$, however we do not see any differential in the particle sizes that would normally be associated with Ostwald ripening (at least on the length scales of our measurements), and it is more likely that particle growth occurs *via* monomer addition from the bulk after the initial nucleation event. However, measurements at earlier time points will be required to confirm this.

The majority of nucleation and growth that occurs in protein solutions occurs *via* heterogeneous nucleation (Buell 2017). However, this is generally associated with heterogeneous particle sizes when nucleation is a low probability event. In this case, nucleation and growth occurs as a result of a first order phase transition, which offers the possibility that homogeneous nucleation may occur. In small volume systems, where more heterogeneous nucleation is expected due to the large surface area of glass, polydispersity in the particle sizes is observed. However, when the glass surface is treated to minimise the surface induced nucleation, monodisperse particle sizes grow. This seems consistent with a homogeneous nucleation process, but further work will be required to confirm this.

5.2.2 P23TR36S expression in a HEK293T/17 mammalian cells

Given how strongly heterogeneous surfaces influenced the phase behaviour of the P23VR58H double mutant, as a next step we examined the expression of a protein that forms condensed phases at low concentrations at physiological temperature, namely P23TR36S. Plasmid DNA coding for this double mutant was used for transfection of a mammalian cell line (HEK293T/17) to determine if the propensity of this double mutant to form crystalline solids at physiological temperature is conserved when expressed in a mammalian cell line.

Plasmid DNA that coded for the HGD-EmGFP fusion protein and designed for propagation in *E. coli* (with optimal expression in mammalian cells) was synthesised by GeneArt (ThermoScientific, Germany). The P23TR36S double mutant was obtained by site directed mutagenesis of HGD plasmid DNA (section 2.2.4). The incorporation of these mutations was confirmed by sequencing which shows the replacement of a Pro (CCC) with a Thr (ACC) and an Arg (CGC) with a Ser (AGC) (figure 5.16).

```
HGD      AGCGACCACCCCAACCTGCAGCCCTACTTGAGCCGCTGCAACTCGGCGCGCGTGGACAGC
         |||||||||  ||||||||||||||||||||||||||||||||||||||  ||||||||||
P23TR36S AGCGACCACACCAACCTGCAGCCCTACTTGAGCCGCTGCAACTCGGCGAGCGTGGACAGC
```

Figure 5.16: *Sequence of P23TR36S plasmid DNA aligned with wild type HGD using BLAST (NCBI).*

Transfection of HEK293T/17 cells was achieved using Lipofectamine 2000 in either 4 or 8 well cell imaging chambers coated with PDL to promote cell adherence (section 2.9.5). Transfection efficiency was typically ~70% (figure 5.17).

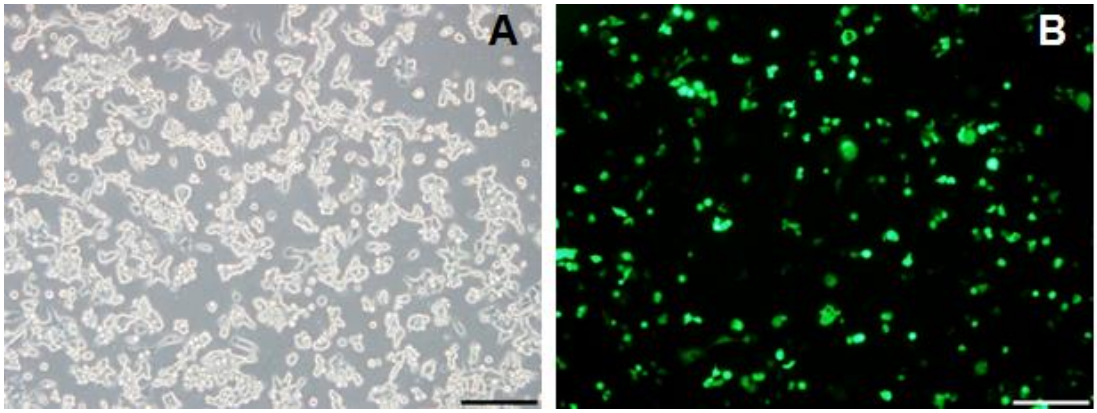

Figure 5.17: *Phase contrast (A) and fluorescence (B) microscopy images of HEK293T/17 cells after transfection with EmGFP-P23TR36S plasmid DNA. Scale bar = 200 μm.*

Previously, EmGFP, HGD fused with EmGFP (EmGFP-HGD) and the P23T single mutant of HGD fused with EmGFP (EmGFP-P23T) were expressed in a mammalian cell line (HEK293T/17) to monitor expression and self-assembly of these proteins inside a more biologically relevant environment (Migas, Quinn et al. 2017). EmGFP and EmGFP-HGD were both evenly distributed within the cytoplasm and no protein condensation was observed (figure 5.18a&b). However, HEK293T/17 cells expressing EmGFP-P23T fusion protein were shown to have regions of greater fluorescence intensity indicating a condensed protein state (figure 5.18c&e).

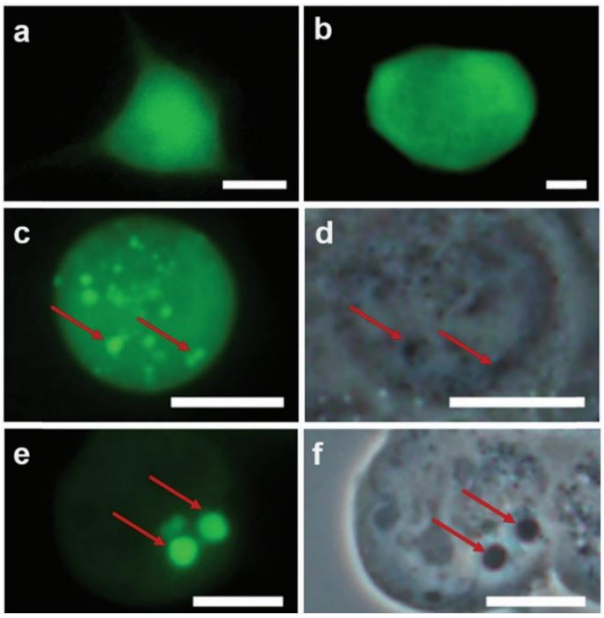

Figure 5.18: *Phase contrast and fluorescence microscopy images of HEK293T/17 cells expressing EmGFP (a), EmGFP-HGD (b), and EmGFP-P23T (c–f). Scale bar = 10 μm. Taken from Migas, Quinn et al. 2017.*

Using the same approach, HEK293T/17 cells were used to express EmGFP fused with P23TR36S (EmGFP-P23TR36S) to investigate the effect the intracellular environment had on the phase behaviour of this double mutant protein.

Phase contrast and fluorescence microscopy images were taken of transiently transfected cells every 24 hours after transfection to monitor aggregation within cells for a total 96 hours (4 days) (figures 5.19–5.22). This was to determine if there was any correlation between aggregate size and time. Regions of increased fluorescence intensity are indicative of the formation of a condensed protein state within the cell. An increase in fluorescence intensity does not always correspond to an increase in the difference between refractive indices of the condensed state and the intracellular environment. However, in some phase contrast images, these aggregates could be distinguished.

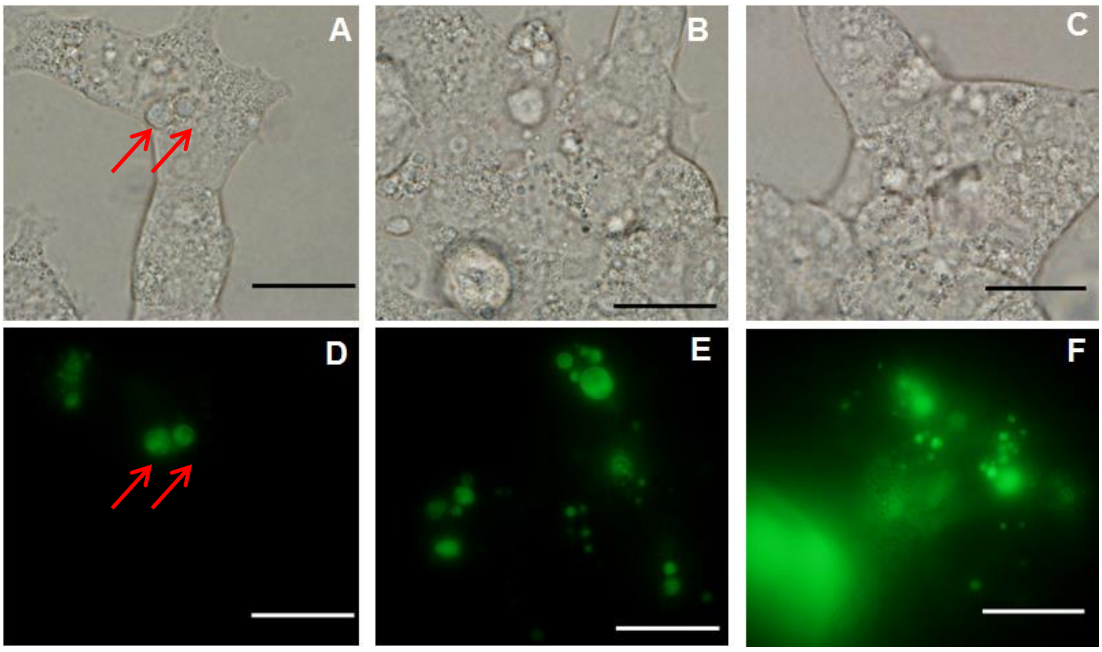

Figure 5.19: *Phase contrast (A–C) and corresponding fluorescence microscopy (D–F) images of HEK293T/17 mammalian cells expressing P23TR36S double mutant of HGD taken 1 day post transfection. Scale bar = 20 μm.*

