

The role of surface modifications in directing protein self-assembly



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

Doctor of Philosophy

By

Alessandro Strofaldi, M.Sc.

Department of Chemistry

Maynooth University

February 2022

Research Supervisors:

Dr. Jennifer J. McManus, Dr. Eithne Dempsey

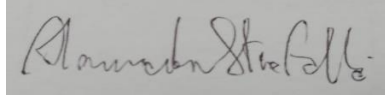
Head of Department:

Prof. Denise Rooney

Declaration

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

Signed:

A handwritten signature in black ink on a light gray rectangular background. The signature appears to be 'Alamin Abdalrhman' written in a cursive style.

Date: 28th of February 2022

Acknowledgements

Nearly five years ago I decided to start a PhD. I would have never imagined that I would encounter so many difficulties, or how nice people would have been in helping me with those.

I wish to thank my Supervisor, Dr. Jennifer McManus, not only for the opportunity granted, but also for the constant help and supervision, the teachings, and her sensitivity to our problems and lives.

I would also like to thank Dr. Eithne Dempsey - for her kindness and helpfulness as my Maynooth Supervisor - Prof. Christiane Schaffitzel, Prof. Imre Berger and Georgia Balchin - for giving me the chance to contribute to their project and for their hard work - Dr. Annela Seddon - for her patience with experiments and thorough explanations. I am also grateful to Kate Kurgan, Alex Hilditch and Rob Barringer for their advice and significant help.

I must mention and thank all the lab crew: Sarah, Judith, Matthew, Mark, Paddy, Stephanie, Jessica, Keith, Emily and Ellen - everything would have been much more difficult without your help, chats and laughs. Special thanks to My (and Cacty!), for her help and inspiration, companion from the very first day and person without whom I probably would not have gotten this far.

Big thanks to all the administrative staff and academics of both Maynooth University and University of Bristol, too many to mention all of them: Anne, Orla, Noel, Ria, Walter, Karen and also Alex, Germinal, Claudia, Andrew. Special mention goes to Diego (and Seamus!) who tried to teach me the most appropriate attitude to science.

I am also grateful to all the postgrads I have met during this path, specially to Stephen, Harlei, Luke, Amanda, Kobi, Caitlyn, Gamma, Michelle and also to Fergus, Yushi, Ioatzin, Max, Levke: thank you for the help and such great moments together.

Thanks should also go to my closest friends too: Marco, Giulia, Leonardo, Gabriele, Giuseppe and to everyone in the circle(s). Also, I cannot forget my M.Sc. friends and Supervisor, endless source of inspiration.

Finally, special thanks goes to Mom, Dad, and Silvia – constantly supporting and patiently standing everything.

Papers arising from this thesis

Strofaldi, A.; Khan, A.; McManus J. J. (2021). “*Surface Exposed Free Cysteine Suppresses Crystallization of Human γ D-Crystallin*” J. Mol. Bio., 433 (22): 167252. doi.org/10.1016/j.jmb.2021.167252

Strofaldi, A.; Seddon, A.; McManus J. J. “*Polymorphic self-assembly of Human γ D-crystallin driven by protein surface anisotropy*” (In preparation)

Abstract

Proteins self-assemble into crystals, gels, amyloid fibrils, amorphous aggregates, dense liquid droplets with implications in several fields such as biotechnology, condensation diseases and the food industry. Self-assembly processes are dictated by protein self-interactions; however, proteins are anisotropic particles, i.e. their surface is chemically heterogeneous. Hence, protein-protein interactions are strongly orientation-dependent and protein self-assembly is dramatically influenced by surface modifications. We examine this topic firstly using a model protein, Human γ D-crystallin (HGD), and then on a novel type of Virus-Like Particle (VLP), ADDomer, a promising vaccine candidate.

Human γ D-crystallin (HGD) is an eye-lens protein and along with other crystallins, it creates the refractive index gradient necessary for proper lens function. Naturally occurring surface modifications of HGD trigger aggregation/crystallization, leading to age-related or hereditary cataract onset. For these reasons, the influence of surface modifications (such as single-point mutations) on HGD self-assembly have been extensively studied making this protein a perfect model to study protein surface anisotropy and its role in directing protein assembly.

Cys-110 is the only surface exposed cysteine of HGD and hence responsible for covalent dimerization, a process that contributes to age-related cataract onset. However, the biological advantage of Cys-110 is not clear. Therefore, we mutated Cys to Met and Ser (C110M and C110S) and studied the mutant protein self-assembly. We found that C110S is recalcitrant to crystallization, while the mutant C110M crystallizes promptly; C110M crystals have the same characteristics as those of HGD, however, we suggest that the different nucleation behaviour of C110M may be due to subtle changes in the water shell of 110th site and its hydrophobicity. In conclusion, Cys-110 has the ability to suppress HGD crystallization; we speculate that the presence of Cys-110 is advantageous compared to either Met or Ser since Met enhances crystallization, hindering protein long-term stability, while Ser would decrease the HGD refractive index increment, which is essential for eye-lens function.

The effects of other single-point mutations on HGD self-assembly are known, for example R58H and R36S enhance crystallization, mutations at the 23rd site (e.g. P23V, P23T) induce the formation of retrograde solubility assemblies (i.e. they melt when cooled). It has been shown that the self-assembly behaviour of the double mutants P23VR36S, P23TR58H and P23VR58H can be predicted by those of the respective single mutants. To generalize this, we studied a novel double mutant, P23VC110M, that both enhances crystallization of the protein due to the C110M

mutation and at higher temperatures forms reversible assemblies with retrograde solubility due to the presence of the P23V mutation. P23VC110M forms both these phases and a new polymorph, needle/plate-shaped crystals, which possibly arise from new crystal contacts involving the two mutation sites. Hence, we confirmed that the self-assembly of double mutants can be predicted from the behaviour of the respective single mutants, but also more complex scenarios can arise. We also examined the morphology of the retrograde solubility assemblies formed due to mutation P23V, which unusually are perfectly spherical. Using P23VC110M, we compared the properties of the amorphous, large, reversible and spherical assemblies to other forms of protein spherical superstructures, i.e. particulates and amyloid spherulites. P23VC110M spherical assemblies are not formed by amyloid fibrils but are amorphous and cannot be ascribed to any of the protein superstructures already known in literature. We therefore suggest that this is a new class of protein assembly.

The effect of surface modifications on protein self-assembly was then assessed for a novel type of Virus-like particle (VLP), ADDomer, based on Human Adenovirus serotype-3 (HAdV-3) penton-base protein. VLPs can be modified to display on their surface multiple copies of an epitope and hence trigger immune-response against a specific disease. However, vaccines often need to be stored in extremely cold conditions, which limits their distribution in remote areas of the world. We have studied the effects of temperature on different variations of ADDomer VLPs to understand the drivers for their self-assembly and their thermostability. The thermal stability of ADDomer was compared to the one of a similar VLP, ChADDomer, derived from Chimpanzee Adenovirus serotype-3 (ChAdV-3); mutants L56C and S57C of ChADDomer were also designed to promote the formation of inter-penton disulphide bridges that could hinder VLP disassembly. A hybrid construct between ADDomer and ChADDomer, Chimera was also tested, along with its disulphide bridge-forming mutant S57C. We found that, of all the VLPs studied, ChADDomer S57C and L56C are the most thermally stable against aggregation and that disulphide-bridge promoting mutations also help stabilizing protein secondary structure. Conversely, ADDomer and Chimera abruptly aggregate. By structural comparison with ChADDomer, we identified the protein regions that may be responsible for this behaviour.

List of abbreviations

A – Absorbance	ESI – Electrospray ionization
A₂₆₀ – Absorbance at 260 nm	ESI-MS – Electrospray Ionization mass spectrometry
A₂₈₀ – Absorbance at 280nm	F_{hkl} – Structure factor
AFM – Atomic Force Microscopy	GFP – Green fluorescent protein
Arg – Arginine	Gln – Glutamine
Asp – Asparagine	Glu – Glutamic acid
ATP – Adenosine Triphosphate	HAdV-2 – Human Adenovirus serotype-2
AU – Arbitrary Units	HAdV-3 – Human Adenovirus serotype-3
Aβ – Amyloid β peptide	Hb – Haemoglobin
B₂₂[*] – Second virial coefficient normalized by the one of hard-spheres	HbS – Sickle Haemoglobin
B₂₂ – Second virial coefficient	HGD – Himan γ D-crystallin
B'₂₂ – Second virial coefficient for a non-ideal gas	HPLC – High performance liquid chromatography
B₂₂^{HS} – Second virial coefficient of hard spheres	I(q) – Intensity of scattered light in function of the scattering vector (SAXS)
BSA – Bovine serum albumin	I₀ – Intensity of scattered light at zero scattering angle (SAXS)
BTPI – Bovine trypsin protease inhibitor	IDP – Intrinsically disordered protein
C – Concentration	IDR – Intrinsically disordered region
C110M – Cys-110 to Met mutant	IEX – Ion exchange chromatography
C110S – Cys-110 to Ser mutant	IPTG – Isopropil- β -D-1-thiogalattopiranoside
CD – Circular dichroism	I_s(t) – Intensity of scattered light at time zero (DLS)
ChAdV-3 – Chimpanzee Adenovirus serotype-3	I_s(t+τ) – Intensity of scattered light at time zero + τ (DLS)
CNT – Classical nucleation theory	J – Nucleation rate (CNT)
CP – Capsid protein	k – partition coefficient (SEC)
Cryo-EM – Cryoelectron microscopy	k₀ – Pre-exponential factor (CNT)
CSA – Chicken serum albumin	k_B – Boltzmann constant
Cys – Cysteine	k_D – Net-interaction parameter (DLS)
d – Interplanar distance	l – Pathlength
D₀ – Ideal diffusion coefficient	L56C – Cys-56 to Leu mutation
D_c – Collective diffusion coefficient	LB – Luria-Bertani broth
DF – Dilution Factor	Leu – Leucine
DLS – Dynamic light scattering	LLPS – Liquid-liquid phase separation
DLVO – Derjaguin–Landau–Verwey–Overbeek	Lys – Lysine
dn/dc – Refractive index increment	mAb – Monoclonal antibody
DNA – Deoxyribonucleic acid	Met – Methionine
DNase I – Deoxyribonuclease I	MPD – 2-Methyl-2,4-pentanediol
dNTP – Deoxynucleoside triphosphate	M_w – Molecular weight
dsDNA – Double stranded DNA	n₀ – Solvent refractive index
DTT – Dithiothreitol	N_A – Avogadro’s number
E. coli – <i>Escherichia Coli</i>	NE-SW – Northeast–Southwest
EDTA – Ethylenediaminetetraacetic acid	
ELCS – Extended law of corresponding states	
EM – Electron microscopy	

NMR – Nuclear magnetic resonance
NW-SE – Northwest–Southeast
OD – Optical density
 $P(r)$ – Pair distribution function
P23T – Pro-23 to Thr mutant
P23TR36S – Pro-23 to Thr and Arg-36 to Ser mutant
P23TR58H – Pro-23 to Thr and Arg-58 to His mutant
P23V – Pro-23 to Val mutant
P23VC110M – Pro-23 to Val and Cys-110 to Met mutant
P23VR58H – Pro-23 to Val and Arg-58 to His mutant
Pb – Penton base protein
PCR – Polymerase chain reaction
PDB – Protein data bank
PDI – Polydispersity index
Phe – Phenylalanine
pI – Isoelectric point
Pro – Proline
 q – Scattering vector
Q – Quadrupole (ESI-MS)
R – Ideal gas constant
R14S – Arg-14 to Ser mutant
R36S – Arg-36 to Ser mutant
R58H – Arg-58 to His mutant
RBC – Red blood cell
 r_c – Critical radius (CNT)
 R_g – Radius of gyration
RGD – Arg-Gly-Asp motif
 R_H – Hydrodynamic radius
RhoGDI – Rho GDP-dissociation inhibitor
RMS – Root mean square
RMSD – Root-mean-square deviation
RNA – Ribonucleic acid
S57C – Ser-57 to Cys mutant
SARS-CoV-2 – Severe acute respiratory syndrome coronavirus-2
SASA – Solvent accessible surface area
SAXS – Small-angle X-scattering
SB – Salt bridge
SCA – Sick cell anemia
SCD – Sick cell disease
SDS – Sodium dodecyl sulfate
SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC – Size exclusion chromatography
SE-HPLC – Size exclusion high performance liquid chromatography
SELS – Ser-Glu-Leu-Ser motif
SEM – Scanning electron microscopy
SER – Surface entropy reduction
Ser – Serine
SLS – Static light scattering
SOC – Super optimal broth with catabolite repression
 T – Temperature
 T_0 – Reference Temperature
 T_{AGG} – Onset temperature of aggregation
 T_c – Critical temperature
TEMED – Tetramethylethylenediamine
ThT – Thioflavin-T
TOF – Time of flight
Tris – Tris(hydroxymethyl)aminomethane
Trp – Tryptophan
 T_{SID} – Onset temperature of scattering intensity drop
 $T_{unfolding}$ – Onset temperature of unfolding
Tyr – Tyrosine
UV – Ultraviolet
UV/Vis – Ultraviolet/Visible
v/v – Volume per volume
 V_0 – Interstitial volume (SEC)
Val – Valine
 V_E – Electrostatic interaction potential
 V_{HS} – Hard sphere excluded volume interaction potential
 V_i – Intra-particle volume
VL – Variable loop
VLP – Virus-like particle
 V_M – Matthew’s coefficient
 V_R – Retention volume
 V_{vdw} – Van der Waals interaction potential
w/v – weight per volume
W42E – Trp-24 to Glu mutant
WAXS – Wide-angle X-ray scattering
WT – Wild-type
 Z – Charge of the particle (DLVO)
 β – Supersaturation (CNT)
 γ – Surface tension (CNT)
 Γ – Decay rate (DLS)
 ΔG – Free energy of Gibbs difference
 ΔG_{crit} – Nucleation free energy barrier
 ΔH – Enthalpy difference

ΔS – Entropy difference
 $\Delta S_{protein}^0$ – Protein entropic contribution to crystallization entropy
 $\Delta S_{solvent}^0$ – Solvent entropic contribution to crystallization entropy
 θ – Ellipticity (CD), scattering angle (DLS) or half of the scattering angle (SAXS, WAXS)
 λ – wavelength
 μ_C – Chemical potential of the crystal phase
 μ_L – Chemical potential of the liquid phase
 v – Protein specific volume
 ζ – Correlation length
 Π – Osmotic pressure
 $\rho(xyz)$ – Electron density distribution
 σ – Hard sphere radius
 φ_{hkl} – Structure factor phase
 Ω – Volume of a molecule
 $g(\tau)$ – Correlation function (DLS)
 ε – Extinction coefficient
 η – Viscosity
 ϕ – Volume fraction
 ϕ_C – Critical volume fraction

Table of Contents

Chapter 1: Introduction	5
1.1 Protein self-assembly	6
1.1.1 Protein structure	7
1.1.2 Protein intermolecular self-assembly	8
1.2 Anisotropic self-assembly	15
1.2.1 Self-assembly of isotropic colloidal particles	16
1.2.2 Phase diagram of isotropic colloidal particles.....	19
1.2.3 Intermolecular interactions in protein solutions	22
1.2.4 Protein phase diagrams.....	24
1.2.5 Anisotropic interactions and protein phase diagrams.....	26
1.3 Anisotropic interactions and protein crystallization	28
1.3.1 Protein crystallization.....	28
1.3.2 Thermodynamics of protein crystallization.....	29
1.3.3 Kinetics of protein crystallization	30
1.3.4 Anisotropic interactions to improve protein crystallization	34
1.4 Self-assembly of virus capsid and anisotropic interactions	34
Chapter 2: Materials and methods	36
2.1 Preparation of buffers and reagents	37
2.1.1 Buffers.....	37
2.1.2 Buffer and reagents for <i>E. coli</i> cultures.....	37
2.1.3 Buffers and reagents for polyacrylamide gel electrophoresis (SDS-PAGE).....	39
2.2 Recombinant protein production.....	40
2.2.1 Plasmid mutation with site-directed mutagenesis	40
2.2.2 Mutated plasmid transformation and amplification	41
2.2.3 Mutated plasmid extraction, purification and sequencing	41
2.2.4 Purified plasmid transformation in <i>E. coli</i> competent cells	42

2.2.5	Recombinant protein expression in bacterial cultures.....	43
2.2.6	Recombinant protein extraction from bacterial cells.....	43
2.3	Techniques for protein purification and purity assessment.....	44
2.3.1	Size exclusion chromatography (SEC).....	44
2.3.2	Ion exchange chromatography (IEX).....	45
2.3.3	High performance liquid chromatography	47
2.3.4	Polyacrylamide gel electrophoresis (SDS-PAGE).....	50
2.3.5	Electrospray ionization mass spectrometry (ESI-MS)	52
2.4	Spectroscopy and scattering techniques for protein characterization	53
2.4.1	UV-Vis spectroscopy	53
2.4.2	Fluorescence emission spectroscopy.....	55
2.4.3	Circular dichroism spectroscopy.....	57
2.4.4	Dynamic Light Scattering (DLS)	58
2.4.5	X-ray scattering.....	60
2.5	Microscopy	62
2.5.1	Cross-polarised light microscopy.....	62
2.5.2	Sample preparation and analysis	63
2.6	Protein crystallization and solubility measurements.....	63
2.6.1	Protein crystals/assemblies growth	63
2.6.2	Protein solubility curve measurements.....	64
2.7	Protein crystallography	64
2.7.1	Theory	64
2.7.2	X-ray diffraction data collection and analysis.....	65
Chapter 3: Surface exposed free cysteine suppresses crystallization of human γD-crystallin		
.....		66
3.1	Background.....	67
3.2	Aim of the study.....	68
3.3	Results.....	68
3.3.1	Production and characterization of Human γ D-crystallin	68
3.3.2	Production and characterization of the mutant C110M.....	70

3.3.3	Production and characterization of the mutant C110S.....	72
3.3.4	Structural characterization of HGD, C110S and C110M.....	74
3.3.5	Protein-protein interactions in the liquid phase of HGD, C110M and C110S..	76
3.3.6	Crystallization of HGD, C110M and C110S.....	77
3.3.7	Solubility curve of C110M crystals and thermodynamic analysis of crystallization.....	78
3.3.8	C110M crystal structure solution from single-crystal X-ray diffraction data ...	80
3.3.9	Crystal contact analysis of C110M structure.....	81
3.4	Discussion.....	84
3.5	Conclusions.....	86
Chapter 4: Polymorphic self-assembly of human γD-crystallin driven by protein surface anisotropy.....		87
4.1	Background.....	88
4.2	Aim of the study.....	89
4.3	Results and discussion.....	90
4.3.1	Production and characterisation of the double mutant P23VC110M.....	90
4.3.2	Structural characterisation of P23VC110M.....	92
4.3.3	Protein-protein interactions in the liquid phase of P23VC110M.....	93
4.3.4	Solubility curves of P23VC110M polymorphs and thermodynamic analysis of crystallization.....	95
4.3.5	Characterisation of P23VC110M spherical assemblies by Light Microscopy..	99
4.3.6	Characterisation of P23VC110M spherical assemblies by ThT binding assay	101
4.3.7	Characterisation of P23VC110M spherical assemblies by SAXS and WAXS	102
4.3.8	Comparison of P23VC110M spherical assemblies with other protein spherical superstructures.....	105
4.4	Conclusions.....	107
Chapter 5: Stability of ADDomer virus-like particles.....		109
5.1	Background.....	110
5.2	Aim of the study.....	111
5.3	Results and discussion.....	112

5.3.1	Comparison of the structural features of ADDomer, ChADDomer and Chimera penton base proteins	112
5.3.2	Preliminary characterisation of VLPs by DLS	114
5.3.3	Temperature stability of ChADDomer VLPs probed by DLS	115
5.3.4	Temperature stability of ADDomer and Chimera VLPs probed by DLS	117
5.3.5	Comparison of VLPs thermal stabilities	119
5.3.6	Thermal stability of VLPs – structural aspects	122
5.3.7	Comparison of the primary structure of ChADDomer, ADDomer and Chimera penton base proteins	124
5.4	Conclusions.....	126
Summary and final conclusions		128
References		131

Chapter 1:

Introduction

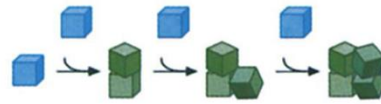
1.1 Protein self-assembly

Proteins are one of the major components of living systems. They are biological macromolecules composed of amino acids chains and their synthesis in both eukaryotic and prokaryotic cells is carried out by ribosomes. In this process, the amino acid chain is built up translating the genetic information contained in a sequence of the organism's genome¹. With 20 different naturally occurring amino acids, proteins are vastly different from each other and can then perform a wide range of very specific tasks. They are involved in DNA replication, metabolism, cell-signalling, immune-response, small molecule transport, energy production and contribute to the survival and life of any biological system¹.