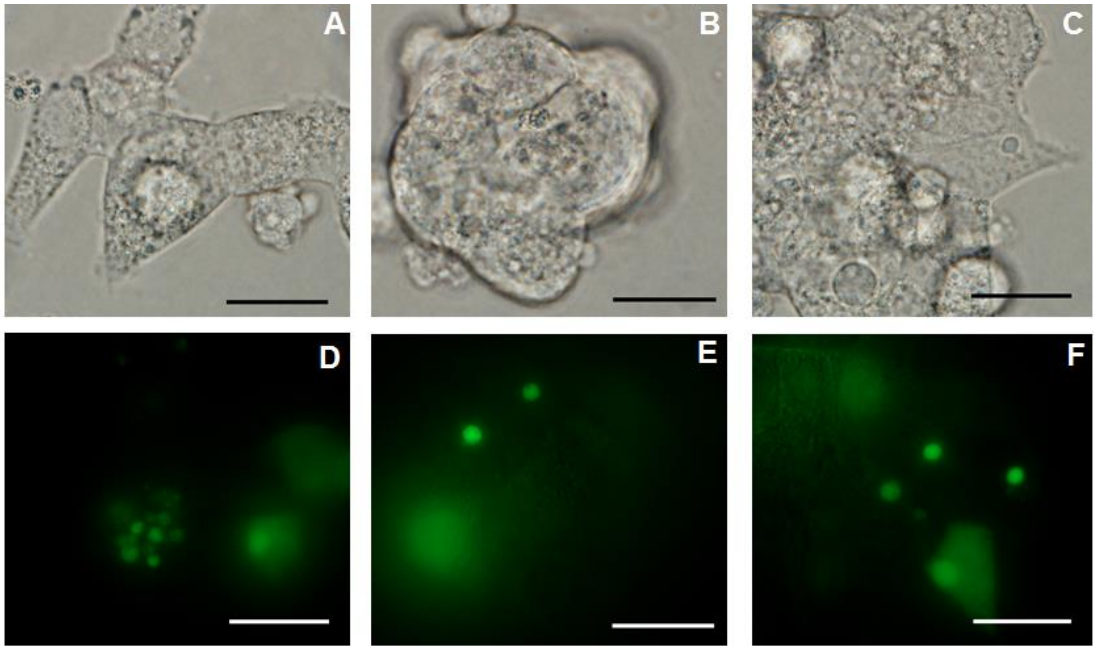

Figure 5.20: *Phase contrast (A–C) and corresponding fluorescence microscopy (D–F) images of HEK293T/17 mammalian cells expressing P23TR36S double mutant of HGD taken 2 days post transfection. Scale bar = 20 μm.*

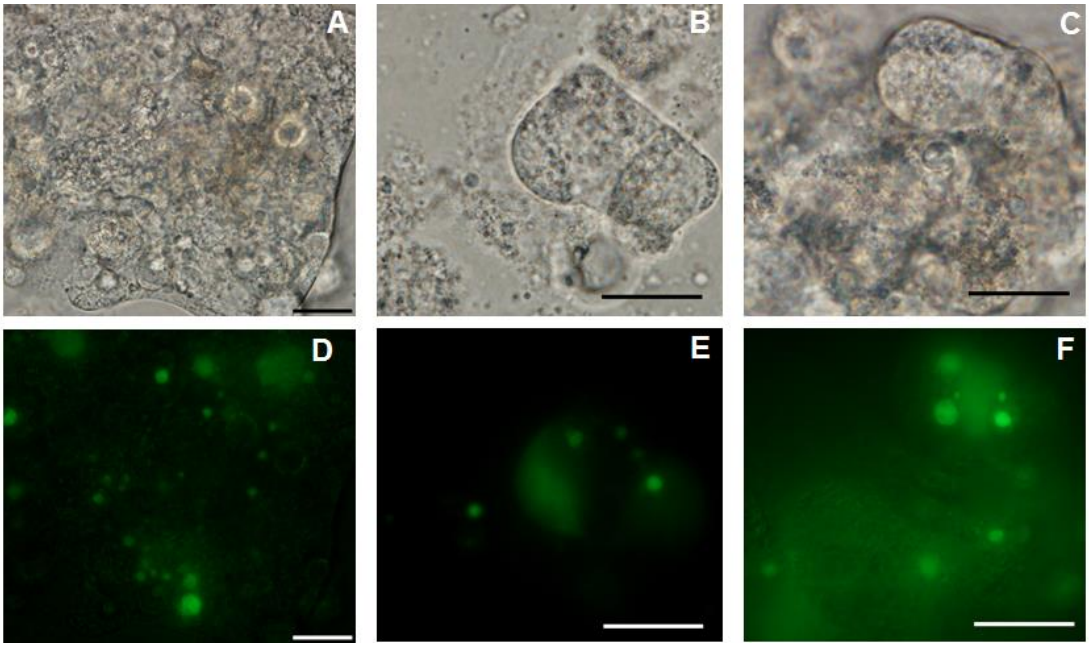

Figure 5.21: *Phase contrast (A–C) and corresponding fluorescence microscopy (D–F) images of HEK293T/17 mammalian cells expressing P23TR36S double mutant of HGD taken 3 days post transfection. Scale bar = 20 μm.*

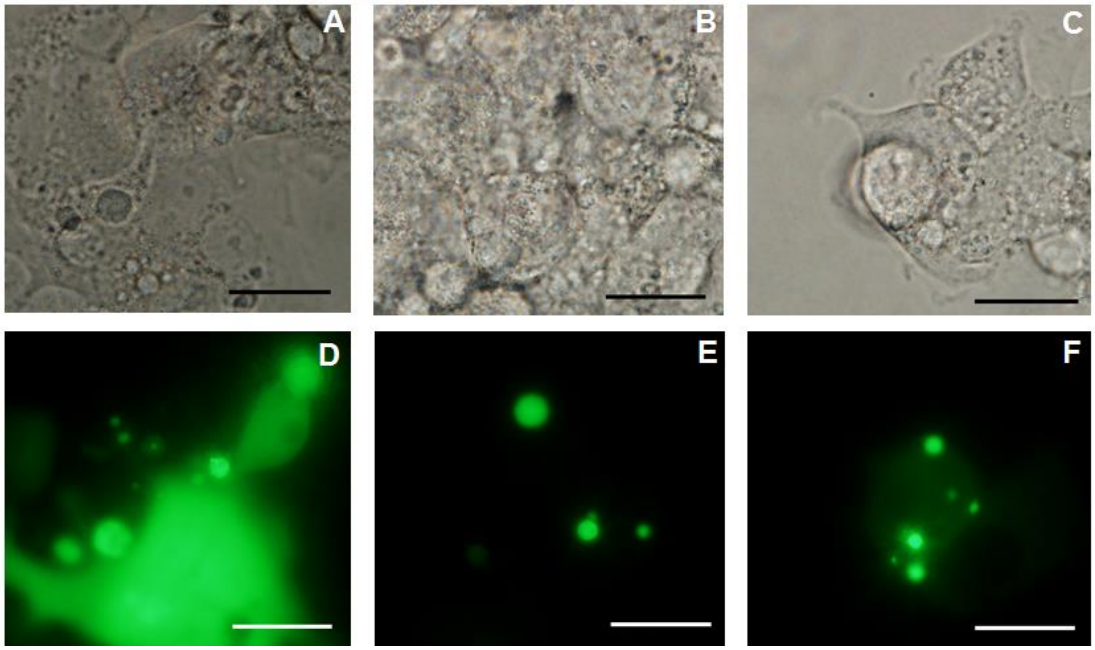

Figure 5.22: *Phase contrast (A–C) and corresponding fluorescence microscopy (D–F) images of HEK293T/17 mammalian cells expressing P23TR36S double mutant of HGD taken 4 days post transfection. Scale bar = 20 μm.*

5.2.2.2 Summary of time resolved study on HEK293T/17 cells transfected with P23TR36S

Figure 5.23 shows the normalised frequency of the size distributions of aggregates formed in cells from days 1 to 4 after transfection and table 5.4 indicates the aggregate statistics from 1 to 4 days following transfection of HEK293T/17 with P23TR36S. As expected, due to the heterogeneous intracellular environment, there was variation in the size distribution of aggregates formed in cells. The majority of aggregates formed in cells between 1 and 4 days after transfection were between 1 and 2 μm in size. The largest aggregates were seen in cells, 2 days after transfection. After this time, an increased amount of cell debris was seen in the culture wells indicating cell death. Cells respond to stresses to either ensure the cell's survival or eliminate damaged cells (Fulda, Gorman et al. 2010). Apoptosis refers to programmed cell death while necrosis refers to unregulated cell death (Elmore 2007). Necrosis can occur in response to cellular stress and is morphologically characterised by a swelling in the cell volume and organelles which in turn lead to cell rupture (Fulda, Gorman et al. 2010). Since the expression of P23TR36S and its subsequent self-assembly within the cell is not a normal feature of biological function, this may have led to cell necrosis.

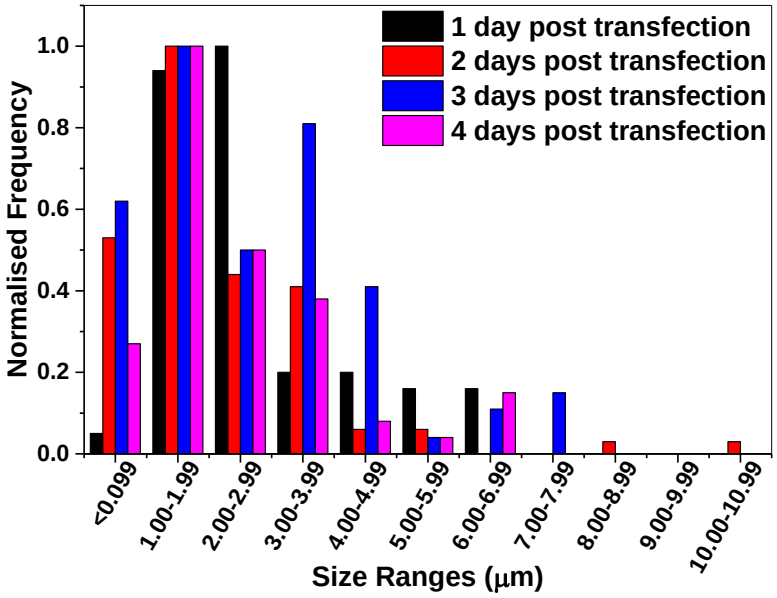

Figure 5.23: *Collated histograms representing the size ranges of the aggregates formed in HEK293T/17 cells 1, 2, 3 and 4 days post transfection.*

Table 5.4: *Statistics on aggregates measured.*

	No. aggregates measured	**Mean size *(μm)***	**Standard deviation *(μm)***	**Minimum diameter *(μm)***	**Maximum diameter *(μm)***
1 day	51	2.66	1.40	0.62	6.54
2 days	87	2.14	1.60	0.42	10.08
3 days	93	2.73	1.75	0.42	7.69
4 days	59	2.11	1.17	0.42	6.86

5.3 Conclusion

Two double mutants of HGD (P23VR58H and P23TR36S) were used to investigate how environmental conditions influence the nucleation and growth of protein condensed phases. These double mutants were used since they are known to form condensed phases with specific morphologies at certain temperatures and concentrations.

The kinetics of the P23VR58H double mutant of HGD, which forms spherical aggregates in 0.1 M sodium phosphate at pH 7.0, were previously investigated in large volumes. Here, the aggregation of P23VR58H has been analysed in smaller volumes where surface effects on nucleation and growth of aggregates are more prevalent. Different surface functionalisation strategies were explored as a means of minimising the surface effects on aggregation and to promote homogeneous nucleation. The growth rates for P23VR58H aggregates were then monitored for the conditions where nucleation proceeded uniformly and resulted in a monodisperse size distribution of aggregates which was consistent with a homogeneous nucleation process. These growth rates were compared to growth rates for previous aggregation studies done using P23VR58H in larger volumes. The growth rates in both volumes at concentrations of 10 mg ml^{-1} were very similar.

The P23TR36S double mutant, which forms rhombic crystals at low concentrations and physiological temperature in 0.1 M sodium phosphate buffer at pH 7.0, was fused with EmGFP and expressed in a mammalian cell line to investigate the influence of the intracellular environment on protein nucleation and growth. Cells transfected with the double mutant were imaged every 24 hours following transfection to monitor EmGFP-P23TR36S aggregation. The intracellular environment did not affect the propensity of the double mutant to form a condensed

state. Unsurprisingly, the heterogeneous intracellular environment gave rise to condensed states of different sizes. The morphology of these condensed states was different to that observed under 0.1 M sodium phosphate pH 7.0 with 20 mM DTT buffer conditions which may be due to the presence of other proteins in the cell.

Summary and final conclusions

This work focusses on different factors that influence protein phase behaviour. This was explored *via* modifications (both chemical modification and mutagenesis) to the surface of human γD-crystallin (HGD) and also by exploring the impact of environment on nucleation and growth of condensed protein phases.

Chemical modification of HGD by fluorescent labelling of specific amino acids (Lys-2 and Cys-110) on the surface of HGD with a (hydrophobic) small molecule fluorescent label was examined using both experiments and simulations. This study found that very low modified protein compositions (as low as x_m = 0.0001) were sufficient to increase T_{ph} significantly (~14 K) and that both the position and type of fluorescent dye used influenced $T_{ph.}$ A numerical model was designed to explain the experimental observations revealed that the increase in LLPS in the presence of modified protein was due to a new increase in attraction in the system.

PEGylated biotin (hydrophilic) was also used to modify HGD at both the Cys and Lys positions on the protein surface. There was little change in T_{ph} relative to unmodified HGD when the modification was performed at the Lys position. However, T_{ph} increased significantly for thiol modified protein at a similar modified protein composition (~10 K). Neither modification had any impact on the structure of the protein relative to unmodified HGD.

Finally, the nucleation and growth of protein aggregates in solution and in cells was probed using double mutants of HGD. For solution based measurements at smaller volumes, surface effects led to polydisperse aggregate formation. To suppress surface effects, different surface treatments were investigated. The chemical nature of the surface treatment did not cause significant differences in the kinetics of aggregation but the quality of the surface modification did. Aggregate growth in cells was demonstrated for the P23TR36S mutant of HGD. Heterogeneous aggregates of GFP labelled protein were observed in cells for up to 4 days. The particles sizes observed were predominantly 1-2 μm in size. Some larger particle sizes were observed but these were very few in number and significant cell death was associated with later particle growth stages.

References

Agarwal, P. and C. R. Bertozzi (2015). "Site-Specific Antibody–Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development." Bioconjugate Chemistry **26**(2): 176-192.

Akkapeddi, P., S.-A. Azizi, et al. (2016). "Construction of homogeneous antibody-drug conjugates using site-selective protein chemistry." Chemical Science **7**(5): 2954-2963.

Alberts, B., A. Johnson, et al. (2002). Molecular Biology of the Cell. New York, Garland Science.

Amin, S., G. V. Barnett, et al. (2014). "Protein aggregation, particle formation, characterization & rheology." Current Opinion in Colloid & Interface Science **19**(5): 438-449.

An, C., G. Huang, et al. (2017). "Emerging usage of electrocoagulation technology for oil removal from wastewater: A review." Science of The Total Environment **579**: 537-556.

Anand, G., S. Sharma, et al. (2010). "Conformational Transitions of Adsorbed Proteins on Surfaces of Varying Polarity." Langmuir **26**(13): 10803-10811.

Anderson, V. J. and H. N. W. Lekkerkerker (2002). "Insights into phase transition kinetics from colloid science." Nature **416**(6883): 811-815.

Andley, U. P. (2007). "Crystallins in the eye: Function and pathology." Progress in Retinal and Eye Research **26**(1): 78-98.

Annunziata, O., O. Ogun, et al. (2003). "Observation of liquid–liquid phase separation for eye lens γS-crystallin." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 970-974.

Antos, J. M. and M. B. Francis (2006). "Transition metal catalyzed methods for site-selective protein modification." Current Opinion in Chemical Biology **10**(3): 253-262.

Asherie, N. (2004). "Protein crystallization and phase diagrams." Methods **34**(3): 266-272.