Proteins can spontaneously form larger structures by self-assembly. In general terms, intermolecular self-assembly is the process through which building-block molecules assemble into higher hierarchical architectures due only to their mutual interactions. This mechanism is widespread in nature and technology and involves objects at all lengthscales² including colloidal particles³, lipid molecules⁴, DNA⁵ and proteins⁶. Self-assembly processes can be classified as static, if the final states are at equilibrium, or dynamic, if they involve energy dissipation (out-of-equilibrium)⁷. Examples of static and dynamic self-assembly processes are protein crystallization² and microtubules formation in cells driven by ATP consumption⁸, respectively.

Protein self-assembly is dictated by weak non-covalent interactions such as hydrogen bonds, van der Waals and hydrophobic interactions; however, the process is also strongly influenced by the environment, e.g. it can occur in the bulk solution or on a surface⁷. Protein self-assembly results in the formation of different types of phases and states such as gels, fibers, amorphous aggregates, liquid droplets, crystals or viral capsids. Ordered solid structures (e.g. crystals) form if the self-assembly process is either reversible or allows protein molecules to rearrange within the initial nuclei (Figure 1.1A); this occurs when the interactions that hold the structure together are of similar strength to the interactions leading to its disruption. In contrast, disordered structures such as glass-like amorphous aggregates arise from irreversible association (Figure 1.1B)⁷. Therefore, the outcomes of self-assembly are also strongly dependent on the kinetics of the process itself⁹.

A Irreversibility gives glasses.



B Reversibility gives crystals

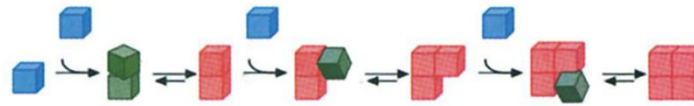


Figure 1.1: Representation of self-assembly processes where A) components interact irreversibly (green) leading to disordered glass-like materials and B) occurring when the association is reversible or components can rearrange leading to ordered structures. The image was readapted from ref. 7.

1.1.1 Protein structure

Proteins are biopolymers consisting of amino acids as building blocks that are arranged to form a linear chain (polypeptide chain). The sequence of amino acids that composes a protein is called primary structure. Amino acids are composed of a carboxylic group (-COOH), a unique side chain, and an amino group (-NH₂). In protein synthesis the -CO group of each amino acid is linked to the -NH group of the next one through a condensation reaction with the elimination of a water molecule (Figure 1.2 A). This reaction results in the formation of a peptide bond that has a planar conformation due to its two resonance structures. The flexibility of the overall polypeptide chain is instead granted by the two torsional angles ϕ , around the bond N-C α , and Ψ , around the bond C α -C(O) (Figure 1.2 B).

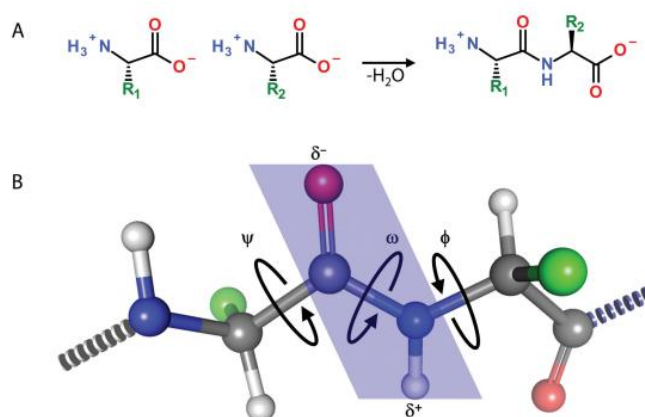


Figure 1.2: A) peptide bond formation by condensation of two amino acids and B) structure of the polypeptide chain. The planar geometry of the peptide group is highlighted in blue. The torsional angles ϕ and Ψ can rotate from 0 to 180° granting flexibility to the polypeptide chain. Image taken from ref. 10.

The backbone structure can fold on itself forming α -helix and β -sheet structures. The α -helix is stabilized by hydrogen bonds formed between the $-\text{CO}$ and $-\text{NH}$ of two peptide bonds placed four residues apart from each other (Figure 1.3 A), while the β -sheets are pleated structures formed by different stretches of the polypeptide chain held together by hydrogen bonds between the polypeptide layers (Figure 1.3 B).

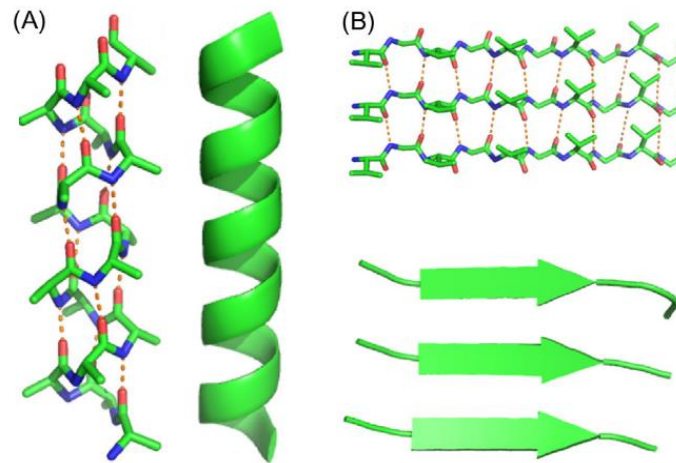


Figure 1.3: Hydrogen bond network and cartoon structure of A) α -helix and B) β -sheet structures. Image taken from ref. 11.

The secondary structures fold further through the interactions between amino acidic side chains to give the protein its global shape (tertiary structure). Finally, if the protein is composed of different subunits (different polypeptide chains), their specific assembly is called quaternary structure.

The polypeptide chain of a protein generally contains from 50 to 1000 amino acidic residues¹² and its folding can generate proteins of several nanometers in radius¹³. Depending on the aforementioned structural properties, proteins can adopt several different shapes. Two common classes of proteins are globular proteins, approximately spherical, and fibrous proteins, with an elongated shape¹

1.1.2 Protein intermolecular self-assembly

To understand the importance of intermolecular protein self-assembly this section will briefly provide common examples related to biological processes, the pathogenesis of some diseases, and how it is involved in the fields of protein crystallization and biotherapeutics development.

1.1.2.1 Functional protein self-assembly

Protein compartmentalization in cells

Cells are basic structures of living systems able to accomplish essential tasks as protein synthesis, DNA transcription and replication and energy production. Each of these processes takes place in different cell subunits, or organelles, bounded by a membrane. In addition, cells also have membraneless organelles¹⁴. These structures are liquid, spherical condensates of proteins and RNA. A large number of these condensates have been studied over the years and this includes P-granules, Nucleoli, Cajan bodies and stress granules¹⁵.

Only recently, it has been shown that biological condensates are formed by a self-assembly process well known in biological physics; liquid-liquid phase separation (LLPS)¹⁶. This phenomenon consists of a spontaneous demixing of the homogeneous liquid into two liquid phases, one with a higher and one with a lower protein concentration than the initial phase¹⁶. The tendency to undergo LLPS *in vivo* has been associated with the presence of intrinsically-disordered regions (IDRs), i.e. domains with high flexibility provided by the low complexity of their primary structure (repetition of few types of amino acids)¹⁴. LLPS enables the organization of biomolecules into compartments while maintaining their ability to diffuse in and out the boundary¹⁴: this mechanism is exploited, for instance, in neurons to control protein concentration without regulating protein synthesis¹⁵, to accelerate biochemical reactions by locally concentrating the reactants¹⁷ or to temporarily store proteins to protect them from stress damage¹⁵.

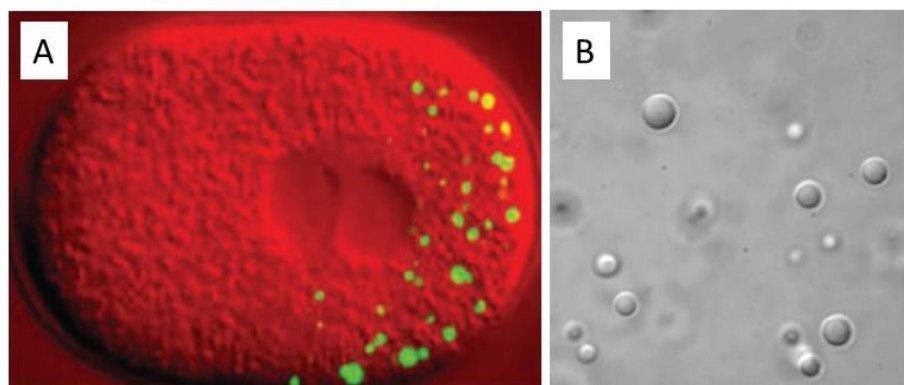


Figure 1.4: A) *In vivo* LLPS: GFP-tagged P-granules (green) in the cytoplasm of a *Caenorhabditis elegans* embryo¹⁸. B) FUS protein undergoing LLPS *in vitro*¹⁵.

Viral capsid

Viruses are infectious agents that exploit living forms to reproduce; since they are less complex than cells, the debate on whether they are living systems is still ongoing¹⁹. Therefore, viruses can be defined as simple nanoscopic entities composed only of a genome (RNA or DNA) enclosed in

a protective protein shell, known as virus capsid. Since they do not possess the complex metabolic machinery belonging to living systems, viruses attack and hijack host cells to reproduce their genome and hence replicate themselves²⁰. Viruses can adopt different morphologies, the most common are rod-like, when their capsid has a helical symmetry, or spherical, when their capsid has an icosahedral symmetry (Figure 1.5C)^{21,22}. For instance, Human Adenoviruses are icosahedral particles with a radius of 65-80nm and their capsid is formed by three components, i.e. hexon, penton base and fiber proteins²³. Viruses can also be classified in enveloped or non-enveloped, depending on whether the capsid is surrounded by one or more lipid bilayers²⁴. Viral capsids form by protein self-assembly both *in vivo* and *in vitro*, either in the presence or absence of genome in their interior (Section 1.4); the mechanism of viral self-assembly is studied for its obvious implications in viral infections but also since penton base proteins can self-assemble alone forming smaller dodecahedral particles, i.e. virus-like particles (VLPs). VLPs can be exploited as vaccine candidates or drug delivery systems (Section 1.1.2.4)²¹.

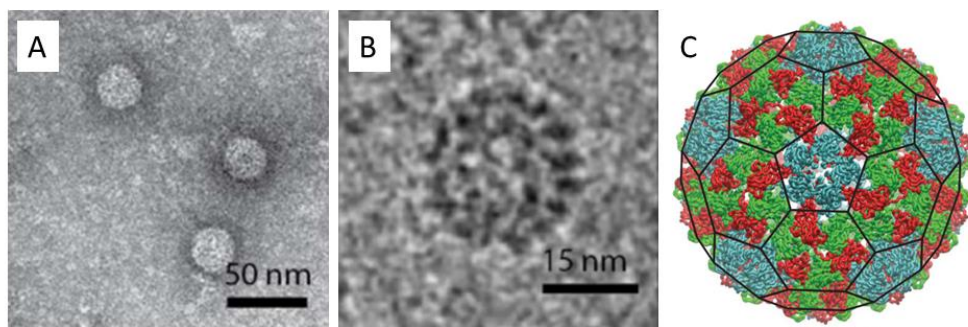


Figure 1.5: A) and B) Cryo-TEM micrographs showing co-assemblies of Cowpea Chlorotic Mottle Virus protein and brome mosaic virus RNA1²⁵. C) illustration of the icosahedral capsid of the same virus²¹.

Protein fibres as scaffolds for cells and tissues

Insoluble protein fibres are natural structural materials that grant motility, flexibility, elasticity and scaffolding properties to cells and tissues²⁶. Their formation occurs *in vivo* by self-assembly of protein monomers with a mechanism dependent on the specific protein involved²⁷. An example is provided by collagens: this is a family of extracellular protein structures with a fibrillar shape that serve as scaffold for connective tissue, skin and bones²⁷. Collagen is formed *in vivo* by the self-assembly of three polypeptide chains of procollagen, its soluble monomeric precursor, folding to form a triple helix (Figure 1.6)²⁷. The association of the three monomeric chains starts at their C-terminus²⁸; when the triple helix is folded, the N- and C-termini form globular domains that are cleaved by the respective metalloproteinases, and collagen molecules are then able to bundle forming fibers²⁹. Other examples of protein that naturally form these types of structures are elastin, keratin, actin, and silk²⁷. Protein fibers or filaments can also be prepared *in vitro* and used for biomedical applications such as tissue repair, wound healing and controllable drug

delivery. They have the advantage, over synthetic polymers that they are non-toxic, biodegradable and biocompatible³⁰. Furthermore, these protein architectures inspire the *de novo* design of synthetic molecules with similar tuneable self-assembly properties like alkyl peptide amphiphiles^{31,32}.

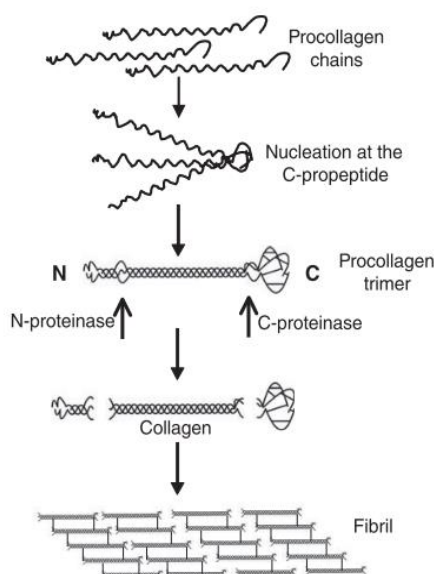


Figure 1.6: Self-assembly mechanism of procollagen monomers into collagen (ref. 29).

1.1.2.2 Aberrant protein self-assembly: condensation diseases

Despite being functional in almost every biological process, protein self-assembly can also result in aberrant phenomena. The expression ‘Protein condensation diseases’ was firstly introduced to relate a series of apparently different conditions to a common source: Alzheimer, Parkinson, Sickle cell anemia and Cataract are indeed all conditions related to the formation of insoluble phases due to net-attractive protein-protein interactions in their respective biological environment³³.

Alzheimer disease

Alzheimer Disease is a neurodegenerative condition recognized as the main cause of dementia in adult patients³⁴. It was firstly observed by Alzheimer in 1906 and it was associated with the presence of insoluble deposits in the brain tissue of patients³⁵. These ‘plaques’ are composed of bundles of fibrils (Figure 1.7) formed by peptide association³⁶. In humans, the peptide responsible for this phenomenon is the Amyloid β peptide ($A\beta$), existing in two variants, $A\beta_{40}$ and $A\beta_{42}$ (40 and 42 amino acid-long)³⁷. It self-associates forming β -sheet structures orthogonal to the elongation axis of the fiber (cross- β structure) which are recognizable from distinct X-Ray diffraction patterns³⁸. The mechanism of fibril formation goes through a pre-fibrillar stage, where

only low molecular weight oligomeric species with an annular shape are formed. The concentration of oligomers then gradually decreases giving rise, in an elongation step, to the fibrils. This usually occurs only after a lag phase, that suggests a nucleation and growth mechanism³⁷. Initially the amyloid plaques were considered to be the cause of the disease, but in the last decades it has been widely demonstrated that only the intermediate oligomeric species are neurotoxic³⁹. This has stimulated the development of several types of therapeutic strategies aimed to inhibit oligomerization and fibrillation⁴⁰. While the majority of early studies focused on the self-assembly of A β , recent studies have also highlighted the correlation between insoluble brain deposits of a second protein – tau – and symptoms of dementia; the roles of these two proteins are still under debate, however they are most likely both involved in Alzheimer's disease pathogenesis^{41,42}.

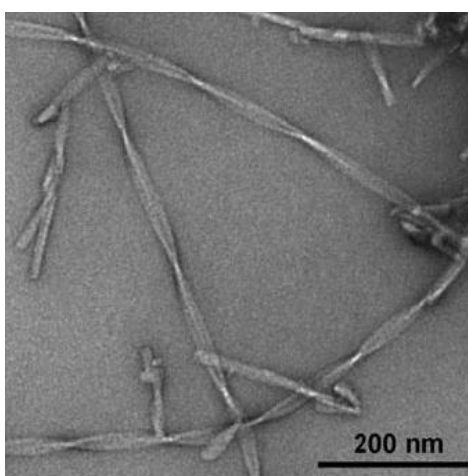


Figure 1.7: Cryo-EM micrograph of fibers formed by A β 40 peptide (ref. 43).