Asherie, N., A. Lomakin, et al. (1996). "Phase Diagram of Colloidal Solutions." Physical Review Letters **77**(23): 4832-4835.

Asherie, N., J. Pande, et al. (1998). "Oligomerization and phase separation in globular protein solutions." Biophysical Chemistry **75**(3): 213-227.

Atkins, P. W. (1988). Physical Chemistry. Oxford, Oxford University Press.

Banerjee, P. R., A. Pande, et al. (2011). "Cataract-associated mutant E107A of human γD-crystallin shows increased attraction to α-crystallin and enhanced light scattering." Proceedings of the National Academy of Sciences **108**(2): 574-579.

Basak, A., O. Bateman, et al. (2003). "High-resolution X-ray crystal structures of human γD crystallin (1.25 A) and the R58H mutant (1.15 A) associated with aculeiform cataract." Journal of Molecular Biology **328**: 1137-1147.

Baslé, E., N. Joubert, et al. (2010). "Protein Chemical Modification on Endogenous Amino Acids." Chemistry & Biology **17**(3): 213-227.

Behrens, S. H. and D. G. Grier (2001). "The charge of glass and silica surfaces." The Journal of Chemical Physics **115**(14): 6716-6721.

Benedek, G. B. (1971). "Theory of Transparency of the Eye." Applied Optics **10**(3): 459-473.

Benedek, G. B., J. Pande, et al. (1999). "Theoretical and experimental basis for the inhibition of cataract." Progress in Retinal and Eye Research **18**(3): 391-402.

Bhattacharjee, S. (2016). "DLS and zeta potential – What they are and what they are not?" Journal of Controlled Release **235**: 337-351.

Bianchi, E., R. Blaak, et al. (2011). "Patchy colloids: state of the art and perspectives." Physical Chemistry Chemical Physics **13**(14): 6397-6410.

Bianchi, E., J. Largo, et al. (2006). "Phase Diagram of Patchy Colloids: Towards Empty Liquids." Physical Review Letters **97**(16): 168301.

Blennow, K., M. J. de Leon, et al. (2006). "Alzheimer's disease." The Lancet **368**(9533): 387-403.

Bloemendal, H., W. de Jong, et al. (2004). "Ageing and vision: structure, stability and function of lens crystallins." Progress in Biophysics and Molecular Biology **86**(3): 407-485.

Blumlein, A. and J. J. McManus (2013). "Reversible and non-reversible thermal denaturation of lysozyme with varying pH at low ionic strength." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(10): 2064-2070.

Bonneté, F., S. Finet, et al. (1999). "Second virial coefficient: variations with lysozyme crystallization conditions." Journal of Crystal Growth **196**(2–4): 403-414.

Boutureira, O. and G. J. L. Bernardes (2015). "Advances in Chemical Protein Modification." Chemical Reviews **115**(5): 2174-2195.

Brangwynne, C. P. (2011). "Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies." Soft Matter **7**(7): 3052-3059.

Brangwynne, C. P. (2013). "Phase transitions and size scaling of membrane-less organelles." The Journal of Cell Biology **203**(6): 875-881.

Brannigan, J. A. and A. J. Wilkinson (2002). "Protein engineering 20 years on." Nat Rev Mol Cell Biol **3**(12): 964-970.

Broide, M. L., C. R. Berland, et al. (1991). "Binary-liquid phase separation of lens protein solutions." Proceedings of the National Academy of Sciences of the United States of America **88**(13): 5660-5664.

Broide, M. L., T. M. Tominc, et al. (1996). "Using phase transitions to investigate the effect of salts on protein interactions." Physical Review E **53**(6): 6325-6335.

Buell, A. K. (2017). "The Nucleation of Protein Aggregates - From Crystals to Amyloid Fibrils." International Review of Cell and Molecular Biology **329**: 187-226.

Cahn, J. W. and J. E. Hilliard (1958). "Free Energy of a Nonuniform System. I. Interfacial Free Energy." The Journal of Chemical Physics **28**(2): 258-267.

Carter, P. (1986). "Site-directed mutagenesis." Biochemical Journal **237**(1): 1-7.

Chalker, J. M., G. J. L. Bernardes, et al. (2009). "Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology." Chemistry – An Asian Journal **4**(5): 630-640.

Chandradoss, S. D., A. C. Haagsma, et al. (2014). "Surface Passivation for Single-molecule Protein Studies." Journal of Visualized Experiments : JoVE(86): 50549.

Chen, Q., P. G. Vekilov, et al. (2004). "Liquid-Liquid Phase Separation in Hemoglobins: Distinct Aggregation Mechanisms of the β6 Mutants." Biophysical Journal **86**(3): 1702-1712.

Cooper, A. (2011). Biophysical Chemistry. Cambridge UK, RSC Publishing.

Cukalevski, R., X. Yang, et al. (2015). "The Ab40 and Ab42 peptides self-assemble into separate homomolecular fibrils in binary mixtures but cross-react during primary nucleation." Chemical Science **6**(7): 4215-4233.

Dalbadie-McFarland, G., L. W. Cohen, et al. (1982). "Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function." Proceedings of the National Academy of Sciences **79**(21): 6409-6413.

Dalby, B., S. Cates, et al. (2004). "Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications." Methods **33**(2): 95-103.

Danielsson, J., X. Mu, et al. (2015). "Thermodynamics of protein destabilization in live cells." Proceedings of the National Academy of Sciences **112**(40): 12402-12407.

De Clerck, L. S., C. H. Bridts, et al. (1994). "Use of fluorescent dyes in the determination of adherence of human leucocytes to endothelial cells and the effect of fluorochromes on cellular function." Journal of Immunological Methods **172**(1): 115-124.

del Castillo, T., J. Marales-Sanfrutos, et al. (2014). "Monovinyl Sulfone β-Cyclodextrin. A Flexible Drug Carrier System." ChemMedChem **9**(2): 383-389.

Delaye, M. and A. Tardieu (1983). "Short-range order of crystallin proteins accounts for eye lens transparency." Nature **302**(5907): 415-417.

Diamantis, N. and U. Banerji (2016). "Antibody-drug conjugates - an emerging class of cancer treatment." Br J Cancer **114**(4): 362-367.

Díaz-Rodríguez, A. and B. G. Davis (2011). "Chemical modification in the creation of novel biocatalysts." Current Opinion in Chemical Biology **15**(2): 211-219.

Dill, K. A. and J. L. MacCallum (2012). "The Protein-Folding Problem, 50 Years On." Science **338**(6110): 1042.

Dorsaz, N., L. Filion, et al. (2012). "Spiers Memorial Lecture: Effect of interaction specificity on the phase behaviour of patchy particles." Faraday Discussions **159**(0): 9-21.

Dorsaz, N., G. M. Thurston, et al. (2008). "Colloidal Characterization and Thermodynamic Stability of Binary Eye Lens Protein Mixtures." The Journal of Physical Chemistry B **113**(6): 1693-1709.

Dumetz, A. C., A. M. Chockla, et al. (2008). "Protein Phase Behavior in Aqueous Solutions: Crystallization, Liquid-Liquid Phase Separation, Gels, and Aggregates." Biophysical Journal **94**(2): 570-583.

Eaton, W. A. and J. Hofrichter (1990). "Sickle Cell Hemoglobin Polymerization." Advances in Protein Chemistry **40**: 63-279.

Elbert, D. L. and J. A. Hubbell (1996). "Surface Treatments of Polymers for Biocompatibility." Annual Review of Materials Science **26**(1): 365-394.

Elmore, S. (2007). "Apoptosis: A Review of Programmed Cell Death." Toxicologic pathology **35**(4): 495-516.

Evans, P., K. Wyatt, et al. (2004). "The P23T Cataract Mutation Causes Loss of Solubility of Folded γD-Crystallin." Journal of Molecular Biology **343**(2): 435-444.

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

Feeling-Taylor, A. R., R. M. Banish, et al. (1999). "Miniaturized scintillation technique for protein solubility determinations." Review of Scientific Instruments **70**(6): 2845-2849.

Feeling-Taylor, A. R., S. T. Yau, et al. (2004). "Crystallization Mechanisms of Hemoglobin C in the R State." Biophysical Journal **87**(4): 2621-2629.

Ferrenberg, A. M. and R. H. Swendsen (1988). "New Monte Carlo technique for studying phase transitions." Physical Review Letters **61**(23): 2635-2638.

Ferrone, F. A. (2016). "Sickle cell disease: Its molecular mechanism and the one drug that treats it." International Journal of Biological Macromolecules **93, Part A**: 1168-1173.

Fink, A. L. (1998). "Protein aggregation: folding aggregates, inclusion bodies and amyloid." Folding and Design **3**(1): R9-R23.

Foffi, G., G. D. McCullagh, et al. (2002). "Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions: Application to protein crystallization." Physical Review E **65**(3): 031407.