Sickle Cell Anemia

Sickle Cell Diseases (SCD) are conditions affecting erythrocytes, or Red Blood Cells (RBC) that are part of the bloodstream. Haemoglobin (Hb) is a tetrameric protein expressed in RBC that binds oxygen delivering it from the lungs to any tissue in the entire body. SCD are caused by single amino acid mutations in the haemoglobin subunit β ⁴⁴. In particular, Sickle Cell Anemia (SCA) is caused by the replacement of the hydrophilic, negatively-charged glutamate (Glu) on the 6th position with valine (Val), a hydrophobic residue (Sickle Haemoglobin, HbS). This mutation has little effect when the protein is in its oxygenated state. However, when the protein has released the oxygen molecule (deoxygenated state), Val6 interacts strongly with a phenylalanine (Phe85) and a leucine (Leu88) of a neighbour haemoglobin molecule forming strong hydrophobic contacts⁴⁵. This enhances protein association inside the RBC leading to the formation of branched fibrils through a double-nucleation mechanism⁴⁶. The new condensed protein phase dramatically alters RBC giving them the well-known crescent-moon shape (Figure 1.8) that modifies their

mechanical and rheological properties in the bloodstream⁴⁴. This results in blood vessels occlusions due to accumulation of the sickle cells that eventually lead to the onset of the disease⁴⁷. It is also significant to note that the very first therapy to fight sickle cell disease – the use of hydroxyurea to inhibit the protein assembly process – was developed following biophysical studies examining the kinetics of aggregation of the deoxygenated form of HbS^{48,49}.

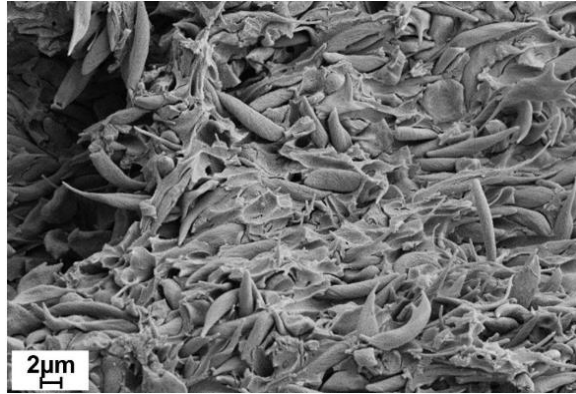


Figure 1.8: SEM micrograph of a blood clot due to sickle cells (ref. 50).

Eye lens and cataract

The eye lens is formed by aligned transparent fiber cells which contain a tightly packed mixture of proteins. In vertebrates, over 90% of the proteins contained in these cells are α , β or γ -crystallins⁵¹. The three classes of crystallins are unevenly distributed in the lens, with α more abundant in the cortex and γ more abundant in the nucleus⁵². In physiological conditions, the interactions between α -crystallins and between β -crystallins are repulsive while they are attractive between γ -crystallins⁵³. This balance maintains the concentration gradient in the eye lens, that ranges from 250 in the cortex to 400 mg/mL in the nucleus, providing the refractive index gradient required to focus the incident light to the retina⁵⁴. Cataract is caused by the presence of protein condensed phases in the eye lens that anomalously scatter light³³. There are two types of cataract depending on the source of protein association. Genetic cataract (also referred to as juvenile, hereditary or congenital cataract) is caused by amino acid mutations of crystallin proteins⁵⁵. In humans, several mutants of γ D-crystallin cause the protein to aggregate/crystallize (Figure 1.9) resulting in cataract onset during childhood^{56,57}. Age-related cataract is instead due to post-translational modifications of eye lens proteins that build up during aging. Indeed, crystallins are subjected to no or very little turnover during an organism's life. Hence, chemical modifications such as intermolecular disulphide bridge formation, carbamoylation, glycosylation, racemization irreversibly increase the propensity of the modified proteins to aggregate in the eye leading to the onset of the disease in late age³³.

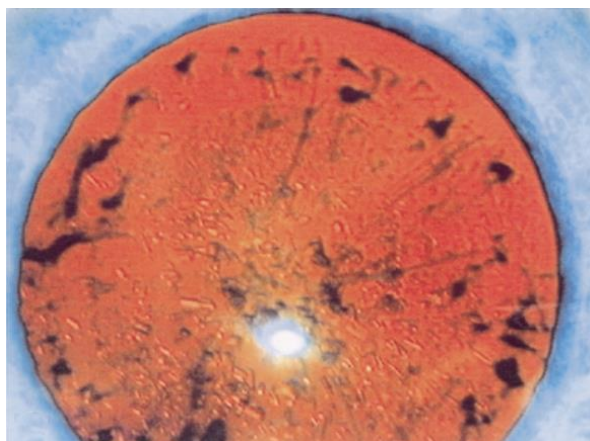


Figure 1.9: Young patient's eye lens before surgery showing numerous protein crystals (ref. 58).

1.1.2.3 Relevance of protein crystallization

The crystal phase is one of the many different solid forms that results from protein self-assembly. Proteins are notoriously recalcitrant to crystallization, and in most cases, crystals are only associated with disease onset (Section 1.1.2.2). It has been proposed that this is the result of an evolutionary negative design that maintains proteins soluble *in vivo* to preserve protein function and cell viability⁵⁹. However, there are also cases in which naturally occurring protein crystals may be biologically useful for protein storage, virion protection, inhibition of proteolysis or solid-state catalysis.^{60,61,62} In addition, protein crystal structure determination by X-Ray crystallography has extensively been used in the last century as the primary tool to investigate protein function and activity and, consequently, to develop new drugs^{63,64}. Despite the advances in recent years of other techniques that allow protein structure determination without the requirement of protein crystals (Cryogenic Electron Microscopy [Cryo-EM] and Nuclear Magnetic Resonance [NMR]), X-ray diffraction produces structures with higher resolutions and will therefore remain the leading technique to obtain protein structure with atomic level details for the foreseeable future⁶³. However, the phenomenon of protein crystallization is still poorly understood due to the heterogeneity of the protein surface, structure and shape. For all of these reasons, better knowledge of protein crystallization, which at present is still mostly achieved by trial and error⁶³, is required.

1.1.2.4 Biotherapeutics development

Proteins are widely employed as therapeutic agents against several conditions including autoimmunity, autoinflammation, cancer, infections, genetic disorders and enzyme deficiency. Examples of important classes of proteins are monoclonal antibodies, plasma proteins, hormones, fusion proteins and coagulation factors^{65,66}. Protein self-assembly can lead to the formation of reversible and irreversible aggregation that can have undesirable effects not only on the efficiency

of the treatment itself, but also on manufacturing, storage and administration of the product⁶⁷. Irreversible protein aggregation can occur following degrading chemical reactions, such as oxidation, disulphide crosslinking, deamidation, peptide bond hydrolysis or thermal and mechanical stress. This can alter the native conformational state of the protein exposing hydrophobic regions that initiate the non-covalent association⁶⁸. In addition, reversible protein self-assembly affects solution viscosity, complicating the delivery by injection⁶⁸. These phenomena limit the number of available bioformulations and the maximum usable protein concentration. Most of the research in this area is focused on understanding the mechanisms of aggregation and proposing new ways to overcome it. The main methods used to inhibit protein aggregation are modifications of either the protein (chemical conjugation or mutagenesis) or solution conditions (addition of excipients such as sugars, cyclodextrin, L-arginine, glycerol, etc.)⁶⁹. The final aim is to increase either the conformational and/or colloidal stability of the protein reducing its aggregation propensity.

A special class of biotherapeutics are Virus-like particles (VLPs). VLPs are virus-derived protein assemblies lacking viral genome and have been studied due to their potential application as drug delivery systems and vaccine candidates²². The most common morphologies of VLPs are rod-shaped, or dodecahedral. VLPs can be engineered to display a specific epitope on their surface resembling the native virus of interest; several copies of the epitope exposed on VLPs surface trigger the immune response allowing the organism to develop immunity to the infectious disease. VLPs are considered safer alternatives to traditional vaccines as they lack of any genetic material⁷⁰. Bioformulations based on VLPs have been proposed as vaccine candidates against several infectious agents such as hepatitis B virus, human papillomavirus and, more recently, SARS-CoV-2^{24,70}. Virus capsid or VLPs stability and immunogenicity can be heavily compromised by external conditions such as temperature or shear stress²⁴; in some cases research is directed to improve their temperature resistance by stabilizing mutations⁷¹ or addition of excipients like sugars⁷² to improve ease of distribution and storing.

1.2 Anisotropic self-assembly

For a time, colloidal models for protein assembly simply considered these biomolecules as spherical objects with a homogeneous surface (isotropic particles), and hence isotropic models derived for colloids were used to understand protein self-assembly; however, protein interactions and the consequent assemblies they form cannot be fully captured by these models^{73,74}. In reality, the chemical heterogeneity of protein surface makes them highly anisotropic particles, which in turn makes protein-protein interactions orientation-dependent⁷⁵. There are in fact several examples where protein self-assembly is dramatically influenced by small variation on protein surface: cataract or sickle cell anemia onset (Section 1.1.2.2), enhanced nucleation of protein

crystals due to mutagenesis^{76,77}, viral capsid assembly^{78,79} among others. Hence, for a complete understanding of protein self-assembly phenomena, interaction models should also include terms to account for the anisotropy of the protein surface. In this section, both simple theories for isotropic colloidal particles and their expansion to include anisotropic interaction will be reviewed.

1.2.1 Self-assembly of isotropic colloidal particles

1.2.1.1 Van der Waals and electrostatic interactions: DLVO theory

Historically, the most widely used theory to describe colloidal solutions behaviour is the DeJaguin – Landau – Verwey – Overbeek (DLVO) theory. Two interacting particles are assumed to be hard spheres with a charge uniformly distributed on their surface and immersed in a medium with dielectric constant ϵ containing ions as point-charges. The centres of mass of the two particles are at distance r from each other. Hence, the pair potential between the two particles, $V(r)$, is expressed as:

$$V(r) = V_{HS}(r) + V_{vdW}(r) + V_E(r) \quad 1.1$$

Where V_{HS} is the contribution due to the hard sphere excluded volume, V_{vdW} is the van der Waals interaction and V_E is due to electrostatic interactions. The excluded volume of a molecule is the volume that is inaccessible to other molecules due to the presence of the first one⁸⁰. As two molecules cannot interpenetrate each other, this gives rise to a repulsive, short-range interaction. The van der Waals force includes three types of interactions depending on the polarity and polarizability of the two molecules: i) two polar molecules (permanent dipoles) interacting through their opposite-charged domains (dipole-dipole interaction), ii) a polar molecule inducing a dipole moment in a neutral molecule (dipole – induced dipole interaction) and iii) two neutral molecules interacting through the formation of instantaneous dipoles from their electron density fluctuations (London dispersion force). This last contribution is the most significant as it is present in all scenarios. The van der Waals force for two small molecules is given by:

$$V_{vdW}(r) = -\frac{C}{r^6} \quad 1.2$$

Where C is the sum of three attractive terms related to the dipole interactions and the r^{-6} dependence indicating it is short-range. For bigger molecules, such proteins, the expression of van der Waals becomes more complex and depends on the dielectric constant of the molecule and the medium and also on the molecule geometry. For two identical “large” particles, van der Waals interactions are always attractive and becomes longer-ranged (r^{-2} dependence)⁸¹.

Colloidal particles immersed in an electrolyte solution can acquire an electrical charge from the presence of ionisable groups on their surface and from ion adsorption; this results in the formation of an electric double layer. The first layer of tightly bounded ions is called the Stern layer. To maintain the electroneutrality of the solution, the charged particle and its adsorbed ions are surrounded by a second diffuse layer of loosely attracted counter-ions subjected both to attraction to the particle and thermal motion. The external limit that divides the diffuse layer from the bulk solution is called slipping plane (Figure 1.10).

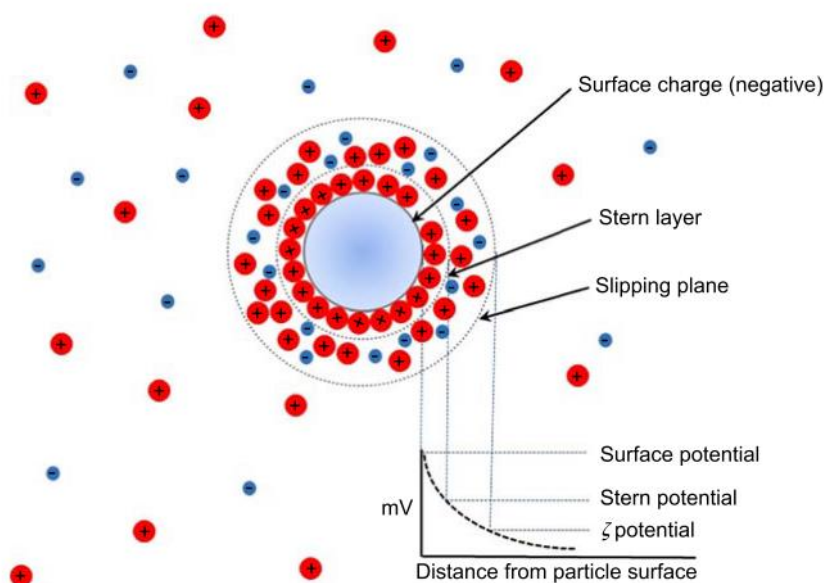


Figure 1.10: Illustration of the electric double layer formed by a negatively charged colloidal particle. The slipping plane is the outer limit that includes the solvent and ions that diffuse with the particle. Image taken from ref. 82.

The electrical potential of the solution exponentially decays with the distance from the particle surface as expressed by the equation⁸¹:

$$V_E(r) = \frac{Z^2 e^2}{\epsilon(1 + \kappa a)} \exp \left[-\frac{\kappa(r - 2a)}{r} \right] \quad 1.3$$

Where Z and a are respectively the charge and radius of the particle, e is the elementary charge and κ^{-1} is the Debye screening length. This is the distance at which the slipping plane occurs and it is inversely proportional to the ion concentration. The potential at the Debye screening length is known as ζ -potential (zeta-potential). In a two-particle system of the same charge, the electric double layer provides a repulsive force between the two particles. At low ionic strength regimes, the thickness of the double layer (κ^{-1}) and ζ -potential are large, giving rise to long-range strongly repulsive electrostatic interactions. Increasing the ionic strength decreases both of these parameters making the electrostatic repulsion weaker and shorter-range (screening effect).

The net interaction potential resulting from the balance of these three terms determines the stability (homogeneous solution) or instability (flocculation) of the colloidal mixture. In Figure 1.11, the net interaction potential is shown along with the contribution of the attractive van der Waals and repulsive electric double-layer potentials (dashed lines). The inset shows how the shape of the potential changes from (a) a stable colloidal solution dominated by long-range repulsive electrostatic interaction to (e) an unstable colloidal solution dominated by short-range attractive van der Waals interactions⁸³.

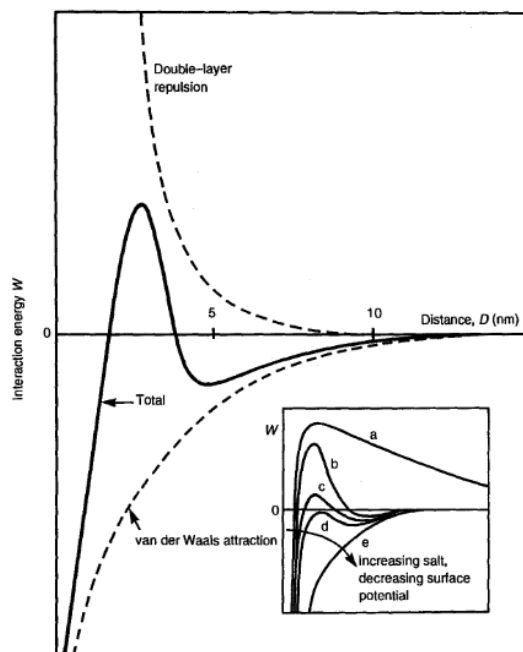


Figure 1.11: DLVO net interaction potential shown with its two contributions: electric-double layer repulsion and van der Waals attraction. The inset shows how increasing salt concentration (from a to e) decreases the repulsive component and increases the attractive component of the potential curve. Image taken from ⁸³.

However, DLVO theory suffers from some limitations:

- It considers the interacting particles only as isotropic spheres. Strong deviations can arise for particles with different shapes, surface roughness and anisotropy such as proteins.
- The only property of the ions considered is the charge: this neglects any dispersion force between them. Ion dispersion forces are among the possible explanations of the salt-specific effects on protein-protein interactions causing salting out (precipitation) or salting in (solubilisation) of proteins as described by the Hofmeister series⁸¹.
- It does not consider hydration and hydrophobic effects.

1.2.1.2 Hydration and hydrophobic effect: non-DLVO forces

Protein molecules are subjected also to non-DLVO forces related to hydration. When a molecule is immersed in an aqueous medium its surface is covered by a solvation layer composed of water

molecules and ions. The solvation process is affected by the hydrophilicity or hydrophobicity of the surface. Two interacting hydrophilic molecules experience a repulsive force due to the positive Gibbs free energy necessary to expel the hydration molecules on their surface (hydration interaction). On the contrary, two hydrophobic molecules will be subjected to an attractive interaction since the Gibbs free energy for the release of water molecules between them is negative (hydrophobic effect)⁸¹. The latter interaction is also related to the structuring of water: when a non-polar molecule comes into contact with bulk water, the water hydrogen bond network is disrupted and immediately reformed, increasing water self-interactions. This leads to a more structured layer of water around hydrophobic molecules (hydrophobic hydration effect)⁸³. Such effects become more complicated for proteins, that have a heterogeneous surface composed of both polar and non-polar amino acid, and are not yet well understood.

Table 1.1: Summary of non-covalent interactions properties. Their strength is usually between 1-40 kJ/mol (weak interactions 1-10 kJ/mol, strong interactions 10-40 kJ/mol)¹⁰.

Interaction	Attractive (-) Repulsive (+)	Strength	Range
van der Waals	-	Weak	Short (small molecules) Longer (large molecules)
Hydration	+	Strong	Short
Hydrophobic effect	-	Strong	Long
Excluded volume	+	Strong	Short
Electrostatic	+	Weak (high ionic strength) Strong (low ionic strength)	Long (low ionic strength) Short (high ionic strength)

1.2.2 Phase diagram of isotropic colloidal particles

The interactions between particles of any type (atoms, small molecules, colloids, etc.) determine the physical state of the system. Hence, the net interaction potential gives rise to the equilibrium phase diagram of the substance that is usually expressed in respect of temperature and volume fraction ϕ (a quantity related to concentration). The interaction potential is not directly obtainable experimentally, thus phase diagrams are determined by simulations using Monte Carlo or perturbation theory starting from reasonable expressions of the potential function^{81,84}. To describe the pairwise interaction between two colloidal particles we have first to explain the simpler case of two isotropic and non-colloidal spheres (hard spheres) such as atoms or small molecules. The simplest form of the potential can be expressed by the Lennard-Jones equation:

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] \quad 1.4$$

Where ε is the depth of the potential well, r the separation between the two particles and σ is the separation between the two particles when the potential reaches 0 (that is the particle radius). The total potential is given by the sum of a long-range attractive term, proportional to $-1/r^6$, and a short-range repulsive term, proportional to $+1/r^{12}$ (Figure 1.12). The attractive part arises from van der Waals interactions and the repulsive part is due to the overlapping of electronic orbitals of the two spheres (Pauli repulsion)⁸⁵.