Fulda, S., A. M. Gorman, et al. (2010). "Cellular Stress Responses: Cell Survival and Cell Death." International Journal of Cell Biology **2010**: 23.

Fusco, D. and P. Charbonneau (2016). "Soft matter perspective on protein crystal assembly." Colloids and Surfaces B: Biointerfaces **137**: 22-31.

Fusco, D., J. J. Headd, et al. (2014). "Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study." Soft Matter **10**(2): 290-302.

Gajraj, A. and R. Y. Ofoli (2000). "Effect of Extrinsic Fluorescent Labels on Diffusion and Adsorption Kinetics of Proteins at the Liquid−Liquid Interface†." Langmuir **16**(21): 8085-8094.

Galkin, O., K. Chen, et al. (2002). "Liquid–liquid separation in solutions of normal and sickle cell hemoglobin." Proceedings of the National Academy of Sciences **99**(13): 8479-8483.

Galvagnion, C., A. K. Buell, et al. (2015). "Lipid vesicles trigger α-synuclein aggregation by stimulating primary nucleation." Nat Chem Biol **11**(3): 229-234.

Gauthier, M. A. and H.-A. Klok (2008). "Peptide/protein-polymer conjugates: synthetic strategies and design concepts." Chemical Communications(23): 2591-2611.

General, I. J., R. Dragomirova, et al. (2012). "Absolute Free Energy of Binding of Avidin/Biotin, Revisited." The Journal of Physical Chemistry B **116**(23): 6628-6636.

George, A. and W. W. Wilson (1994). "Predicting Protein Crystallization from a Dilute Solution Property." Acta Cryst. **D50**: 361-365.

Giegé, R. (2013). "A historical perspective on protein crystallization from 1840 to the present day." FEBS Journal **280**(24): 6456-6497.

Glotzer, S. C. and M. J. Solomon (2007). "Anisotropy of building blocks and their assembly into complex structures." Nat Mater **6**(7): 557-562.

Graw, J. (2009). "Genetics of crystallins: Cataract and beyond." Experimental Eye Research **88**(2): 173-189.

Grzybowski, B. A., C. E. Wilmer, et al. (2009). "Self-assembly: from crystals to cells." Soft Matter **5**(6): 1110-1128.

Gu, L. and Z. Guo (2013). "Alzheimer's Ab42 and Ab40 peptides form interlaced amyloid fibrils." J Neurochem **126**: 7.

Gunton, J. D., A. Shiryayev, et al. (2007). Protein condensation: Kinetic pathways to crystallization and disease, Cambridge University Press.

Gunzenhäuser, J., R. Wyss, et al. (2014). "A Quantitative Approach to Evaluate the Impact of Fluorescent Labeling on Membrane-Bound HIV-Gag Assembly by Titration of Unlabeled Proteins." PLoS ONE **9**(12): e115095.

Haaga, J., E. Pemberton, et al. (2016). "Phase diagram of a model of the protein amelogenin." The Journal of Chemical Physics **145**(8): 085105.

Hagan, M. F. (2014). Modeling Viral Capsid Assembly. Advances in Chemical Physics: Volume 155, John Wiley & Sons, Inc.**:** 1-68.

Hagen, M. H. J. and D. Frenkel (1994). "Determination of phase diagrams for the hard-core attractive Yukawa system." The Journal of Chemical Physics **101**(5): 4093-4097.

Hai, P., X. Yifan, et al. (2015). "A simple procedure to improve the surface passivation for single molecule fluorescence studies." Physical Biology **12**(4): 045006.

Hawe, A., M. Sutter, et al. (2008). "Extrinsic Fluorescent Dyes as Tools for Protein Characterization." Pharmaceutical Research **25**(7): 1487-1499.

Haynes, C. A. and W. Norde (1994). "Globular proteins at solid/liquid interfaces." Colloids and Surfaces B: Biointerfaces **2**(6): 517-566.

Hejtmancik, J. F. (2008). "Congenital cataracts and their molecular genetics." Seminars in Cell & Developmental Biology **19**(2): 134-149.

Herman, B. (1998). Fluorescence Microscopy. Oxon, Taylor & Francis.

Hillger, F., D. Nettels, et al. (2007). "Detection and Analysis of Protein Aggregation with Confocal Single Molecule Fluorescence Spectroscopy." Journal of Fluorescence **17**(6): 759-765.

Hofer, T., J. D. Thomas, et al. (2008). "An engineered selenocysteine defines a unique class of antibody derivatives." Proceedings of the National Academy of Sciences of the United States of America **105**(34): 12451-12456.

Hong, P., S. Koza, et al. (2012). "Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates." Journal of Liquid Chromatography & Related Technologies **35**(20): 2923-2950.

Horwitz, J., M. P. Bova, et al. (1999). "Lens [alpha]-crystallin: Function and structure." Eye **13**(3b): 403-408.

Hsieh, P.-C. and R. Vaisvila (2013). Protein Engineering: Single or Multiple Site-Directed Mutagenesis. Enzyme Engineering: Methods and Protocols. J. C. Samuelson. Totowa, NJ, Humana Press**:** 173-186.

Hua, B., K. Young Han, et al. (2014). "An Improved Surface Passivation Method for Single-Molecule Studies." Nature Methods **11**(12): 1233-1236.

Huang, B., M. Bates, et al. (2009). "Super-Resolution Fluorescence Microscopy." Annual Review of Biochemistry **78**(1): 993-1016.

Hunt, J. A. and V. M. Ingram (1958). "Allelomorphism and the chemical differences of the human haemoglobins A, S and C." Nature **181**(4615): 1062-1063.

Iqbal, K., F. Liu, et al. (2016). "Tau and neurodegenerative disease: the story so far." Nat Rev Neurol **12**(1): 15-27.

Iqbal, K., F. Liu, et al. (2010). "Tau in Alzheimer Disease and Related Tauopathies." Current Alzheimer research **7**(8): 656-664.

Irving, S. S., B. G. Harwood, et al. (1982). "Thiol-beta-lactamase: Replacement of the active-site serine of RTEM

beta-lactamase by a cysteine residue." Proc Natl Acad Sci USA **79**: 4.

Ishimoto, C. and T. Tanaka (1977). "Critical Behavior of a Binary Mixture of Protein and Salt Water." Physical Review Letters **39**(8): 474-477.

Israelachvili, J. (1992). Intermolecular & Surface Forces. Suffolk, Academic Press.

James, S. (2015). The self-assembly of proteins: Probing anisotropic protein-protein interactions using phase diagrams. Doctorate, National University of Ireland Maynooth.

James, S. and J. J. McManus (2012). "Thermal and Solution Stability of Lysozyme in the Presence of Sucrose, Glucose, and Trehalose." The Journal of Physical Chemistry B **116**(34): 10182-10188.

James, S., M. K. Quinn, et al. (2015). "The self assembly of proteins; probing patchy protein interactions." Physical Chemistry Chemical Physics **17**(7): 5413-5420.

Ji, F., J. Jung, et al. (2013). "The Human W42R γD-Crystallin Mutant Structure Provides a Link between Congenital and Age-related Cataracts." The Journal of Biological Chemistry **288**(1): 99-109.

Jones, R. A. L. (2002). Soft Condensed Matter. Oxford, Oxford University Press.

Joshi, N. S., L. R. Whitaker, et al. (2004). "A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation." Journal of the American Chemical Society **126**(49): 15942-15943.

Karthika, S., T. K. Radhakrishnan, et al. (2016). "A Review of Classical and Nonclassical Nucleation Theories." Crystal Growth & Design **16**(11): 6663-6681.

Kasten, F. H. (1999). Introduction to Fluorescent Probes: Properties, History and Applications. Fluorescent and Luminescent Probes for Biological Activity D. B. Sattelle. London, W.T. Mason.

Kelly, S. M., T. J. Jess, et al. (2005). "How to study proteins by circular dichroism." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1751**(2): 119-139.

Kelly, S. M. and N. C. Price (2000). "The Use of Circular Dichroism in the Investigation of Protein Structure and Function." Current Protein & Peptide Science **1**(4): 349-384.

Kern, N. and D. Frenkel (2003). "Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction." The Journal of Chemical Physics **118**(21): 9882-9889.

Kingshott, P. and H. J. Griesser (1999). "Surfaces that resist bioadhesion." Current Opinion in Solid State and Materials Science **4**(4): 403-412.

Kmoch, S., J. Brynda, et al. (2000). "Link between a novel human γD-crystallin allele and a unique cataract phenotype explained by protein crystallography." Human Molecular Genetics **9**(12): 1779-1786.

Kueltzo, L. A., B. Ersoy, et al. (2003). "Derivative absorbance spectroscopy and protein phase diagrams as tools for comprehensive protein characterization: A bGCSF case study." Journal of Pharmaceutical Sciences **92**(9): 1805-1820.