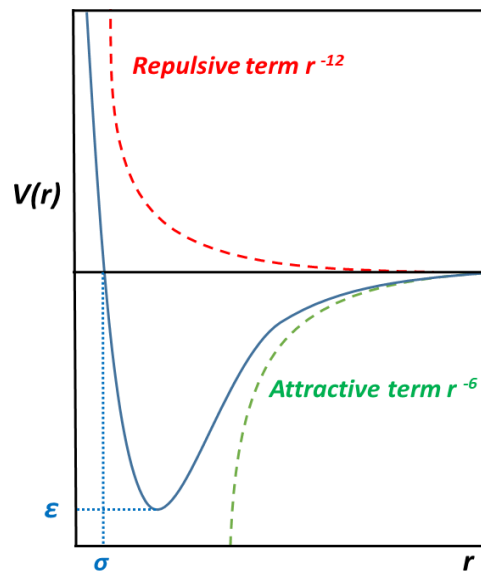


Figure 1.12: Generic illustration of the total interaction potential resulting from the two opposite contributions of a short-range repulsive and a long-range attractive terms.

The phase diagrams for simple non-colloidal substances such as Argon or generic molecular fluids, can be reproduced using a square-well or Lennard-Jones potentials that accounts for long-range interactions^{86,87}. This results into three different phases: a fluid F, that can be either dense (liquid, L) or dilute (vapour, V), and a solid S. Phase diagrams are then composed of the respective monophasic regions L, S, V and biphasic regions $V+L, F+S, V+S$ (Figure 1.13A).

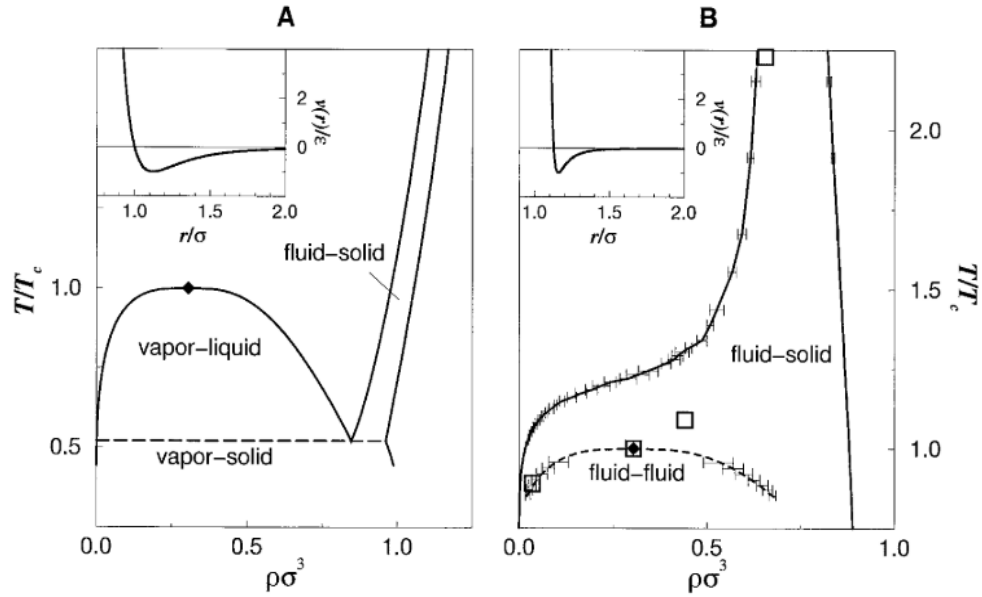


Figure 1.13: Typical phase diagrams of A) simple molecular substances with long-range attraction and B) colloids with short-range attraction. T_c is the critical temperature (maximum temperature of the fluid-fluid biphasic region) and ρ the density. The x-axis can be related to concentration or volume fraction, ϕ . The insets show the long-range and short-range potentials used to calculate the phase diagrams respectively (ref. 86).

The biphasic region $V+L$ in Figure 1.13A is stable in respect of the solidification, i.e. it lies above the phase boundary between F and $F+S$ (liquidus line). This means that a molecular fluid when cooled goes from vapour (V) to vapour-liquid ($V+L$) to vapour-solid ($V+S$). The Lennard-Jones equation can be replaced or modified to account for a shorter range of interaction between the particles (insets of Figure 1.13A and B). If the range of the interaction is reduced to approximately less than one-quarter of the particle diameter⁸⁶, the phase diagram changes drastically to the one shown in Figure 1.13B. This is the situation of colloidal systems such as proteins (here the vapour and liquid phases correspond respectively to a dilute and dense fluid phase [F] since the thermodynamic formalism is changed by the application of the McMillan-Mayer approximation, described in section 1.2.3), where the order of the phase transitions is inverted: upon cooling, colloids go from dilute fluid (F) to fluid-solid ($F+S$) and then to dense fluid-dilute fluid ($F+F$)⁸⁷. This makes the $F+F$ biphasic region metastable in respect of solidification, i.e. this region lies below the liquidus line. The phenomenon of the metastable formation of two fluid (liquid) phases, one diluted and one dense is indeed notorious in colloidal and protein science and known as liquid-liquid phase separation (LLPS).

Although this type of model can reproduce the phase diagram of simple isotropic colloids, proteins are much more complex systems due to their shape, flexibility and heterogeneous surface and hence, calculating their equilibrium phase diagrams from atomistic details is still prohibitive^{73,88}. Nevertheless, it is possible to gain at least some information about the net

interaction potential; indeed, the strength of protein-protein interactions in specific solution conditions can be quantified by an experimental parameter, the second virial coefficient⁷³.

1.2.3 Intermolecular interactions in protein solutions

Historically, protein-protein interactions in solution were measured by osmometry. In an osmometric measurement a solution containing protein and salt is placed in a compartment that is in contact with a second one containing only the salt solution. The two vessels are separated by a semi-permeable membrane that can be penetrated by salt ions but not by protein molecules. To reach equilibrium, the protein solution compartment will be subjected to an influx of the salt solution from the second compartment; the pressure to be exerted on the protein solution to stop this flow is called osmotic pressure¹⁰. For ideal solutions, i.e. solution where there are no interactions between molecules, the dependence of osmotic pressure Π , on protein mass concentration C (g/L), would be linear and expressed by Van't Hoff equation:

$$\frac{\Pi}{RT} = \frac{C}{M_W} \quad 1.5$$

where R is the ideal gas constant, T the temperature and M_W the molecular weight of the protein. In reality, positive or negative deviations from the straight line are observed (Figure 1.14)⁵² and can be described applying a virial expansion to equation 1.5:

$$\frac{\Pi}{RT} = \frac{C}{M_W} + B_{22}C^2 + \dots \quad 1.6$$

Where B_{22} is the second virial coefficient (with units $mol \cdot L/g^2$). Usually, the truncation after the second term of the equation is enough to approximate the deviation. Here, the second virial coefficient B_{22} is analogous to the one of a non-ideal van der Waals gas that accounts for the non-ideality of the system due to intermolecular interactions.

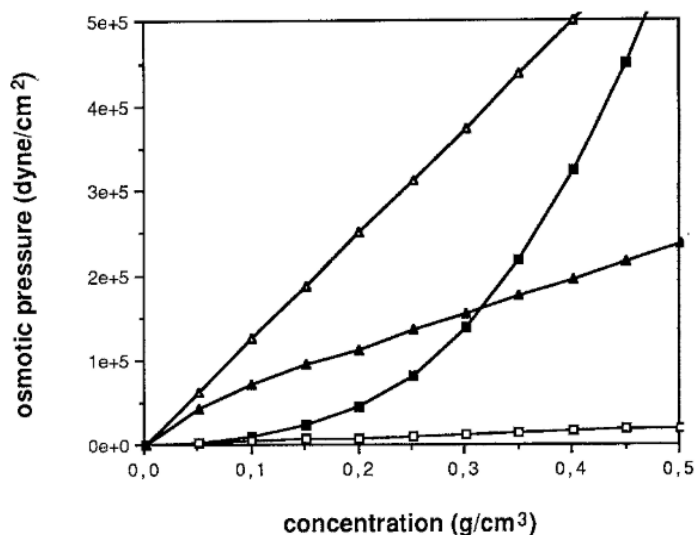


Figure 1.14: Deviations from ideality of the osmotic pressure for bovine γ -crystallins (triangles) and α -crystallins (squares): the straight lines (open symbols) represent the ideal cases calculated from Equation 1.5 while the non-linear trends (filled symbols) are obtained experimentally⁵².

Therefore, B_{22} is a measure of net protein-protein interactions in specific solution conditions; the value of B_{22} is negative when interactions are attractive and positive when they are repulsive⁵³. To date, more frequently used techniques to measure the second virial coefficient are Static Light Scattering (SLS) and Small-angle X-Ray scattering (SAXS); protein-protein interactions can be also evaluated by Dynamic Light Scattering (DLS) by measuring k_D , the net interaction parameter. This quantity is related to B_{22} but it also includes an hydrodynamic contribution⁸⁹. Numerous studies have measured B_{22} as a function of solution conditions such as pH and salt concentration demonstrating that protein-protein interactions can be qualitatively predicted in the framework of DLVO theory^{89,90,91}. However, the limitations discussed above reduce the applicability of this theory (e.g. it is not valid for high concentration of electrolytes, it is not able to distinguish for ions identity, it does not consider hydration forces).

As a measure of the net interaction between protein molecules, B_{22} is also useful for the description of protein phase diagrams. The second virial coefficient of protein solutions is analogous to the second virial coefficient developed from the statistical thermodynamic treatment of non-ideal gases in vacuum. The McMillian-Mayer framework allows the treatment of the complex protein environment, e.g. composed of water molecules (solvent), salt ions (solute 1) and protein (solute 2), as a pseudo one-component system made of the protein molecules (solute) and a continuum background of water and ions (solvent)^{92,93}. The solvent background is then accounted for in the interaction potential between the two interacting protein particles⁸¹. This converts the thermodynamic properties of a non-ideal gas in vacuum to protein solution in a

complex medium and vice versa (e.g. converts pressure to osmotic pressure). The expression of the second virial coefficient for a non-ideal gas is⁸⁰:

$$B'_{22} = -\frac{1}{2} \int_0^{\infty} [e^{-\frac{V(r)}{kT}} - 1] 4\pi r^2 dr \quad 1.7$$

where $V(r)$ is the potential of mean force (analogous to the net interaction potential in protein solutions) and k the Boltzmann constant. B'_{22} can be simply related with the experimental second virial coefficient measured in protein solutions by $B'_{22} = B_{22}M_W^2/N_A$ where N_A is the Avogadro's number.

1.2.4 Protein phase diagrams

Generally, a phase diagram is a map of the physical form (liquid, solid, vapour, etc.) that a compound can adopt by changing the conditions of the systems (temperature, pressure, concentration, etc.). As already mentioned, protein self-assembly leads to a broad range of phases and states (gels, crystals, amorphous aggregates, fibrils, etc.). Although many of them are non-equilibrium states, the knowledge of equilibrium phase diagrams of proteins is essential to fully understand self-assembly phenomena. Experimental protein phase diagrams strongly resemble the one computed for isotropic colloidal particles (Figure 1.13B). A generic sketch is shown in Figure 1.15.

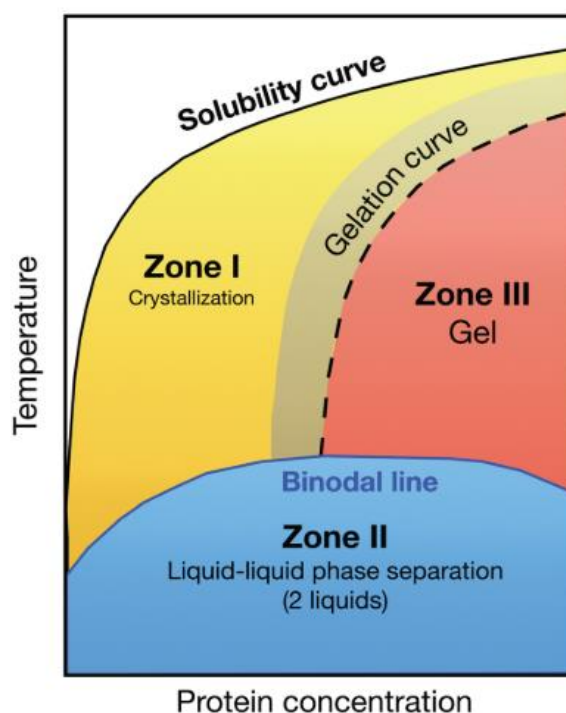


Figure 1.15: Generic phase diagram of proteins: Zone I and II are respectively the L+S and L+L (LLPS) biphasic regions. Zone III indicates the gelation region (non-equilibrium process). The figure was firstly published in ref. 94 and readapted by authors of ref. 95.

The solubility curve, or liquidus line, separates the liquid phase region (L , white zone in Figure 1.15) from the liquid and solid biphasic region ($L+S$, Zone I) where crystallization is thermodynamically favourable. The solubility curve is not a property of the protein itself, but of the equilibrium between the liquid phase and the specific solid phase it forms: if the protein crystallizes in different polymorphs, these will exhibit different solubility curves⁹⁴. In rare cases, the solubility curve is inverted (i.e. the protein equilibrium solubility decreases when increasing the temperature)^{96,97,98,99}, or independent of temperature¹⁰⁰. At low temperatures some protein solutions become cloudy forming small droplets that eventually sink creating a second liquid phase. This process is called liquid-liquid phase separation (LLPS, Zone II) and is due to a spontaneous demixing of the protein solution into coexisting concentrated and diluted phases. Many proteins undergo LLPS; some of them are Lysozyme^{94,101,102}, γ -crystallins^{54,103}, haemoglobin^{98,104}, bovine pancreatic trypsin inhibitor (BPTI)¹⁰⁵, monoclonal antibodies (mAb)^{106,107} and several intrinsically disordered proteins (IDP)^{108,109,110}. LLPS is analogous to the fluid-fluid separation for colloids (section 1.2.2) and it is metastable with respect to solidification. The maximum temperature at which LLPS occurs is known as critical temperature, T_c ; the protein volume fraction at T_c is the critical volume fraction ϕ_c . At very high concentrations, protein solutions can also form a gel-like state characterized by a disordered network (Zone III); however, this is not an equilibrium phase and it is obtained by kinetic arrest of the system dynamics for thermodynamic conditions where the crystal phase is still the most stable form⁸¹.

The position and shape of equilibrium phase boundaries (LLPS and solubility curve) depend on the strength of the attractive interactions between proteins that can be changed by varying solution conditions such as pH, salt type and salt concentration⁹²: if the attraction between proteins increases, the critical temperature of LLPS also increases and the solubility curve is shifted to lower protein volume fractions^{94,111, 112}. This is also confirmed by simulations showing that the critical volume fraction and temperature are respectively determined by the range and the magnitude of protein-protein interactions^{75,113}.

It has been observed that experimental LLPS curves of some different globular proteins in several solution conditions collapse on a universal curve when normalized by the critical parameters T_c and ϕ_c ¹¹⁴. This means that systems with different interaction potentials can be compared if such scaling is applied⁸¹ and demonstrates the universality of the generic protein phase diagram⁹⁵. This is known as extended law of corresponding states (ELCS)¹¹⁵ and in principle would allow the location of the LLPS boundary to be predicted, only knowing the reduced second virial coefficient, B_{22}^* , since this parameter is only weakly influenced by the interaction potential⁷³ (n.b. B_{22}^* is the second virial coefficient normalized by the value for hard spheres, B_{22}^{HS} , which only depends on the hard sphere diameter, σ ⁹⁵).

1.2.5 Anisotropic interactions and protein phase diagrams

The phase diagrams of globular proteins can be partially reproduced by isotropic models. However, while these models capture the critical parameters, they fail to reproduce the width of the LLPS curve and the accurate position of the solubility line^{73,75}. To account for anisotropic interactions in simulations, proteins can be modelled as particles with surface patches. This method is known as coarse-graining and it allows to represent particles at various levels of detail (with atomistic models being the most detailed) (Figure 1.16).

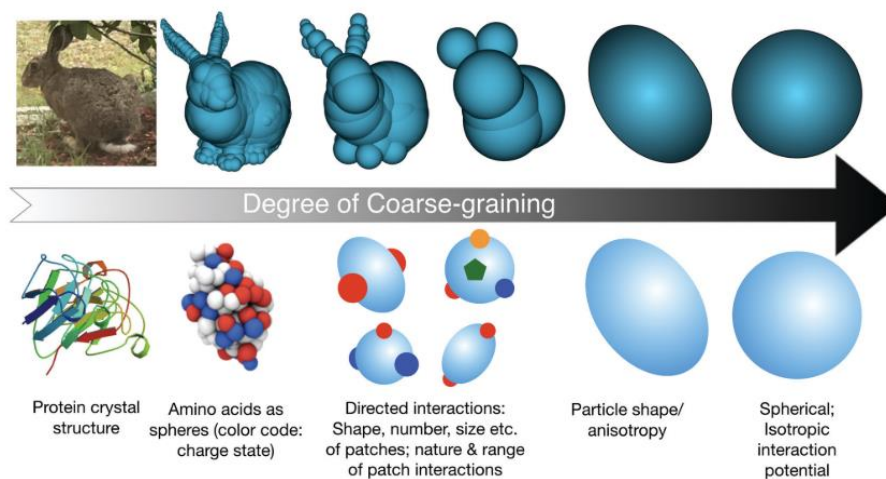


Figure 1.16: Illustration of protein coarse-grained models at different levels of detail. Image taken from ref. 95.

Unfortunately, calculating full protein phase diagrams from atomistic models is computationally prohibitive and hence patchy particles are employed instead⁹⁵. In these models two particles can interact only for certain orientations, i.e. when their respective patches are aligned. Orientation-dependant interactions can be defined in different ways, changing for example the number, size and shape of the patches or including different patch types^{73,74,75}. In general, simulations that consider anisotropic protein interactions better capture protein phase diagram than their isotropic counterparts (Figure 1.17)^{73,75,116}.