Kulkarni, A. M., N. M. Dixit, et al. (2003). "Ergodic and non-ergodic phase transitions in globular protein suspensions." Faraday Discussions **123**(0): 37-50.

Kuznetsova, I. M., K. K. Turoverov, et al. (2014). "What Macromolecular Crowding Can Do to a Protein." International Journal of Molecular Sciences **15**(12): 23090-23140.

Leckband, D. and J. Israelachvili (2001). "Intermolecular forces in biology." Quarterly Reviews of Biophysics **34**(2): 105-267.

Levinthal, C. (1968). "Are there pathways for protein folding?" J. Chim. Phys. **65**: 44-45.

Li, W., B. A. Persson, et al. (2015). "Charge-Induced Patchy Attractions between Proteins." The Journal of Physical Chemistry B **119**(2): 503-508.

Liu, C., N. Asherie, et al. (1996). "Phase separation in aqueous solutions of lens γ-crystallins: Special role of γ_s." Proceedings of the National Academy of Sciences **93**: 377-382.

Liu, C., A. Lomakin, et al. (1995). "Phase Separation in Multicomponent Aqueous-Protein Solutions." The Journal of Physical Chemistry **99**(1): 454-461.

Liu, H., S. Garde, et al. (2005). "Direct determination of phase behavior of square-well fluids." The Journal of Chemical Physics **123**(17): 174505.

Liu, H., S. K. Kumar, et al. (2007). "Vapor-liquid coexistence of patchy models: Relevance to protein phase behavior." The Journal of Chemical Physics **127**(8): 084902.

Lodish, H., A. Berk, et al. (2000). Molecular Cell Biology. New York, WH Freeman.

Lomakin, A., N. Asherie, et al. (1996). "Monte Carlo study of phase separation in aqueous protein solutions." The Journal of Chemical Physics **104**(4): 1646-1656.

Lomakin, A., N. Asherie, et al. (1999). "Aeolotopic interactions of globular proteins." Proceedings of the National Academy of Sciences **96**(17): 9465-9468.

Mackay, D., U. Andley, et al. (2004). "A missense mutation in the γD crystallin gene (CRYGD) associated with autosomal dominant "coral-like" cataract linked to chromosome 2q." Molecular Vision **10**: 8.

Malfois, M., F. Bonneté, et al. (1996). "A model of attractive interactions to account for fluid–fluid phase separation of protein solutions." The Journal of Chemical Physics **105**(8): 3290-3300.

Mazia, D., G. Schatten, et al. (1975). "Adhesion of cells to surfaces coated with polylysine. Applications to electron microscopy." The Journal of Cell Biology **66**(1): 198-200.

McManus, J. J., P. Charbonneau, et al. (2016). "The physics of protein self-assembly." Current Opinion in Colloid & Interface Science **22**: 73-79.

McManus, J. J., A. Lomakin, et al. (2007). "Altered phase diagram due to a single point mutation in human γD-crystallin." Proceedings of the National Academy of Sciences **104**(43): 16856-16861.

McPherson, A. and B. Cudney (2006). "Searching for silver bullets: An alternative strategy for crystallizing macromolecules." Journal of Structural Biology **156**(3): 387-406.

Messina-Baas, O. M., L. M. Gonzalez-Huerta, et al. (2006). "Two affected siblings with nuclear cataract associated with a novel missense mutation in the CRYGD gene." Molecular Vision **12**: 6.

Migas, U. M., M. K. Quinn, et al. (2017). "Protein self-assembly following in situ expression in artificial and mammalian cells." Integrative Biology **9**(5): 444-450.

Mittag, J. J., S. Milani, et al. (2014). "Simultaneous measurement of a range of particle sizes during Aβ1–42 fibrillogenesis quantified using fluorescence correlation spectroscopy." Biochemical and Biophysical Research Communications **448**(2): 195-199.

Möller, J., S. Grobelny, et al. (2014). "Reentrant Liquid-Liquid Phase Separation in Protein Solutions at Elevated Hydrostatic Pressures." Physical Review Letters **112**(2): 028101.

Muschol, M. and F. Rosenberger (1995). "Interactions in undersaturated and supersaturated lysozyme solutions: Static and dynamic light scattering results." The Journal of Chemical Physics **103**(24): 10424-10432.

Muschol, M. and F. Rosenberger (1997). "Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization." The Journal of Chemical Physics **107**(6): 1953-1962.

Nagel, R. L. and C. Lawrence (1991). "The distinct pathobiology of sickle cell-hemoglobin C disease: Therapeutic implications." Hematology/Oncology Clinics of North America **5**(3): 433-451.

Nakanishi, K., T. Sakiyama, et al. (2001). "On the adsorption of proteins on solid surfaces, a common but very complicated phenomenon." Journal of Bioscience and Bioengineering **91**(3): 233-244.

Nandrot, E., C. Slingsby, et al. (2003). "Gamma-D crystallin gene (CRYGD) mutation causes autosomal dominant congenital cerulean cataracts." Journal of Medical Genetics **40**(4): 262-267.

Nicolaides, D. and A. D. Bruce (1988). "Universal configurational structure in two-dimensional scalar models." Journal of Physics A: Mathematical and General **21**(1): 233.

Noro, M. G., N. Kern, et al. (1999). "The role of long-range forces in the phase behavior of colloids and proteins." EPL (Europhysics Letters) **48**(3): 332.

Ostuni, E., R. G. Chapman, et al. (2001). "A Survey of Structure−Property Relationships of Surfaces that Resist the Adsorption of Protein." Langmuir **17**(18): 5605-5620.

Ouberai, M. M., K. Xu, et al. (2014). "Effect of the interplay between protein and surface on the properties of adsorbed protein layers." Biomaterials **35**(24): 6157-6163.

Oya, T., N. Hattori, et al. (1999). "Methylglyoxal Modification of Protein: CHEMICAL AND IMMUNOCHEMICAL CHARACTERIZATION OF METHYLGLYOXAL-ARGININE ADDUCTS." Journal of Biological Chemistry **274**(26): 18492-18502.

Pace, C. N., J. M. Scholtz, et al. (2014). "Forces Stabilizing Proteins." FEBS letters **588**(14): 2177-2184.

Pan, W., L. Filobelo, et al. (2009). "Viscoelasticity in Homogeneous Protein Solutions." Physical Review Letters **102**(5): 058101.

Pande, A., O. Annunziata, et al. (2005). "Decrease in Protein Solubility and Cataract Formation Caused by the Pro23 to Thr Mutation in Human γD-Crystallin." Biochemistry **44**(7): 2491-2500.

Pande, A., K. S. Ghosh, et al. (2010). "Increase in Surface Hydrophobicity of the Cataract-Associated P23T Mutant of Human γD-Crystallin Is Responsible for

Its Dramatically Lower, Retrograde Solubility." Biochemistry **49**(29): 6122-6129.

Pande, A., D. Gillot, et al. (2009). "The Cataract-Associated R14C Mutant of Human γD-Crystallin Shows a Variety of Intermolecular Disulfide Cross-Links: A Raman Spectroscopic Study." Biochemistry **48**(22): 4937-4945.

Pande, A., J. Pande, et al. (2001). "Crystal cataracts: Human genetic cataract caused by protein crystallization." Proceedings of the National Academy of Sciences **98**(11): 6116-6120.

Pande, A., J. Pande, et al. (2000). "Molecular basis of a progressive juvenile-onset hereditary cataract." Proceedings of the National Academy of Sciences **97**(5): 1993-1998.

Pande, J., A. Lomakin, et al. (1995). "Oxidation of gamma II-crystallin solutions yields dimers with a high phase separation temperature." Proceedings of the National Academy of Sciences **92**(4): 1067-1071.

Pellicane, G., D. Costa, et al. (2004). "Theory and simulation of short-range models of globular protein solutions." Journal of Physics: Condensed Matter **16**(42): S4923.

Pellicane, G., G. Smith, et al. (2008). "Molecular Dynamics Characterization of Protein Crystal Contacts in Aqueous Solutions." Physical Review Letters **101**(24): 248102.

Philo, J. S. and T. Arakawa (2009). "Mechanisms of Protein Aggregation." Current Pharmaceutical Biotechnology **10**(4): 348-351.

Polakis, P. (2016). "Antibody Drug Conjugates for Cancer Therapy." Pharmacological Reviews **68**(1): 3-19.

Pusey, M., J. Barcena, et al. (2015). "Trace fluorescent labeling for protein crystallization." Acta Crystallographica. Section F, Structural Biology Communications **71**(Pt 7): 806-814.

Pusey, P. N. and W. van Megen (1986). "Phase behaviour of concentrated suspensions of nearly hard colloidal spheres." Nature **320**(6060): 340-342.

Quang, L. J., S. I. Sandler, et al. (2014). "Anisotropic Contributions to Protein–Protein Interactions." Journal of Chemical Theory and Computation **10**(2): 835-845.

Quigley, A. and D. R. Williams (2015). "The second virial coefficient as a predictor of protein aggregation propensity: A self-interaction chromatography study." European Journal of Pharmaceutics and Biopharmaceutics **96**: 282-290.

Quinn, M. K., N. Gnan, et al. (2015). "How fluorescent labelling alters the solution behaviour of proteins." Physical Chemistry Chemical Physics **17**(46): 31177-31187.

Richards, D. P., C. Stathakis, et al. (1999). "Labeling effects on the isoelectric point of green fluorescent protein." Journal of Chromatography A **853**(1–2): 21-25.

Rivillas-Acevedo, L., A. Fernández-Silva, et al. (2015). Function, Structure and Stability of Human Gamma D Crystallins: A Review. Physical Biology of Proteins and Peptides: Theory, Experiment, and Simulation. L. Olivares-Quiroz, O. Guzmán-López and E. H. Jardón-Valadez. Cham, Springer International Publishing**:** 81-98.

Roberts, C. J. (2007). "Non-native protein aggregation kinetics." Biotechnology and Bioengineering **98**(5): 927-938.

Roberts, C. J. (2014). "Protein Aggregation and Its Impact on Product Quality." Current Opinion in Biotechnology **0**: 211-217.

Rodwell, V. W., K. M. Botham, et al. (2015). Harper's illustrated biochemistry. New York, McGraw-Hill Education LLC.

Romano, F., E. Sanz, et al. (2011). "Crystallization of tetrahedral patchy particles in silico." The Journal of Chemical Physics **134**(17): 174502.

Romano, F., P. Tartaglia, et al. (2007). "Gas–liquid phase coexistence in a tetrahedral patchy particle model." Journal of Physics: Condensed Matter **19**(32): 322101.

Roosen-Runge, F., F. Zhang, et al. (2014). "Ion-activated attractive patches as a mechanism for controlled protein interactions." Sci. Rep. **4**.

Rosenbaum, D., P. C. Zamora, et al. (1996). "Phase Behavior of Small Attractive Colloidal Particles." Physical Review Letters **76**(1): 150-153.

Ross, C. A. and S. J. Tabrizi (2011). "Huntington's disease: from molecular pathogenesis to clinical treatment." The Lancet Neurology **10**(1): 83-98.

Rovigatti, L., D. d. l. Heras, et al. (2013). "Computing the phase diagram of binary mixtures: A patchy particle case study." The Journal of Chemical Physics **138**(16): 164904.

Royall, C. P., W. C. K. Poon, et al. (2013). "In search of colloidal hard spheres." Soft Matter **9**(1): 17-27.

Sahoo, H. (2012). "Fluorescent labeling techniques in biomolecules: a flashback." RSC Advances **2**(18): 7017-7029.

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

Sear, R. P. (1999). "Phase behavior of a simple model of globular proteins." The Journal of Chemical Physics **111**(10): 4800-4806.

Seed, B. (2001). Silanizing Glassware. Current Protocols in Cell Biology, John Wiley & Sons, Inc.

Selkoe, D. J. and M. B. Podlisny (2002). "Desciphering the genetic basis of Alzheimer's disease." Ann. Rev. Genomics Human Genet. **3**: 33.

Sélo, I., L. Négroni, et al. (1996). "Preferential labeling of α-amino N-terminal groups in peptides by biotin: application to the detection of specific anti-peptide antibodies by enzyme immunoassays." Journal of Immunological Methods **199**(2): 127-138.

Serrano, M. D., O. Galkin, et al. (2001). "Are protein crystallization mechanisms relevant to understanding and control of polymerization of deoxyhemoglobin S?" Journal of Crystal Growth **232**(1): 368-375.

Sharma, K. K. and P. Santhoshkumar (2009). "Lens aging: Effects of crystallins." Biochimica et Biophysica Acta (BBA) - General Subjects **1790**(10): 1095-1108.

Shaw, D. J. (1992). Introduction to Colloid and Surface Chemistry. Oxford, UK, Butterworth-Heinemann.

Shentu, X., K. Yao, et al. (2003). "Special fasciculiform cataract caused by a mutation in the γD-crystallin gene." Molecular Vision **10**: 7.

Shiels, A. and J. F. Hejtmancik (2013). "Genetics of human cataract." Clinical Genetics **84**(2): 120-127.

Sletten, E. M. and C. R. Bertozzi (2009). "Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality." Angewandte Chemie International Edition **48**(38): 6974-6998.

Spicer, C. D. and B. G. Davis (2014). "Selective chemical protein modification." Nat Commun **5**.

Stearns, N. A., S. Zhou, et al. (2016). "The Use of Poly-L-Lysine as a Capture Agent to Enhance the Detection of Antinuclear Antibodies by ELISA." PLoS ONE **11**(9): e0161818.

Steinberg , M. H. (1999). "Management of Sickle Cell Disease." New England Journal of Medicine **340**(13): 1021-1030.

Stephanopoulos, N. and M. B. Francis (2011). "Choosing an effective protein bioconjugation strategy." Nat Chem Biol **7**(12): 876-884.

Streets, A. M. and S. R. Quake (2010). "Ostwald Ripening of Clusters during Protein Crystallization." Physical Review Letters **104**(17): 178102-178102.

Taratuta, V. G., A. Holschbach, et al. (1990). "Liquid-liquid phase separation of aqueous lysozyme solutions: effects of pH and salt identity." The Journal of Physical Chemistry **94**(5): 2140-2144.

Thevenot, P., W. Hu, et al. (2008). "SURFACE CHEMISTRY INFLUENCE IMPLANT BIOCOMPATIBILITY." Current topics in medicinal chemistry **8**(4): 270-280.

Thomson, J. A., P. Schurtenberger, et al. (1987). "Binary liquid phase separation and critical phenomena in a protein/water solution." Proceedings of the National Academy of Sciences **84**(20): 7079-7083.

Thurston, G. M. (2006). "Liquid-liquid phase separation and static light scattering of concentrated ternary mixtures of bovine α and γB crystallins." The Journal of Chemical Physics **124**(13): -.

Thurston, G. M., D. L. Hayden, et al. (1997). "Quasielastic light scattering study of the living human lens as a function of age." Current Eye Research **16**(3): 197-207.

Tokumasu, F., G. A. Nardone, et al. (2009). "Altered Membrane Structure and Surface Potential in Homozygous Hemoglobin C Erythrocytes." PLoS ONE **4**(6): e5828.

Tsien, R. Y. (1998). "The Green Fluorescent Protein." Annu Rev Biochem **67**: 45.

Tsypin, M. M. and H. W. J. Blöte (2000). "Probability distribution of the order parameter for the three-dimensional Ising-model universality class: A high-precision Monte Carlo study." Physical Review E **62**(1): 73-76.

Turecek, P. L., M. J. Bossard, et al. (2016). "PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs." Journal of Pharmaceutical Sciences **105**(2): 460-475.

Vekilov, P. G. (2010). "Nucleation." Crystal Growth & Design **10**(12): 5007-5019.

Vekilov, P. G. (2010). "Phase transitions of folded proteins." Soft Matter **6**(21): 5254-5272.

Vekilov, P. G. (2012). "Phase diagrams and kinetics of phase transitions in protein solutions." Journal of Physics: Condensed Matter **24**(19): 193101.

Vekilov, P. G. and A. A. Chernov (2003). "The Physics of Protein Crystallization." Solid State Physics **57**: 1-147.

Vekilov, P. G., A. R. Feeling-Taylor, et al. (2002). "Intermolecular interactions, nucleation, and thermodynamics of crystallization of hemoglobin C." Biophysical Journal **83**(2): 1147-1156.

Velev, O. D., E. W. Kaler, et al. (1998). "Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen." Biophysical Journal **75**(6): 2682-2697.

Veronese, F. M. (2001). "Peptide and protein PEGylation: a review of problems and solutions." Biomaterials **22**(5): 405-417.

Virnau, P. and M. Müller (2004). "Calculation of free energy through successive umbrella sampling." The Journal of Chemical Physics **120**(23): 10925-10930.

Walsh, C. T., S. Garneau-Tsodikova, et al. (2005). "Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications." Angewandte Chemie International Edition **44**(45): 7342-7372.

Walsh, G. and R. Jefferis (2006). "Post-translational modifications in the context of therapeutic proteins." Nat Biotech **24**(10): 1241-1252.

Wang, B., C. Yu, et al. (2011). "A Novel CRYGD Mutation (p.Trp43Arg) Causing Autosomal Dominant Congenital Cataract in a Chinese Family." Human Mutation **32**(1): E1939-E1947.