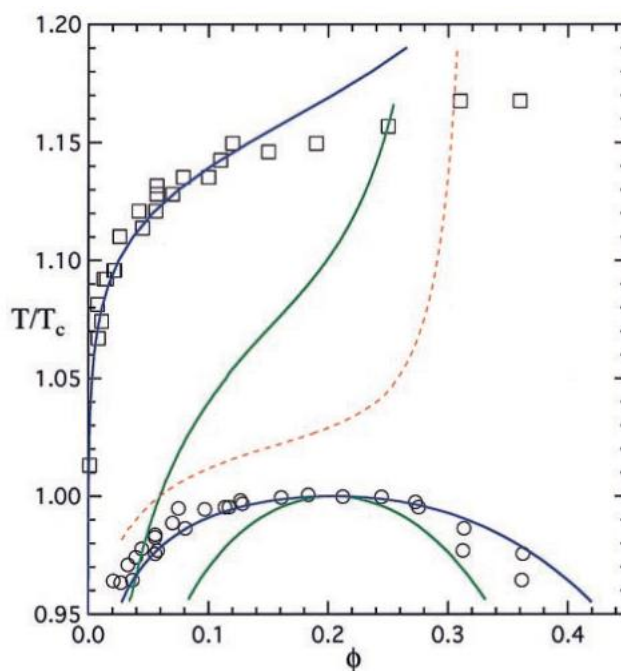


Figure 1.17: Theoretical predictions of a protein phase diagram assuming isotropic interactions (green lines) and a certain degree of anisotropy (blue lines). Open symbols are experimental data points of the solubility curve (squares) and LLPS (circle) of γ_{IIIb} -crystallin.

As shown in Figure 1.17 the solubility curve and the width of LLPS is quantitatively reproduced considering protein-protein interactions as anisotropic. However, the LLPS curve of monoclonal antibodies (mAbs) is asymmetric and much wider than that of model globular proteins such as Lysozyme or γ -crystallins; it was proposed that this is due to mAbs being Y-shaped and possibly due to their flexibility¹⁰⁶. However, a recent study found out that the particular shape of immunoglobulin G (IgG) contributes only marginally to LLPS boundary broadening, which is instead caused by the anisotropic nature of attractive interactions between the domains of the protein¹¹⁷.

Nevertheless patchy particle models have been successfully used to describe the phase behaviour of globular proteins: recently, the phase diagrams of single point mutants of Rubredoxin were reproduced with reasonable agreement parametrizing the patches from the knowledge of the protein crystal contacts¹¹⁸. Furthermore, the solubility curve inversion caused by a single surface mutation in Human γ D-crystallin (HGD) could be explained by a single temperature-dependent interaction in a patchy particle model¹¹⁹. Also, further experiments on HGD revealed that attaching hydrophobic patches on a protein surface increases protein-protein attraction raising the critical temperature of the LLPS boundary to an extent dependent on the position of the patch^{120,121}.

Despite being unsuitable to calculate whole protein phase diagrams, atomistic simulations can still be used to calculate protein-protein interactions (B_{22}) with models derived from X-Ray

protein structures^{122,123,124}. Neal *et al.* demonstrated that the value of B_{22} is orientation-dependent, and it is mostly determined by a few surface-complementary protein configurations¹²⁵. Experiments on lysozyme have shown that even single amino acid replacements can change the surface charge and hydrophobicity affecting the net protein-protein interactions (measured by B_{22})^{126,127}. However, this is not always the case since some single point mutations in γ -crystallins affect the crystallization behaviour and solubility curve maintaining the net protein-protein interaction in the liquid phase unaffected (measured by LLPS boundary)^{96,128}. To explain this, it has been proposed that anisotropic interactions caused by these specific single-point mutations are averaged out in the liquid phase and only engaged in the crystal phase⁹⁶.

1.3 Anisotropic interactions and protein crystallization

1.3.1 Protein crystallization

Proteins are very diverse and therefore there is no unique strategy for crystallization. However, it is well known that certain conditions are necessary: i) the formation of protein crystals must be thermodynamically possible, i.e. the protein concentration has to exceed the crystal solubility (supersaturated solution) and ii) the kinetics of crystallization from the supersaturated solution must be favourable for the process to eventually occur. These observations can be summarized in a generic crystallization phase diagram (Figure 1.18)¹²⁹.

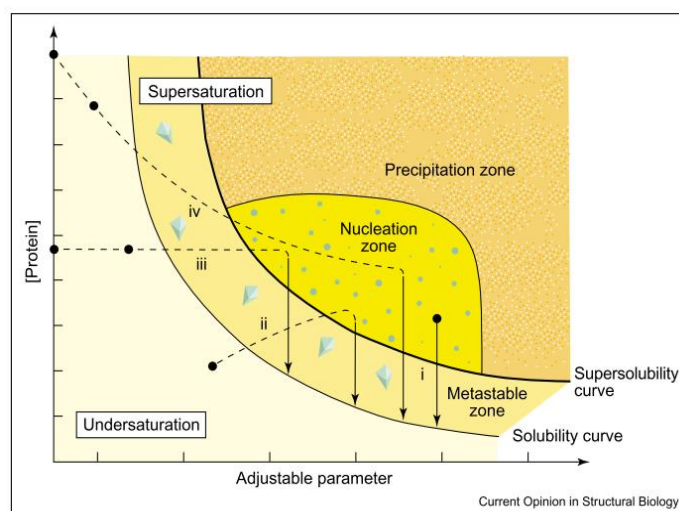


Figure 1.18: Sketch of a typical crystallization phase diagram: the arrows indicate the pathways of i) batch, ii) vapour diffusion, iii) dialysis and iv) free interface crystallization experiments. Image taken from ref. 130.

Unlike the protein phase diagrams previously discussed, the phase boundaries of crystallization phase diagrams are not at equilibrium, i.e. they are strongly dependent on the kinetics and crystallization method. They map the physical state of the protein against precipitant/salt

concentration (c/c) instead of temperature (c/T) (precipitants are additives used to increase protein supersaturation). The position of the supersolubility curve and the precipitation zone (non-equilibrium boundaries) is not well-defined since it also depends on the kinetics of the process⁶⁴. To obtain protein crystals it is necessary to exceed the solubility curve (metastable zone); unfortunately, this is not sufficient and also the kinetic barrier for crystal nucleation from the supersaturated solution has to be overcome (nucleation zone). Indeed, the crystalline phase is an ordered state and requires protein molecules to find the appropriate orientation while interacting: if the protein or precipitant concentration is too high, the kinetics of the process is too fast, resulting in amorphous (disordered) aggregation (precipitation zone)⁶³. This also means that protein-protein interactions have to be only moderately attractive: George and Wilson proposed that exists a universal range of B_{22} values where protein crystallization is more likely to occur¹³¹.

However, this fine balance of conditions varies from protein to protein and is still achieved by a trial and error approach: it involves screening hundreds of solution conditions, changing protein concentration, ionic strength, salt type, pH, precipitation additives, temperature using several different techniques such as vapour diffusion, batch and dialysis crystallization^{63,132}. Crystallization phase diagrams, such as the one sketched in Figure 1.18 can be built up during the screening process helping the experimenter to search and/or optimize the crystallization conditions^{133, 134}. Despite this, the trial-and-error approach has been facilitated over the years by the use of easy-to use crystallization kits, crystallization robots and microfluidics^{63,135}, but it still demands long times and significant resources. Therefore, a better understanding of the thermodynamic and kinetics of protein crystallization is needed to at least narrow the screening space. The computational efforts of soft matter physics in reproducing protein phase diagram (c/T) described above, along with the experimental knowledge of protein-protein interactions (B_{22} , DLVO theory) are the best means at present to describe the thermodynamic starting conditions required for protein crystallization.

1.3.2 Thermodynamics of protein crystallization

The formation of protein crystals from a supersaturated solution is only possible for those conditions where the crystal phase is thermodynamically stable, then when the Gibbs free energy of crystallization ΔG_{cryst}^0 is negative. This can be expressed by Gibbs-Helmholtz equation valid at constant pressure:

$$\Delta G_{cryst}^0 = \Delta H_{cryst}^0 - T\Delta S_{cryst}^0 \quad 1.8$$

The Gibbs free energy of crystallization must be negative for the phase transition to occur. For proteins, this has been shown to be only moderately negative, between -10 and -100 kJ mol^{-1} ¹³⁶, differently from inorganic salts¹³⁷. If ΔG_{cryst}^0 becomes positive the process is not thermodynamically possible, explaining why small variations in solution conditions are known to dramatically influence protein crystallization¹³⁷.

It is also possible to evaluate the single contributions to ΔG_{cryst}^0 . Indeed, at constant temperature, protein crystallization would be favoured by very negative ΔH_{cryst}^0 or very positive ΔS_{cryst}^0 . However, the crystallization enthalpy has been measured for some protein^{138,139,140}, and its values are generally only moderately negative. Therefore, the enthalpic stabilization is not the major thermodynamic contribution that allows proteins to crystallize¹³⁷. This is in also in agreement with the knowledge that protein crystals are built up only by weak, non-covalent interactions^{141,142}. Therefore, the entropy value has to be as high as possible to allow ΔG_{cryst}^0 to be negative. This is counterintuitive since the crystalline ordered phase must have a lower entropy than the liquid phase (due the loss of rotational and translational degrees of freedom¹⁴³) making ΔS_{cryst}^0 negative and then disfavouring crystallization¹³⁷. The favourable contribution must then come from the solvent; Equation 1.8 can be also written as:

$$\Delta G_{cryst}^0 = \Delta H_{cryst}^0 - T(\Delta S_{protein}^0 + \Delta S_{solvent}^0)_{cryst} \quad 1.9$$

Since $\Delta S_{protein}^0$ is negative, ΔG_{cryst}^0 must be decreased by very high and positive values of $\Delta S_{solvent}^0$. It is known that water molecules are structured around a protein surface creating a shell of a few Angstroms with different properties than bulk water^{144,145}. During crystallization, the release of water molecules increases the $\Delta S_{solvent}^0$ contribution, with the result of decreasing the overall ΔG_{cryst}^0 of the process¹³⁷. The two entropic contributions can be divided and $\Delta S_{protein}^0$ measured by monitoring the growth of protein crystals from atomic force microscopy (AFM). The value of $\Delta S_{protein}^0$ can be as high as -100 $\text{J mol}^{-1} \text{K}^{-1}$ ^{137,139} while $\Delta S_{solvent}^0$ has been estimated to be between 100 and 600 $\text{J mol}^{-1} \text{K}^{-1}$ and corresponds to the release of approximately 5-30 water molecules per protein¹³⁶. Therefore, the exclusion of surface water during the creation of protein-protein interfaces is most likely to be the thermodynamic driving force of protein crystallization^{136,139}.

1.3.3 Kinetics of protein crystallization

Once that the conditions required for crystallization are satisfied (i.e. the solution is supersaturated, and in the $L+S$ region of the c/T phase diagram), the new solid phase can spontaneously emerge from the metastable homogeneous liquid; this process is called nucleation.

Nucleation determines what polymorphs may form, the number of crystals and their size^{146,147} and it is dramatically influenced by slight changes in solution conditions. However, nucleation timescales can be very large, so that crystallization may remain experimentally inaccessible¹⁴⁶. Therefore, in addition to the thermodynamic requirements, the kinetics of nucleation must be fast enough to allow crystallization. The simplest model developed to evaluate the kinetics of this process is the Classical Nucleation Theory (CNT). Considering the nucleus as a sphere, the Gibbs free energy of nucleation is given by Equation 1.10¹⁴⁸:

$$\Delta G = -\frac{4\pi r^3}{3\Omega} k_B T \ln \beta + 4\pi r^2 \gamma \quad 1.10$$

where r is the radius of the nucleus, Ω is the volume of a single molecule, k_B is the Boltzmann constant, T the temperature, β the supersaturation (the concentration divided by the equilibrium one at the same temperature), and γ the surface tension of the nucleus. From the first term in Equation 1.10 (volume term) we can see that nucleation is enhanced for high values of the nucleus radius and of supersaturation. The second term (surface term) relates to the unfavourable creation of a new surface in the liquid and hence hinders nucleation; this also depends on r , but it scales more slowly. The competition between the favourable and unfavourable terms generates a maximum in the function as shown in Figure 1.19:

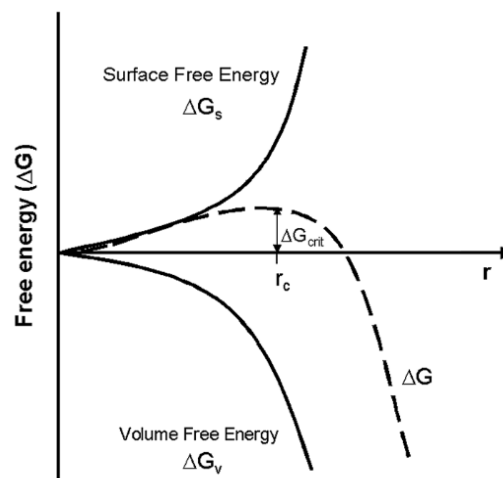


Figure 1.19: Nucleation free energy barrier ΔG_{crit} . Its height depends on the opposite contributions of the surface and volume terms. Image taken from ref. 149.

The maximum value of the function, ΔG_{crit} , gives the height of the nucleation free energy barrier. Its existence explains why even if the solution meets the thermodynamic requirements, crystallization does not occur unless certain values of supersaturation are reached¹⁵⁰. The nucleation rate, J , depends on the height of the nucleation free energy barrier and can be expressed as¹⁴⁸:

$$J = k_0 \exp\left(-\frac{\Delta G_{crit}}{k_B T}\right) \quad 1.11$$

where the pre-exponential factor k_0 is the product of solution concentration, the rate of attachment of new molecules to the nucleus and a probabilistic factor (Zeldovich factor)¹⁴⁷.

This treatment is strictly valid only for homogeneous nucleation, i.e. occurring in the bulk of the solution. However, nucleation is observed most commonly at the interface of a solid surface like the walls of the vessel or around impurities in the solution (heterogeneous nucleation). In this cases, the nucleation barrier is decreased by an extent proportional to the contact angle between the nucleus and the surface¹⁴⁸.

It has been shown that in some cases CNT overestimates the homogeneous nucleation rate of protein crystals by up to ten orders of magnitude¹⁴⁷. CNT is a generic model and applies to several different systems. For instance, it can be used to describe the nucleation of the solid/crystal phase from a metastable liquid as well as to nucleation of liquid droplets from a metastable vapour. However, while the latter phase transition involves only a density change, the former requires a change in density and structure (crystallinity). Both of these properties are order parameters and are used to unequivocally identify the phase that the system adopts. CNT implicitly assumes that the nucleus is formed from a dilute metastable solution by simultaneous fluctuations of both the order parameters, so that by the time the nucleus reaches the critical radius, r_c , it is already ordered and crystalline (one-step mechanism)¹⁴⁷. However, Wolde and Frenkel demonstrated by simulations that this mechanism describes protein crystal nucleation only far away from the LLPS critical point (either at lower or higher temperatures than LLPS critical temperature T_c). Instead, in this vicinity the two order parameters are separated: first, the density increases leading to the formation of the dense liquid droplets, and then crystallization (increase of structure) occurs within these droplets (two-step mechanism)^{86,147}. These two pathways are compared in Figure 1.20:

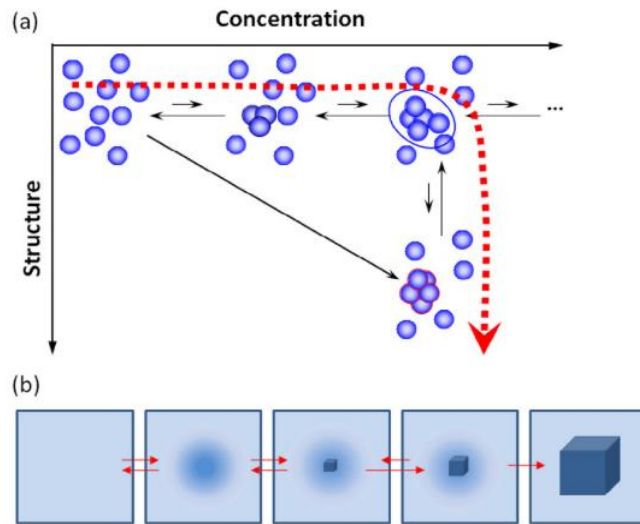


Figure 1.20: A) Comparison of one-step (black arrow) and two-step mechanisms (dashed red arrow) on the density (or concentration)-structure plane. B) Crystal growing from a dense liquid droplet as described by the two-step mechanism. Image taken from ref.147.

Both simulations and experiments show that the nucleation rate is increased in the vicinity of the critical point supporting the two-step mechanism^{86,151} and indeed it has been shown that protein crystals can grow from dense liquid droplets¹⁵². This case occurs when the dense liquid is a stable phase in respect of the dilute liquid phase (Figure 1.21, lower curve). However, a more common situation is that the dense liquid is unstable with respect to the dilute solution (Figure 1.21, upper curve): when this happens, the dense liquid is formed within mesoscopic clusters from which crystallization occurs¹⁴⁷.

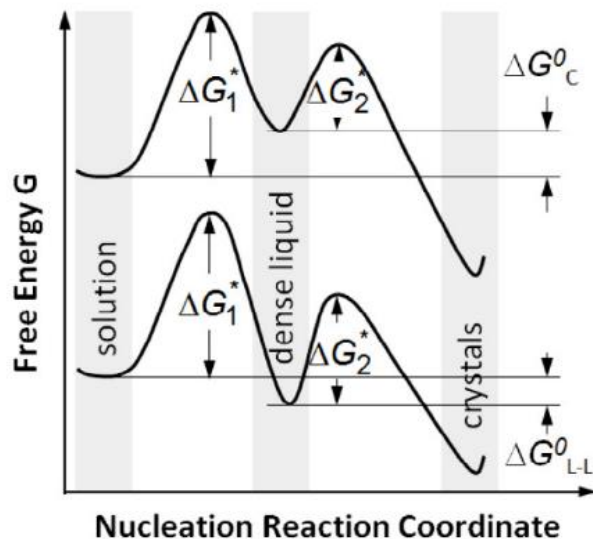


Figure 1.21: Two-step nucleation of protein crystals occurs A) from mesoscopic clusters or B) within dense liquid droplets if the dense liquid phase is respectively unstable or stable in respect of the dilute liquid. Image taken from ref.147.

1.3.4 Anisotropic interactions to improve protein crystallization

Protein crystals are supported by weak intermolecular interactions. However, not all of the surface of the protein is involved, and the interface formed by two protein molecules in a crystal is known as crystal contact. These interfaces are purely artefacts of the crystallization process and therefore have a different role from the biologically relevant contacts formed under physiological conditions¹⁵³. The formation of crystal contacts is dictated to a certain degree from geometrical complementarity and by weak intermolecular non-covalent interactions between the two surfaces (hydrogen bonds, salt bridges, van der Waals and hydration forces)¹⁴². Crystal contacts are generally hydrophilic and polar¹⁴². The energy involved in their formation is only a few $k_B T$ (at room temperature 1 $k_B T$ corresponds to 2.479 kJ/mol), much less than the binding energies associated with biologically relevant interfaces¹⁴². This allows the intermolecular bonds to be broken and reformed until the proper orientations necessary for the ordered crystal phase are obtained¹⁴¹. On the contrary, interactions that are too strong lead to the formation of protein gels or amorphous aggregate¹⁴¹.

The interactions involved in protein crystal contacts are inherently anisotropic, i.e. they depend on the heterogeneous amino acid composition and shape of protein surface. Hence, variations of these interfaces can dramatically affect nucleation, growth, solubility and polymorphism of protein crystals. This fact was empirically known in the past and, if a protein from a species was not crystallisable, an homologous (the same protein from another species that differs only of few amino acids) would have been tried⁷⁶. This approach has been abandoned and instead protein surface modifications, such as reductive methylation of lysines or site-directed mutagenesis, have been used to increase protein propensity to crystallize^{76,154,155}. In recent years it has been shown that decreasing the side-chain conformational entropy of flexible surface amino acids, such as lysine or glutamic acid, by mutation to alanine often results in increasing the likelihood of crystallization. This approach, known as Surface Entropy Reduction (SER), diminishes the entropy loss of protein crystallization and consequently decreases the Gibbs free energy of the process¹⁵⁵. The SER strategy has been firstly demonstrated by a mutagenesis study on the globular domain of the protein RhoGDI: even a single lysine to alanine mutation promoted crystallization and the crystal structures revealed that the mutated residues were effectively involved in crystal contacts¹⁵⁶. Ever since, the SER strategy has been widely used to obtain protein crystals¹⁵⁷ further demonstrating the sensitivity of anisotropic protein-protein interactions.

1.4 Self-assembly of virus capsid and anisotropic interactions

As previously stated, viruses are supramolecular assemblies of protein, forming rod-shape, or icosahedral structures. In infected cells, viral particles assembly is assisted by the host cellular machinery by the consumption of adenosine triphosphate (ATP)²⁰; however, virus structure can

also be obtained *in vitro* by self-assembly both in the presence or absence of a genome. In the first case, only protein-protein interactions are involved, and the process proceeds via a nucleation and growth mechanism, i.e. an initial ordered nucleus forms and then other protein subunits add until the capsid is completed¹⁵⁸. However, if the nucleic acid is present, its negatively charged backbone interacts with the positively charged proteins. The competition between protein-protein and protein-nucleic acid interactions results in two different self-assembly mechanisms: at high ionic strengths, when protein-protein interactions prevail, the capsid self-assembly still occurs via nucleation and growth; at low ionic strengths instead, when the protein-nucleic acid interactions are stronger, the capsid is formed by an *en masse* mechanism, where the protein subunits adsorb onto the genome in a disordered way and then rearrange to form the ordered capsid structure^{21,159}.

In nature, viral capsids are perfect geometrical structures and their formation is driven by the building of specific interfaces between capsid proteins (CPs). Several studies show that these interfaces are extremely sensitive to single-point mutations, hindering capsid formation or stability^{160,161}; more specifically, recent work revealed that only mutations on interface-conserved residues (so called hot-spots) significantly impact capsid formation^{162,163}. Hence, given the high specificity of the process, viral capsid assembly provides a further example of the importance of anisotropy in protein self-assembly.

Given the importance of anisotropic interactions in protein self-assembly, the aim in this thesis is to understand them further; in chapter 3, the effects of a specific surface amino acid of HGD, Cys-110, on the self-assembly behaviour is studied while, in Chapter 4, we investigated how two independent surface mutations on the same protein generate a more complex, polymorphic, self-assembly behaviour. Finally, in chapter 5, the effects of surface modifications on the thermostability of ADDomer virus like particles are studied.

Chapter 2:

Materials and Methods

2.1 Preparation of buffers and reagents

2.1.1 Buffers

All buffers were prepared dissolving the appropriate quantity of salts in Milli-Q water. The pH was adjusted with concentrated hydrochloric acid (HCl, $M_w=36.46$ g/mol) (Fisher Scientific, UK) or sodium hydroxide (NaOH, $M_w=40.00$ g/mol) (Fisher Scientific, UK) when required. Sodium azide (NaN_3 , $M_w=65.01$ g/mol) (Fisher Scientific, UK) was added to any buffer at a concentration of 0.02% (w/v) to inhibit microbial growth. Prior to use, buffers were filtered through 0.45 μm nylon filters (Millipore, Ireland) and degassed under vacuum with stirring. When sterile conditions were needed, buffer and reagents were filtered through 0.22 μm Millex-GV syringe driven filter units (Millipore, Ireland).

2.1.1.1 Sodium acetate

0.275 M Sodium acetate was prepared either at pH 4.5 or 4.8 by adding 16.50g of glacial acetic acid (CH_3COOH , $M_w=60.05$ g/mol) (Fisher Scientific, UK) and 4.48g of NaOH per litre of water. Sodium acetate buffer at pH 4.8 was prepared with the same amount of glacial acetic acid, 6.70 g of NaOH and supplemented with 19.02 g of sodium chloride (NaCl, $M_w=58.44$ g/mol, 0.32 M) (Fisher Scientific, UK) per litre of water.

2.1.1.2 Sodium phosphate

0.1 M Sodium phosphate buffer was prepared dissolving 14.2 g of sodium phosphate dibasic anhydrous (Na_2HPO_4 , $M_w=141.96$ g/mol) (Fisher Scientific, UK) and 15.60 g of sodium phosphate monobasic dihydrate (NaH_2PO_4 , $M_w=156.01$ g/mol) (Fisher Scientific, UK) per litre of water. The pH was brought to 7 using concentrated NaOH.

2.1.1.3 TE buffer

TE buffer was prepared by dissolving 0.029 g of ethylenediaminetetraacetic acid (EDTA) ($\text{C}_{10}\text{H}_{16}\text{N}_2\text{O}_8$, $M_w=292.23$ g/mol, 1 mM) (Fisher Scientific, USA) and 0.12 g of tris(hydroxymethyl)aminomethane (Tris-HCl) ($\text{C}_4\text{H}_{11}\text{NO}_3$, $M_w=121.14$ g/mol, 10 mM) (Fisher Scientific, UK) in 100 mL of Milli-Q water and bringing the pH to 8 with NaOH.

2.1.2 Buffer and reagents for *E. coli* cultures

All buffers, media and reagents for *E. coli* cultures were sterilized before use. Sterilization was carried out with a SX-500E TOMY autoclave (Seiko, Japan) at 121°C, 0.212 MPa pressure for 20 minutes. For small volumes, sterilization was carried out by filtering the solution through 0.22 μm Millex-GV syringe driven filters.

2.1.2.1 Ampicillin stock solution

100mg/mL Ampicillin stock solutions were prepared by dissolving 1 g of Ampicillin salt ($C_{16}H_{18}N_3NaO_4S$, $M_w=371.39$ g/mol) (Fisher Scientific, USA) in 10 mL of ultrapure water. The solutions were filter sterilized and stored at $-20^{\circ}C$.

2.1.2.2 IPTG stock solution

1M Isopropyl β -D-1-thiogalactopyranoside (IPTG) stock solutions were prepared by dissolving 2.38 g of IPTG ($C_9H_{18}O_5S$, $M_w=238.31$ g/mol) (Fisher Scientific, UK) in 10 mL of ultrapure water. Solutions were filter sterilized and stored at $-20^{\circ}C$.

2.1.2.3 LB broth

Luria-Bertani broth (LB) was prepared by dissolving 25 g of LB Miller powder (Fisher Scientific, USA) per litre of water and autoclaved. Ampicillin was added to a final concentration of 100 μ g/mL when the medium temperature cooled to below $50^{\circ}C$.

2.1.2.4 LB Agar plates

LB Agar plates were prepared dissolving 37 g of LB Agar granules (Fisher Scientific, UK) in 1 litre of ultrapure water and autoclaved. Ampicillin was added after autoclaving to a final concentration of 100 μ g/mL when the medium cooled to below $50^{\circ}C$. The LB Agar was then poured into sterile petri dishes and allowed to settle. Plates were sealed and stored at $-4^{\circ}C$ and used within one week.

2.1.2.5 NZY+ broth

NZY⁺ was prepared dissolving 10 g of NZ Amide (Sigma Aldrich, USA), 5 g of yeast extract (Oxoid, England) and 5 g of NaCl per litre of water. The pH was adjusted to 7.5 and the media was then autoclaved. After sterilization the media was enriched with previously filter sterilized 12.5 mL of 1 M magnesium chloride ($MgCl_2 \cdot 6H_2O$, $M_w=203.30$ g/mol. 4.06 g per 20 mL) (Fisher Scientific, USA), 12.5 mL of 1M magnesium sulfate (Mg_2SO_4 , $M_w=120.361$ g/mol) (Fisher Scientific, USA) and 10 mL of 2 M glucose ($C_6H_{12}O_6$, $M_w=180.16$ g/mol)

2.1.2.6 Lysis buffer

Lysis buffer for bacterial cells was prepared by dissolving 0.0788 g of Tris-HCl (50 mM) (Fisher Scientific, UK), 0.0146 g of NaCl (25 mM) and 0.0074 g of EDTA (2 mM) (Fisher Scientific, USA) in 10 mL of ultrapure water.

2.1.3 Buffers and reagents for polyacrylamide gel electrophoresis (SDS-PAGE)

2.1.3.1 Tris Buffer

1 M Tris buffer, pH 8.8 and 0.5 M Tris buffer pH 6.8 were prepared by dissolving 12.11 g and 6.06 g respectively of tris(hydroxymethyl)aminomethane base salt ($M_w=121.1$ g/mol) (Fisher Scientific, UK) in 100 mL of Milli-Q water. The proper pH was reached by titrating with a concentrated HCl solution.

2.1.3.2 Sample buffer

Sample buffer was prepared by mixing 1.2 mL of 0.5M Tris pH 6.8 buffer, 4.8 mL of Milli-Q water, 1 mL of Glycerol (Sigma Aldrich, USA), 2 mL of 10% (w/v) Sodium dodecyl sulfate (SDS, $\text{NaC}_{12}\text{H}_{25}\text{SO}_4$, $M_w=288.37$ g/mol) (Fisher Scientific, UK) and 0.5 mL of 0.1% (w/v) of bromophenol blue ($\text{C}_{19}\text{H}_{10}\text{Br}_4\text{O}_5\text{S}$, $M_w=669.96$ g/mol) (Sigma Aldrich, Germany). Reducing sample buffer was freshly prepared by mixing 475 μL of sample buffer and 25 μL of β -mercaptoethanol (0.05% (v/v), $\text{C}_2\text{H}_6\text{OS}$, $M_w=78.13$ g/mol) (Sigma Aldrich, Belgium).

2.1.3.3 Destaining solution

De-staining solution was prepared by mixing Milli-Q water, glacial acetic acid and methanol (CH_3OH , $M_w=32.04$ g/mol) (Sigma Aldrich, UK) in a 60:10:30 ratio.

2.1.3.4 Molecular weight standards solution

2.1.1.1 Molecular weight standards solution

Molecular weight standards for SDS-PAGE low range solutions was purchased from BioRad (USA) (Figure 2.1). 10 μL aliquots were made by diluting the stock solution 1:20 with the stock sample buffer and were stored at -20°C until use. Each aliquot was heated to 95°C for 5 minutes, cooled and briefly centrifuged before use.

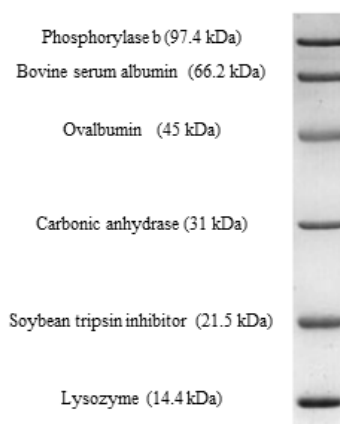


Figure 2.1: SDS-PAGE low molecular weight standard proteins

2.2 Recombinant protein production

2.2.1 Plasmid mutation with site-directed mutagenesis

Human γ D-crystallin (HGD) and its mutant P23V plasmids (already available in the lab) were used as templates, with appropriate primers, to mutate cysteine on position 110 to methionine by site-directed mutagenesis. This allowed plasmids coding for the expression of the single mutants C110M, C110S and the double mutant P23VC110M. The oligonucleotide primers used were synthesized by Life technologies (Ireland) and are listed below:

C110M:

Forward Primer: 5'- C ACT GAG GAC TGC TCC ATG CTT CAG GAC CGC TTC CG - 3'

Reverse Primer: 5'- CG GAA GCG GTC CTG AAG CAT GGA GCA GTC CTC AGT G -3'

C110S:

Forward Primer: 5'- CT GAG GAC TGC TCC AGT CTT CAG CGC-3'

Reverse Primer: 5'- GCG GTC CTG AAG ACT GGA GCA GTC CTC AG-3'

Site-directed mutagenesis was performed with the QuickChange II site-directed mutagenesis kit (Stratagene, USA) following the manufacturer guidelines. The polymerase chain reaction (PCR) solution was prepared mixing the reagents in a 0.5 mL PCR tube (Eppendorf, Germany) on ice and bringing the final volume to 50 μ L with MilliQ water.

Table 2.1: Components of PCR solutions

Solution component	Quantity
10X reaction buffer	5.00 μ L
dsDNA template	50 ng
Forward primer	125 ng
Reverse primer	125 ng
dNTP mix	1.00 μ L
Pfu Ultra HF DNA polymerase	1.00 μ L

The solution was gently mixed and subjected to temperature cycles with a ³Prime thermal cycler (Bibby Scientific, UK) using the steps in Table 2.2.

Table 2.2: PCR cycling parameters.

Step	Duration (minutes)	Temperature (°C)	Cycles
1 (lid preheating)	2	105	1
2	1	95	1
3 (denaturation)	0.5	95	16
4 (annealing)	1	65	
5 (extension)	9	68	

The reaction tube was then incubated on ice for 2 minutes. 1 μ L of Dpn I restriction enzyme was added and the amplification reaction was gently mixed by pipetting. The solution was then centrifuged for 1 minute and incubated at 37°C for one hour to completely digest parental DNA.

2.2.2 Mutated plasmid transformation and amplification

Mutated plasmid DNA was then transformed in XL1-Blue supercompetent cells (Agilent, USA) for nicks repair as per manufacturer's instructions. XL1-Blue supercompetent cells were thawed on ice and 50 μ L of the suspension was added to a pre-chilled polypropylene 15 mL falcon tube. 1 μ L of the Dpn I-digested DNA was added and the entire reaction was gently mixed and incubated on ice for 30 minutes. The transformation reaction was heat-pulsed for 45 seconds in a 42°C water bath and then incubated on ice for 2 minutes. The transformation reaction was then divided into two LB agar plates containing ampicillin and incubated overnight at 37°C. A single colony from each transformation plate was used to inoculate a 4 mL LB culture containing ampicillin to amplify plasmid DNA. This culture was grown for 8 hours at 37 °C with shaking at 300 RPM. 25 mL LB culture containing ampicillin was inoculated with 50 μ L of the starter culture and incubated at 37°C with shaking a 300 RPM overnight. These amplification steps were carried out in vessels at least 4 times bigger than the culture volume to ensure cells grew up to 3-4 x10⁹ cells/mL, necessary to obtain a satisfactory plasmid yield. The cell pellet was harvested by centrifugation at 7000 RPM for 15 min at 4 °C and used for plasmid extraction and purification.

2.2.3 Mutated plasmid extraction, purification and sequencing

Plasmid DNA was extracted from cells and purified after the amplification steps using a QIAGEN Plasmid Purification Midi Kit (QIAGEN GmbH, Germany) as per manufacturer's instructions. The bacterial pellet was resuspended in 4 mL of buffer P1 containing RNase A. 4 mL of buffer P2 was added, the solution was thoroughly mixed and incubated at room temperature for 5 minutes. 4 mL of pre-chilled buffer P3 was added, the solution mixed and incubated on ice for 15 minutes. The mixture was then separated from parental DNA by centrifuging at 13000 RPM for

30 minutes at 4°C. Small particles were further removed from the supernatant with a second centrifugation step for 15 minutes. After equilibration of the QIAGEN-tip 100 with 4 mL of buffer QBT the supernatant from centrifugation steps was loaded onto the column and allowed to flow through. Low salt conditions provided by buffer QBT allowed the plasmid DNA to bind to the resin. Impurities were washed away from the column with 2 washes of 10 mL buffer QC. Plasmid DNA elution was ensured by loading the column with 5 mL of buffer QF that provides high salt conditions. Plasmid DNA was then precipitated by adding 3.5 mL of room temperature isopropanol ($\text{CH}_3\text{CHOHCH}_3$, $M_w=60.1$ g/mol) (Sigma Aldrich, UK) and mixing the solution. After centrifugation at 13000 RPM for 30 minutes at 4°C a small amount of white precipitate appeared on the bottom of the falcon tube. The supernatant was carefully decanted and the precipitate washed with 2 mL of room temperature 70% (v/v) ethanol solution ($\text{CH}_3\text{CH}_2\text{OH}$, $M_w=46.07$ g/mol) (Sigma Aldrich, UK) to remove salts. The solution was subjected to another centrifugation step at 13000 RPM for 10 minutes. After decanting the supernatant, the plasmid DNA precipitate was air-dried for 5-10 minutes and redissolved in 0.5-1 mL of TE buffer. Purified plasmid DNA concentration and purity was determined spectrophotometrically using a Molecular Devices SpectraMax M2e (Molecular Devices, USA) using equation 2.1.

$$\text{concentration } (\mu\text{g/mL}) = A_{260} \cdot DF \cdot 50 \mu\text{g/mL} \quad 2.1$$

where A_{260} is the absorbance at 260 nm and DF is the dilution factor. The purity of DNA was calculated from the ratios A_{260}/A_{280} and A_{260}/A_{230} . When these ratios are ≥ 1.8 and ≥ 2 respectively, DNA is considered pure. An aliquot of purified plasmid corresponding to approximately 600 ng was washed with Amicon-0.5 Ultra centrifugal device with a 10kDa membrane (Fisher Scientific, UK) to wash EDTA away and resuspended in water to a final concentration of approximately 30 ng/ μL . The plasmid solution so prepared was sequenced with T7 promoter primer using an automated capillary DNA sequencer (Medical Research Council Protein Phosphorylation and Ubiquitylation Unit, University of Dundee, Scotland).