Wang, W., S. Nema, et al. (2010). "Protein aggregation—Pathways and influencing factors." International Journal of Pharmaceutics **390**(2): 89-99.

Wang, W. and C. J. Roberts (2010). Aggregation of therapeutic proteins. New Jersey.

Wang, Y., R. F. Latypov, et al. (2014). "Quantitative Evaluation of Colloidal Stability of Antibody Solutions using PEG-Induced Liquid–Liquid Phase Separation." Molecular Pharmaceutics **11**(5): 1391-1402.

Wang, Y., A. Lomakin, et al. (2011). "Phase separation in solutions of monoclonal antibodies and the effect of human serum albumin." Proceedings of the National Academy of Sciences **108**(40): 16606-16611.

Wang, Y., A. Lomakin, et al. (2010). "Phase behavior of mixtures of human lens proteins Gamma D and Beta B1." Proceedings of the National Academy of Sciences **107**(30): 13282-13287.

Weber, K., J. R. Pringle, et al. (1972). [1] Measurement of molecular weights by electrophoresis on SDS-acrylamide gel. Methods in Enzymology, Academic Press. **Volume 26:** 3-27.

Wentzel, N. and J. D. Gunton (2008). "Effect of Solvent on the Phase Diagram of a Simple Anisotropic Model of Globular Proteins." The Journal of Physical Chemistry B **112**(26): 7803-7809.

Whitesides, G. M. and M. Boncheva (2002). "Beyond molecules: Self-assembly of mesoscopic and macroscopic components." Proceedings of the National Academy of Sciences of the United States of America **99**(8): 4769-4774.

Whitesides, G. M. and B. Grzybowski (2002). "Self-Assembly at All Scales." Science **295**(5564): 2418.

Whitford, D. (2005). Proteins: Structure and Function. West Sussex, John Wiley & Sons Ltd.

Wilding, N. B. (1997). "Simulation studies of fluid critical behaviour." Journal of Physics: Condensed Matter **9**(3): 585.

Williamson, J., J. Goldman, et al. (2009). "Genetic Aspects of Alzheimer Disease." The neurologist **15**(2): 80-86.

Winter, G., A. R. Fersht, et al. (1982). "Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding." Nature **299**(5885): 756-758.

Wistow, G. (2012). "The human crystallin gene families." Human Genomics **6**(1): 26-26.

Wolde, P. R. t. and D. Frenkel (1997). "Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations." Science **277**(5334): 1975.

Wu, M. H. (2009). "Simple poly(dimethylsiloxane) surface modification to control cell adhesion." Surface and Interface Analysis **41**(1): 11-16.

Zhang, J., R. E. Campbell, et al. (2002). "Creating new fluorescent probes for cell biology." Nat Rev Mol Cell Biol **3**(12): 906-918.

Zhang, Y. and V. N. Gladyshev (2007). "High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis." Nucleic Acids Research **35**(15): 4952-4963.

Zheng, Q., M. F. Juette, et al. (2014). "Ultra-stable organic fluorophores for single-molecule research." Chemical Society Reviews **43**(4): 1044-1056.

Appendix

This appendix provides data tables for phase diagrams presented in Chapters 3 and 4.

Table A1: *Data table for unlabelled HGD.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
Unmodified HGD	0.0	0.1449	276.575	0.025
		0.1365	276.3	0.05
		0.1246	276.075	0.0725
		0.11235	275.8	0.1
		0.1043	275.025	0.125
		0.09765	274.875	0.125
		0.0896	274.25	0.15
		0.0847	274.075	0.295
		0.0805	273.85	0.3
		0.0749	274.15	0.65
		0.0714	273.775	0.375
		0.0651	273.425	0.425
		0.0574	272.775	0.525
		0.05355	272.525	0.625
		0.049	272.025	0.775
		0.0427	270.625	1.075
		0.0371	270	1.2
		0.0315	268.9	1.25

Table A2: *Data table for phase diagram in figure 3.7.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
Fluorescently Labelled HGD (*Amine Modified*)	0.002	0.03089	268.4	1.15
		0.035	270.05	1.35
		0.0404	272.8	1
		0.04515	272.8	1
		0.05117	274.8	0.8
		0.05904	276.1	0.4
		0.06601	274.8	0.65
		0.07219	276.25	0.8
		0.077	276.15	0.4
		0.0385	274.95	1.8
		0.035	274.7	1.55
	0.006	0.077	281.05	0.4
		0.07349	280.2	0.45
		0.07	279.05	0.5
		0.06649	278.15	0.6
		0.06299	277.95	0.6
		0.0595	277.55	0.6
		0.05599	276.65	0.8
		0.05246	276.7	0.75
		0.049	275.9	0.65
		0.04549	275.9	0.45
		0.042	275.1	0.45
		0.0385	274.25	0.6
	0.0075	0.077	287.05	0.5
		0.07	283.6	0.25
		0.0637	280.95	0.8
		0.05775	279.15	0.6
		0.05022	277.35	0.6
		0.04359	274.55	0.6
		0.0357	270.25	0.8
	0.01	0.077	292.95	0.6
		0.07105	290.7	0.85
		0.06286	286.7	0.85
		0.0574	282.1	1.55
		0.05194	279.5	0.85
		0.03578	271.9	1.65
		0.0315	270.15	1.1

Table A3: *Data table for phase diagram in figure 3.12.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
Fluorescently Labelled HGD (*Thiol Modified*)	0.0001	0.07	283.05	2.7
		0.0665	282.5	3.15
		0.063	281.25	1.4
		0.0595	280.1	2.15
		0.056	280.05	2.2
		0.0525	279.45	2.1
		0.049	278.7	1.65
		0.0455	278	2.05
		0.042	277.9	1.85
		0.0385	274.95	1.8
		0.035	274.7	1.55
	0.0002	0.07	289.05	2.2
		0.0665	288.85	2.3
		0.063	287.35	2
		0.0595	286.7	2.45
		0.056	285.5	2.85
		0.0525	284.9	3.35
		0.049	283.9	3.55
		0.0455	281.25	2.4
		0.042	282.9	1.15
		0.0385	279.2	2.55
		0.035	278.65	2.9

Table A4: *Data table for phase diagram in figure 3.18.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
FITC labelled HGD	0.0002	0.07	285.1	0.55
		0.063	283.45	1.4
		0.056	280.45	1.3
		0.049	278.9	1.45
		0.042	277.75	1.7

Table A5: *Data table for phase diagram in figure 4.8.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
Biotinylated HGD (*Amine Modified*)	0.62	0.1349	277.45	0.2
		0.1278	277.95	0.3
		0.1207	277.6	0.25
		0.1136	276.3	0.35
		0.1065	275.6	0.25
		0.0994	276.05	0.5
		0.0923	275.7	0.55
		0.0852	274.9	0.35
		0.0781	274.85	0.45
		0.071	274.4	0.45
		0.0639	273.95	0.4
		0.0568	273.45	0.3
		0.05325	273.1	0.45
	0.86	0.0703	274.35	1.1
		0.0666	274.2	0.55
		0.0629	273.135	0.05
		0.0555	273.1	0.3
		0.0518	273.07	0.02
		0.0481	273.07	0.03
		0.0444	271.85	0.2
		0.0407	271.45	0.2
		0.037	270.85	0.3

Table A6: *Data table for phase diagram in figure 4.9.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
Biotinylated HGD (*Thiol Modified*)	0.47	0.074	279.65	0.4
		0.0703	279.4	0.45
		0.0666	279	0.65
		0.0629	278.55	0.6
		0.0592	278.2	0.65
		0.0555	277.9	0.75
		0.0518	277.75	0.9
		0.0481	277.35	0.8
		0.0444	276.85	0.9
		0.0407	276.55	1.2
		0.037	275.95	0.9
	0.89	0.0777	282.95	0.4
		0.074	282.45	0.5
		0.0703	282.1	0.45
		0.0666	281.95	0.6
		0.0629	281.45	0.7
		0.0592	281.55	1.3
		0.0555	281.45	0.9
		0.0518	280.95	0.9
		0.0481	280.7	1.25
		0.0444	280.1	1.55
		0.0407	279.35	1.5
		0.037	279.7	2.25

Table A7: *Data table for phase diagram in figure 4.19.*

Modification	**Modified Protein Fraction (x_m)**	**Volume Fraction (ϕ)**	**Temperature (K)**	**Error (K)**
P23V single mutant of HGD	0.5	0.06807	273.15	0.8
		0.0644	273	0.55
		0.06037	272.65	0.6
		0.05635	272.35	0.5
		0.05232	272.15	0.6
		0.0483	271.7	0.45
		0.04428	271.05	0.5