2.2.4 Purified plasmid transformation in *E. coli* competent cells

All purified plasmids, containing coding sequences for ampicillin resistance and Isopropyl β -D-1-thiogalactopyranoside (IPTG) mediated protein expression, were transformed in BL21-Gold (DE3) competent *E. coli* cells (Agilent, USA) suitable for protein production. 100 μL of competent cells were thawed on ice and transferred in a 15 mL falcon polypropylene tube. 50 ng of HGD plasmid were added to the competent cells and the mixture were incubated on ice for 30 minutes. The transformation reaction was heat-pulsed in a 42°C water bath for 20 seconds and then incubated on ice for 2 minutes. 0.9 mL of preheated (42°C) SOC medium was added and the transformation reaction was incubated at 37°C for 1 hour shaking at 250 rpm in a Innova 42

Incubator (New Brunswick Scientific, USA). The transformation reaction was then plated on LB Agar plate containing ampicillin and incubated overnight at 37°C. Stocks of successfully transformed cells were prepared picking one colony from the transformation plate to inoculate 100 mL of LB broth containing ampicillin and incubating it overnight at 37°C shaking at 225 RPM. The stocks were stored at -80°C.

2.2.5 Recombinant protein expression in bacterial cultures

Stocks of *E. coli* cells containing the appropriate plasmid were used to streak a freshly prepared LB Agar plate containing ampicillin. The streaked plate was then incubated overnight at 37°C overnight until ampicillin-resistant colonies appeared. Cells from a single colony from the plate were taken to inoculate a 200 mL LB culture. The culture was grown overnight at 37°C shaking at 225 RPM and then used to inoculate 6 large flasks containing 1.2 L of LB broth. The cultures were grown at 37°C with shaking at 225 RPM for approximately four hours until the optical density (OD) was between 0.8 and 1. 1 mL of 1M IPTG was then added to each flask to trigger protein expression and was allowed to continue for 4-5 hours. The cultures were then centrifuged in a Heraeus Megafuge 40R centrifuge (ThermoFisher Scientific, USA) using 750mL Nalgene bottles and a TX-750 swinging bucket rotor (ThermoFisher Scientific, USA) at 3000 RPM for 10 minutes. The supernatant was discarded and the pelleted cells were stored at -80°C.

2.2.6 Recombinant protein extraction from bacterial cells

Pelleted cells were thawed in a water bath at 30°C. 10 mL of lysis buffer supplemented with 155 mg of dithiothreitol (DTT, $M_w=154.35$ g/mol) and 1 Roche Protease Inhibitor Cocktail tablet (Roche Diagnostics, USA) was divided into the 4 x 750mL Nalgene bottles. The cells were re-suspended in the lysis buffer by vortexing, combined into one bottle and incubated at room temperature for 2 hours. 160 μ L of 50 mg/mL lysozyme solution (Calbiochem, Canada) were added. The cell suspension was then vortexed and incubated at room temperature for a further 30 minutes. The suspension was subjected to four freeze-thaw cycles using liquid nitrogen and a water bath at 30°C. 2 mL of freshly prepared 1mg/mL deoxyribonuclease I (DNase I, bovine pancreas) (Calbiochem, Canada) was mixed with 1mL of 1 M magnesium sulfate and added to the cell suspension. This was mixed thoroughly and incubated at room temperature for 1-1.5 hours ensuring no viscous material was present. The pH of the cell suspension was then adjusted to 4.5 using glacial acetic acid, transferred to a 25 mL tube and centrifuged overnight at 8.500 RPM at 4 °C in a Heraeus 3R+ Multifuge centrifuge. The supernatant containing the soluble protein of interest was immediately used in the first purification step.

2.3 Techniques for protein purification and purity assessment

2.3.1 Size exclusion chromatography (SEC)

2.3.1.1 Theory

Size exclusion or gel filtration chromatography (SEC) is an analytical technique that allows the separation of molecules depending on their hydrodynamic radius. This technique finds its major application in the field of macromolecules such as proteins or polymers. In the case of proteins, SEC is often used for purification, purity analysis and aggregation studies. The stationary phase used in size exclusion columns is a porous resin that traps small molecule and allows large molecules to elute first. This selective phenomenon permits the separation of a polydisperse samples of macromolecules in its single components. SEC differs from all the other chromatographic techniques since no adsorption to the stationary phase is involved. The thermodynamics of a generic chromatographic separation can be described considering the change in Gibbs free energy of the process¹⁶⁴:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = RT \ln k \quad 2.2$$

where ΔG^0 , ΔH^0 , ΔS^0 and k are respectively the change in Gibbs free energy, enthalpy, entropy and partition coefficient between the stationary and mobile phase due to adsorption phenomena. Ideally, SEC separation is totally entropic since there is no change in enthalpy between the two phases due to the absence of interactions with the column resin ($\Delta H^0=0$). For this reason equation 2.2 becomes:

$$\ln k = -\frac{\Delta S^0}{R} \quad 2.3$$

As can be noted in equation 2.3, the partition coefficient k is not related to temperature directly. However, temperature can influence mobile phase viscosity and sample diffusivity impacting on the separation. The partition coefficient k can be written as:

$$k = \frac{V_R - V_0}{V_i} \quad 2.4$$

where V_R , V_0 and V_i are respectively the retention volume of the analyte, the interstitial volume and the intra-particle volume. k will range from 0, when the analyte is completely excluded, to 1, when the pore space can be occupied completely from the analyte¹⁶⁴. The partition coefficient is linked to the hydrodynamic radius of the molecules. It is then possible to relate the elution volume (or time) to the molecular weight of the particle building a calibration curve with molecules of known molecular weight.

2.3.1.2 Instrumentation

For preparative purposes a XK50/100 column (GE Healthcare Bio-Sciences, Sweden) connected with an AKTA purifier chromatographic system (GE Healthcare Bio-Sciences, Sweden) was used. The chromatographic system was composed of 2 pumps, a fraction collector, spectrophotometric and conductivity detectors. The column was packed with Sephacryl S-100 High Resolution media (GE Healthcare Bio-Sciences, Sweden), a packing material composed of cross-linked copolymer of allyl dextran and N,N-methylene bisacrylamide that allows separation of molecules in the molecular weight range of 1000-100000Da. The column was packed and always used at a flow rate of 2 mL/min, the maximum pressure limit set by the chromatographic system was 1 MPa. The column was equilibrated with 5 column volumes of 0.275 M sodium acetate buffer pH 4.5 containing 0.02% (w/v) of sodium azide and stored in the same buffer. The column was washed either before purifying a new protein or after 5 uses with the same protein with 1 column volume of 0.2 M NaOH and re-equilibrated as previously described.

2.3.1.3 Sample preparation and elution

The above-mentioned apparatus was exclusively used for the purification of HGD and its mutants. The cell lysate (approximately 20 mL) resulting from the procedure described in paragraph 2.2.6 was filtered through 0.22 μ m Millex-GV syringe driven filters and loaded on to the size exclusion column. The sample was isocratically separated with 0.275 M sodium acetate pH 4.5 as the buffer, accordingly to the following method.

Table 2.3: Method for SEC separation of HGD and its mutants.

Step	Volume (mL)	Fraction volume (mL)
1	0-400	0
2	400-2200	22

2.3.2 Ion exchange chromatography (IEX)

2.3.2.1 Theory

Ion exchange chromatography (IEX) separates proteins depending on their net charge. The overall charge of a protein in a specific environment is either due to the chemical composition of its surface, and so to the primary structure, and to the pH of the solvent. Indeed, the isoelectric point (pI) of a protein is defined as the pH at which the protein has no net charge due to the balance of negative and positive surface groups: at $\text{pH} > \text{pI}$ the protein will have a negative net-charge whereas at $\text{pH} < \text{pI}$ it will be positively charged. Using anionic (cation-exchange) or cationic

(anion-exchange) resins and varying pH and salt condition during the elution it is so possible to separate protein with different isoelectric points¹. An Ion exchange chromatography procedure is composed of four steps:

- **Equilibration:** the ion exchange column is washed with a buffer with low ionic strength to ensure the highest amount of proteins will bind to the resin.
- **Sample application and washing of unbound proteins:** the sample is loaded on column and the equilibration buffer is flown to remove any protein with low binding affinity to the column.
- **Elution:** the ionic strength of the buffer is slowly increased to displace protein molecules from the ionic resin. This is usually obtained with a linear gradient elution by mixing the initial, low ionic strength solvent with a second, high ionic strength solvent.
- **Regeneration:** After the elution of the target protein, a high ionic strength solvent is pumped through to ensure all analytes are removed from the resin. After this step the column can be equilibrated again with the low ionic strength solvent.

2.3.2.2 Instrumentation

A XK50/60 column (GE Healthcare Bio-Sciences, Sweden) was packed with SP Sepharose Fast Flow media (GE Healthcare Bio-Sciences, Sweden), a cation exchange resin composed of 6% cross-linked agarose beads and sulphopropyl cation exchange groups. The column, connected to AKTA purifier chromatographic system, was packed and always used with a flow rate of 10 mL/min. The maximum pressure limit set by the chromatographic system was 1 MPa. The column was washed and regenerated after each purification with one column volume of 1M NaOH

2.3.2.3 Sample preparation and elution

IEX was used as a second step purification for HGD and mutant batches. The IEX column was equilibrated before use with 3 column volumes of 0.275 M sodium acetate pH 4.5. Fractions collected from the previous SEC separation step corresponding to the protein of interest were combined and loaded into the IEX column. Unbound proteins were washed with three column volumes of the equilibration buffer. Protein separation was obtained using 0.275 M sodium acetate pH 4.8 + 0.325 M NaCl as linear gradient elution solvent. Table 2.4 and Figure 2.2 show the used method in detail.

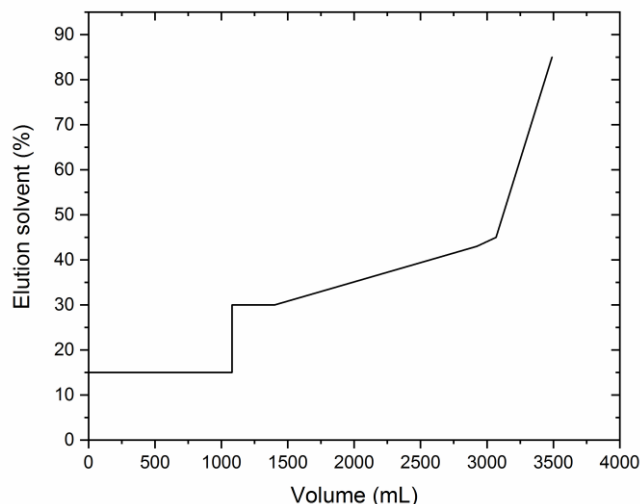


Figure 2.2: IEX method for HGD and mutants purification.

Table 2.4: IEX method for HGD and mutants purification.

Step	Volume (mL)	Elution solvent concentration (%)	Fraction volume (mL)
1	0-1080	15-15	0
2	1080-1401	30-30	0
3	1401-2925	30-43	22
4	2925-3069	43-45	22
5	3069-3489	45-85	0

2.3.3 High performance liquid chromatography

2.3.3.1 Theory

High performance liquid chromatography is a technique used to separate and quantify components of a mixture for analytical or semi-preparative purposes. HPLC provides higher resolution and faster analysis than conventional chromatography and it is often used to assess the purity of protein samples. The basic components of an HPLC system are a pump, that provides high-pressure flow of the solvent used for the analysis, an autosampler and/or injection valve, to introduce the sample, the column, that separates the compounds in the mixture depending on the physical properties of the stationary phase, and a detector, that measures a physical quantity used for the detection and the quantification of eluted compounds.

2.3.3.2 Instrumentation

For protein purity assessments and analysis, a size exclusion column, Superdex 75 5/150 GL (GE Healthcare, Sweden) was connected to a HPLC Shimadzu CBM-20A chromatographic system (SE-HPLC). The stationary phase of the column used is constituted of a mixture of cross-linked agarose and dextrane. The Shimadzu HPLC system used is composed of the following modules: LC-20AT solvent delivery module, DGU-20A₅ online degasser, SIL-20AHT autosampler, SPD-M20A UV/Vis photodiode array detector and CBM-20A communication bus module. Proteins were detected at a wavelength of 280nm.

2.3.3.3 Sample preparation and elution

Protein samples were prepared in 0.1 M sodium phosphate buffer, pH 7 by directly dissolving protein powders (for molecular weights standards) or dialysing with Amicon Ultra-4 centrifugal filters with a 10kDa cut-off membrane (Millipore, Ireland) (HGD and mutants samples). Samples were filtered through 0.22 µm Millex-GV filters into borosilicate glass HPLC vials (Chromacol, Germany) closed with rubber septum caps (ThermoScientific, USA). Protein concentrations between 0.5 and 15 mg/mL were used, depending on the aim of the analysis. HGD samples were supplemented with 20 mM DTT to inhibit covalent dimer formation. The following parameters were used for the analysis:

Table 2.5: HPLC analysis parameters.

Sample Volume	5-20 µL
Flow Rate	0.150 mL/min
Duration	40 min
Pressure limit	15 bar

2.3.3.4 Molecular weight calibration plot

To determine the relationship between the molecular weight and the retention time of HGD and its mutants on a Superdex 75 15/150 column, a calibration plot was determined using protein standards of known molecular weight. Samples were prepared in 0.1 M sodium phosphate pH 7 at concentrations of approximately 0.5-1 mg/mL and filtered through 0.22 µm Millex-GV syringe driven filters prior use. γ -Globulins, Bovine serum albumin (BSA), Chicken serum albumin (CSA), α -chymotrypsinogen A were purchased from Sigma Aldrich (USA) and Lysozyme was purchased from Calbiochem (Canada). Protein retention times and molecular weights are listed in Table 2.6 and the calibration plot in Figure 2.3.

Table 2.6: Retention times and molecular weights of protein standards on Superdex 75 5/150.

Protein	Molecular weight(kDa)	Retention Time (min)
γ -Globulin	150	7.90
Bovine serum Albumin (dimer)	132.9	7.83
Bovine serum Albumin (monomer)	66.4	8.85
Chicken serum Albumin (dimer)	88.6	8.45
Chicken serum Albumin (monomer)	44.3	9.80
α -chymotrypsinogen A	25.6	11.14
Lysozyme	14.4	13.10

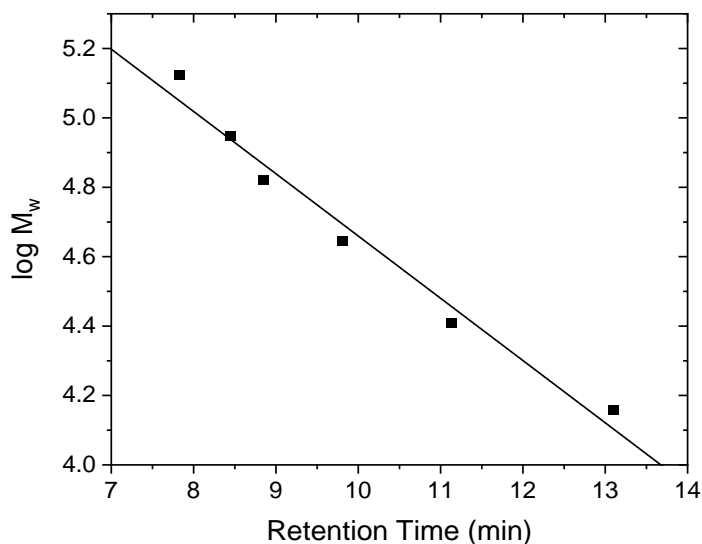


Figure 2.3: Molecular weight calibration plot of Superdex 75 5/150

The linear fit of the points yields the calibration equation:

$$y = 6.454 - 0.179x \quad 2.5$$

where $y = \log_{10}M_w$ and x is the retention time. HGD and its mutants elute at 12.5 minutes which corresponds to a molecular weight of about 16.5 kDa. The difference of about 4 kDa in molecular weight from the true value (20.6 kDa) is due to interactions of the protein with the column resin that makes the size exclusion separation not thermodynamically ideal.

2.3.3.5 Concentration calibration plot

SE-HPLC was used to determine the concentration of HGD samples in the presence of DTT, a reducing agent that contributes to absorption in the UV range. A calibration curve was built accurately determining the concentration of HGD samples without DTT spectrophotometrically and running these samples on the Superdex 75 5/150 size exclusion column. The total area of dimer and monomer was normalized for the injection volume and related to the concentration measured, the calibration plot is shown in Figure 2.4.

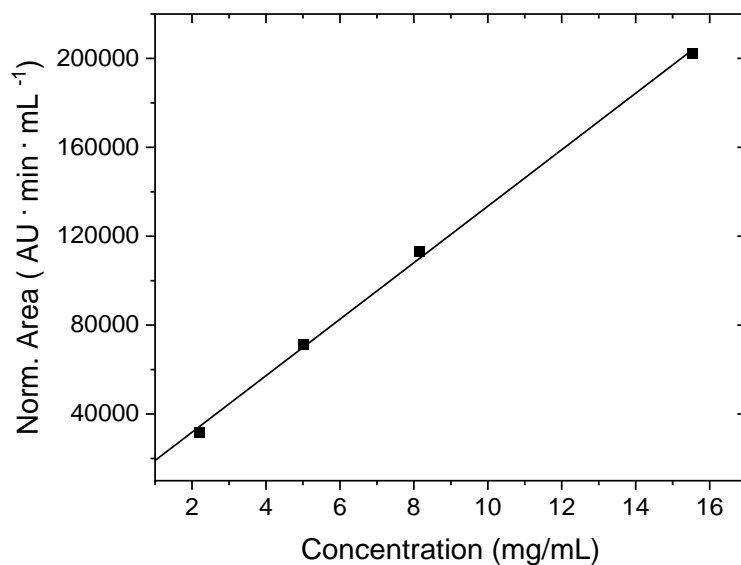


Figure 2.4: Normalized total peaks area as a function of HGD concentration

To determine the concentration of unknown HGD samples containing DTT, an aliquot of the solution was diluted to have a concentration between 1-15 mg/mL and run on the SE-HPLC column. The sum of dimer and monomer areas was divided by the injection volume and related to the concentration through equation 2.6 obtained fitting points in Figure 2.4.

$$y = 6393.44 + 12712.27x \quad 2.6$$

Where y is the normalized area in $\text{AU} \cdot \text{min} \cdot \text{mL}^{-1}$ and x is the concentration in mg/mL .

2.3.4 Polyacrylamide gel electrophoresis (SDS-PAGE)

2.3.4.1 Theory

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is an analytical technique that allows the separation of proteins on a polyacrylamide gel depending on their molecular weight. In native electrophoretic techniques molecules migrate through the medium

due to the application of an electric potential difference and separate depending on the molecular weight to net charge ratio. Sodium dodecyl sulphate binds to protein molecules allowing their denaturation and the formation of a negative net charge regardless of their isoelectric point (pI) and buffer conditions. In the presence of this reagent the electrophoretic mobility is no more influenced by the net charge or the shape of the protein molecule but solely on the molecular weight¹.

2.3.4.2 Preparation and casting of gels

The stacking and running gels were prepared at 4 and 12.6% Acrylamide/Bis (37.5:1) (C₃H₅NO, M_w= 71.08 g/mol) (BioRad, USA) respectively, according to the recipe outlined in Table 2.7:

Table 2.7: Stacking and running gel composition for SDS-PAGE.

Stacking Gel (4%)		Running Gel (12.6%)	
Component	Volume (mL)	Component	Volume (mL)
Milli-Q water	3.05	MilliQ water	3.15
0.5 M Tris buffer pH 6.8	1.25	1 M Tris buffer pH 8.8	2.50
10% (w/v) SDS	0.05	10% (w/v) SDS	0.1
Acrylamide/Bis	0.65	Acrylamide/Bis	4.20

The solutions were degassed for 15 minutes. The polymerization reaction was initiated by adding 10% (w/v) ammonium persulfate (N₂H₈S₂O₈, M_w=228.18 g/mol) (Fisher Scientific, UK) and tetramethylethylenediamine (TEMED, C₆H₁₆N₂, M_w= 116.24 g/mol) (Fisher Scientific, UK) outlined in Table 2.8.

Table 2.8: Polymerization initiators quantities for stacking and running gel of SDS-PAGE.

	Stacking Gel	Running Gel
Component	Volume (μL)	
10% (w/v) ammonium persulfate	50	25
TEMED	5	5

A short and a spacer plate were assembled on a casting stand and secured with a casting frame. Immediately after adding initiator reagents, the running gel solution was poured in between the

two plates. A small amount of 2-methyl-2-butanol ($C_5H_{12}O$, $M_w=88.15$ g/mol) (Fisher Scientific, UK) was used to overlay the running gel solution to remove bubbles. After approximately 40 minutes the gel formed, the alcohol was washed away with Milli-Q water and the remaining water gently removed with a Kim wipe. The stacking gel solution, freshly supplemented with polymerization initiator reagents, was poured on the top of the running gel and combs were soaked in it. The solutions were allowed to settle and gels were stored at 4°C for no more than one week.

2.3.4.3 Sample preparation and gel running

Protein samples were prepared in 0.1 M sodium phosphate pH 7 using Amicon-4 Ultra centrifugal filters to have concentrations between 0.2 and 0.8 mg/mL. 10 μ L of each sample were then mixed with the same amount of the sample buffer (supplemented with 0.05% (v/v) of β -mercaptoethanol when reducing conditions were required), boiled for 5 minutes at 95°C, briefly centrifuged and loaded on to the gel.

2.3.5 Electrospray ionization mass spectrometry (ESI-MS)

2.3.5.1 Theory

Mass spectrometry is an analytical technique that permits the detection of components in a sample depending on their mass to charge ratio (m/z). Several types of mass spectrometers exist and are characterized by different ionization sources, analyzers and detectors. In the case of proteins, one of the most used set ups is composed of an electrospray ionization source (ESI) and two analyzers in tandem, a quadrupole (Q) and a time of flight (TOF). The analyte solution is passed through a narrow needle and nebulized by applying an electric potential difference. The solvent in the droplets is increasingly evaporated until desolvated multicharged ions enter the first analyser. The quadrupole (Q) and time of flight (TOF) analyzers select and direct ions to the detector depending on their m/z ratio resulting in an Intensity versus m/z spectrum.

2.3.5.2 Sample preparation and analysis

Protein samples were prepared at a concentration of 0.5 mg/mL in 0.1 M sodium phosphate buffer pH 7 using Amicon-4 Ultra centrifugal units and filtered through 0.22 μ m Millex-GV syringe driven filters. Samples were analysed in FingerPrints Proteomics Facility, College of Life Sciences, University of Dundee, Scotland, UK with a LC-ESI-Q-TOF-MS instrument.

2.4 Spectroscopy and scattering techniques for protein characterization

2.4.1 UV-Vis spectroscopy

2.4.1.1 Theory

UV-Vis spectroscopy measures the absorption of electromagnetic radiation from a sample in the range of wavelengths between 200 and 780 nm. When certain molecules are irradiated with a monochromatic light of a specific wavelength they are subjected to an electronic transition: the interaction with the photon promotes one electron from a ground to an excited state. This phenomenon arises when the energy of the photon is equal to the energy gap between the ground and excited state of the electron. Scanning the entire range of ultraviolet-visible light allows the promotion of electrons with different potential energy generating an absorption spectrum of the molecule. The absorption of a solution is expressed as the ratio between the intensity of incident light, I_0 , and reduced intensity of light after passing through the sample, I_t . Absorbance is defined as:

$$A = \log \frac{I_0}{I_t} \quad 2.7$$

It is found to have a linear relationship with the concentration of absorbing species according to Beer-Lambert law:

$$A = \epsilon lc \quad 2.8$$

where ϵ is the extinction coefficient of the molecular species, l the pathlength and c the concentration. This can be used to determine the concentration of protein samples since amino acids such as tryptophan (Trp), phenylalanine (Phe) and tyrosine (Tyr) absorb in the far UV range. The absorption spectra of these amino acids are shown in Figure 2.5. The strong absorption bands in the far UV range are due to peptide bond absorption¹⁶⁵.

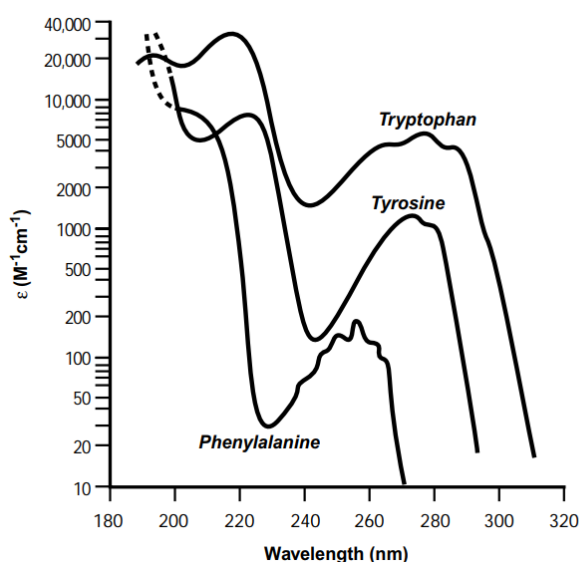


Figure 2.5: Absorption spectra of tryptophan, tyrosine and phenylalanine¹⁶⁵.

The characteristic absorption spectrum of a protein is composed of the contributions of each absorbing amino acid it contains. The microenvironment of tryptophan, tyrosine and phenylalanine residues will influence their wavelength of maximum absorption and so the protein absorption spectrum profile. Second-derivative UV spectroscopy can resolve spectral bands of individual amino acids from protein spectra. This allows changes in the microenvironments of absorbing residues to be monitored and so changes in protein conformation can be detected¹⁶⁶. Hence, second-derivative UV spectroscopy in the range of absorption of aromatic amino acids, can be used to probe protein tertiary structure: the position of the emission maxima undergoes a blue-shift (at lower wavelengths) for higher solvent exposures of the buried aromatic amino acids to the polar solvent and a red-shift (at lower wavelengths) for the opposite phenomenon¹⁶⁷.

2.4.1.2 Sample preparation and analysis

For accurate protein concentration determination, aliquots of the stock samples were diluted from 10 to 1000 times to yield an absorbance value at 280 nm between 0.1 and 1 using a pathlength of 1cm. The same buffer the samples were dissolved in was used as reference. Measurements were repeated at least three times and the average value was recorded. The errors were quantified by calculating the standard deviation and were typically < 2%. For HGD and its mutants, an extinction coefficient of $2.09 \text{ mg}^{-1} \text{ cm}^{-1} \text{ mL}^{97}$ was used to calculate the concentration of the samples. Concentration was occasionally expressed in protein volume fraction ϕ :

$$\phi = c \cdot v \quad 2.9$$

Where c is protein concentration in mg/mL and v is the partial specific protein volume in mL/mg (for HGD and its mutants $v = 0.00071$ mL/mg¹⁶⁸).

The concentration of virus like particle (VLP) samples was measured using as extinction coefficients 1.16, 1.15 and 1.18 mL mg⁻¹ cm⁻¹ for ChADDomer mutants, ADDomer and Chimera respectively.

Second-derivative UV-Vis spectra were obtained from protein solutions of approximately 0.1 mg/mL in 0.1 M sodium phosphate pH 7 in non-reducing conditions. Spectra were recorded between 250 and 350 nm with a wavelength interval of 0.1 nm and a scan rate of 120 nm/min. Derivation and smoothing were performed with Origin 2018 software using Savitzky-Golay filter (polynomial order=2, window size=5). Spectra were either obtained using a Molecular Devices SpectraMax M2e spectrophotometer, a Beckman-Coulter DU800 or a Perkin-Elmer λ 35.

2.4.2 Fluorescence emission spectroscopy

When a molecule interacts with a photon of appropriate energy this results in the promotion of one electron into an excited state; the phenomenon generates an absorption spectrum and is the basis of UV-Vis spectroscopy. The excited electron can partially dissipate its energy as heat (non-radiative relaxation) and then decays back to the ground state releasing a photon of less energy than the initial absorbed one; this generates an emission spectrum and the phenomenon is called fluorescence emission. Fluorescence emission spectroscopy is performed by exciting a sample at the wavelength of its maximum absorption and recording its emission spectrum; the technique is termed as 'intrinsic' if the analyte is naturally fluorescent and 'extrinsic' if a fluorescent probe needs to be added to the sample to study the analyte.

Fluorescence emission spectroscopy is widely used to study biological macromolecules. In a protein, aromatic residues (tryptophans, phenylalanines and tyrosines) generate intrinsic fluorescence emission spectra and the overall profile is found to be sensitive to solvent polarity. Indeed, aromatic amino acids can be more or less buried in the protein structure depending on solution conditions; this results in shifts in the emission spectra making fluorescence emission spectroscopy a powerful tool to study protein tertiary structure¹⁶⁵.

2.4.2.1 Thioflavin T binding assay

Thioflavin T (ThT) is a compound widely used to assess the formation of amyloid fibrils by fluorescence extrinsic emission spectroscopy. When bound to amyloid fibrils, ThT shows a significant increase in fluorescence emission of several orders of magnitude¹⁶⁹.

Amyloid fibrils are formed by a cross- β -sheet structure, i.e. a supersecondary structure where a core of β -strands is perpendicularly aligned to the fibril axis (Figure 2.6A)¹⁷⁰. The cross- β -sheet structure gives rise to a characteristic X-Ray diffraction pattern with reflections at 4.8 and 8-10Å corresponding respectively to the inter-strand and inter-sheet distances. It has been proposed that ThT binds along the surface amino acid side-chains perpendicularly to the β -strands axis (Figure 2.6B)¹⁷⁰. According to the most widely accepted mechanism, the two aromatic rings of ThT (Figure 2.6C) are free to rotate around the central carbon-carbon bond when the molecule is in solution; the rotation quenches the electronic excited states causing little or no fluorescence emission. However, when ThT interacts with fibrils, its rotation is blocked and the excited states are not quenched generating a high intensity of fluorescence emission¹⁷¹.

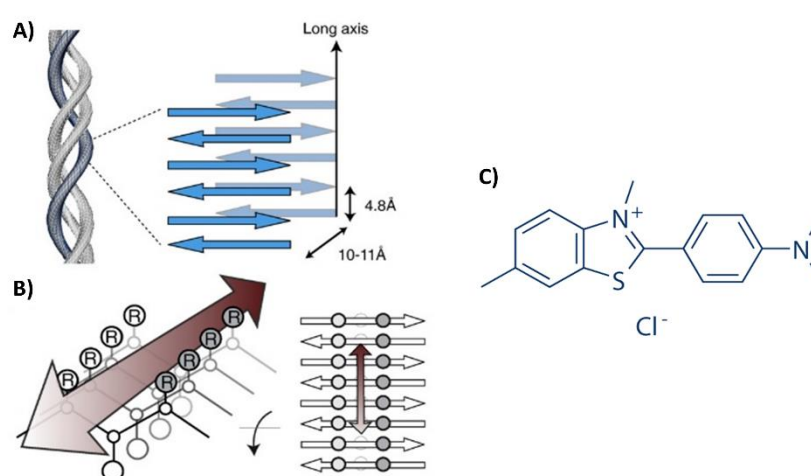


Figure 2.6: A) The cross- β -sheet structure of amyloid fibrils; the inter-sheet and inter-strands distances are 4.8 and 10 Å as measured by X-Ray diffraction; B) orientation of ThT molecule (grey arrow) when binding to the cross- β -sheet structure; C) chemical structure of a ThT molecule. Insets A) and B) are taken from ref. 169.

ThT binding is commonly considered to be specific for amyloid fibrils; however, ThT has been demonstrated to bind also in α -helix regions¹⁷² or hydrophobic cavities in non-fibrillar proteins¹⁷³. The assay is therefore not foolproof¹⁷⁴ and careful controls when studying fibrillation are essential.

2.4.2.2 Sample preparation and analysis

Samples for fluorescence emission spectroscopy were prepared in 0.1 M sodium phosphate pH 7 and non-reducing conditions at a concentration of 0.3 mg/mL for HGD and its mutants and in 50 mM Tris pH 7.5 at concentration of 0.2-0.3 mg/mL for virus-like particles (VLPs). Two different excitation wavelengths were used: 280 nm, to excite all the aromatic residues and 295 nm, to selective excite only tryptophan residues. The emission spectra were recorded between 250 and 450 nm with a wavelength interval of 1 nm with a Molecular Devices SpectraMax M2e

spectrometer after baseline subtraction of the buffer reference. To obtain the wavelength of maximum emission of VLPs samples, first-order derivation, smoothing (Savitzky-Golay filter, polynomial order=2, Points of window=25) and interpolation were applied to the spectra.

ThT binding assay samples were prepared mixing P23VC110M protein solutions and stocks of ThT in 0.1M Sodium phosphate pH 7 giving final concentrations of 30 mg/mL and 10 μ M respectively. ThT concentration was determined by UV-Vis spectroscopy at 412 nm using an extinction coefficient of 36000 M⁻¹ cm⁻¹. The samples were then incubated for few hours at 30-37 °C to grow spherical aggregates. 70 μ L aliquots were pipetted into Nunc™ F96 MicroWell™ Black Polystyrene Fluorescence Plates (ThermoScientific, Ireland). ThT Fluorescence emission of samples and controls were recorded using an excitation and emission wavelengths respectively of 440 and 482 nm in a Molecular Devices SpectraMax M2e microplate reader.

2.4.3 Circular dichroism spectroscopy

2.4.3.1 Theory

Circular dichroism spectroscopy is one the main techniques to study conformations of macromolecules. In a spectropolarimeter, linearly polarized light waves are decomposed into two equal circularly polarized waves, left and right-handed. Optically active molecules, or else molecules missing any roto-reflection axis S_n , absorb the left and right-handed circularly polarized light with different extinction coefficients. After passing through a chiral sample, these two components of light are recomposed giving rise to an elliptical polarized wave. This results in a measurable quantity called ellipticity θ , that can be converted to an extinction coefficient difference between the left handed and right handed circularly polarized light $\Delta\mathcal{E}$ ¹⁷⁵. Conventional circular dichroism spectroscopy is performed in the UV range. This means a sample to be analysed with this technique must absorb in this wavelength range. Circular dichroism spectroscopy is a simple and efficient technique to study structure in protein samples: protein dichroic behaviour arises from their inherent chiral nature (L-amino acids) and from the higher hierarchy chiral structures they adopt (secondary and tertiary structures). Using the peptide bond absorption region (far-UV, 180-240 nm) information on the secondary structure of the protein can be obtained (relative contribution of α -helix, β -sheet and random coil) while the near-UV region (240-320nm), responsible for the absorption of tryptophan, tyrosine and phenylalanine residues, gives information on the tertiary structure of the protein¹⁷⁵.

2.4.3.2 Sample preparation and analysis

Protein samples were prepared at a concentration of 0.7 mg/mL in 0.1 M sodium phosphate buffer pH 7 using Amicon-4 Ultra centrifugal units and filtered through 0.22 μ m Millex-GV syringe driven filters. For VLPs, samples were prepared in 50 Mm Tris pH 7.5 at a concentration of 1

mg/mL. Samples were analysed either at the Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland, UK on a JASCO J-810 spectropolarimeter (C110M and C110S) or at the Bristol Center for Synthetic Biology, University of Bristol, England, UK on JASCO J-810 or 815 spectropolarimeters (HGD, P23VC110M, VLPs). Spectra were normalised by sample concentration, pathlength, total number of residues and, where necessary, smoothed by Savintzky-Golay filter (polynomial order=1; points of window=9 for VLPs or =15 for HGD and P23VC110M). The melting curves for VLPs were smoothed by the same method but using points of window=99.

2.4.4 Dynamic Light Scattering (DLS)

2.4.4.1 Theory

Dynamic light scattering is a technique that measures the diffusion coefficient of particles in solution. The diffusion coefficient can be used to deduce the hydrodynamic radius of the particle and, if measured at different concentrations, intermolecular interactions. Macromolecules in solution are constantly subjected to Brownian motion, thus, considering an infinitesimal volume of the sample, particles will diffuse in and out of this volume. This gives rise to infinitesimal concentration fluctuations that corresponds to fluctuations of the scattered intensity. These variations can be analysed over time using an autocorrelation function:

$$g(\tau) = \frac{\langle I_s(t) \cdot I_s(t + \tau) \rangle}{\langle I_s(t) \rangle^2} \quad 2.10$$

Where $I_s(t)$ and $I_s(t+\tau)$ are the intensity of light scattered at time 0 and after a fixed period of time τ , and the brackets $\langle \rangle$ indicate the averaging over time. When the scattered intensity at time τ is the same as at the beginning of the experiment the autocorrelation function is equal to 1. At longer delay times the autocorrelation function starts to decay exponentially with τ . If the sample is monodisperse equation 2.10 can be rewritten as:

$$g(t) = A + B \exp(-2\Gamma\tau) \quad 2.11$$

Where A and B are two pre-exponential factors linked to the autocorrelation baseline and Γ , the decay rate, can be expressed as follows:

$$\Gamma = D_0 q^2 \quad 2.12$$

Where D_0 is the diffusion coefficient of the particle at infinite dilution and q the scattering vector:

$$q = \frac{4\pi n_0}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad 2.13$$

Equation 2.11 can be rewritten as:

$$\ln g(\tau) = \ln AB - 2D_0 q^2 \tau \quad 2.14$$

The diffusion coefficient is then the slope of the linearized autocorrelation function¹⁷⁵. This quantity can be used to calculate the hydrodynamic radius of a (spherical) particle using Stokes-Einstein equation:

$$R_H = \frac{k_B T}{6\pi\eta D_0} \quad 2.15$$

where η is the viscosity of the solution, T the temperature and k_B the Boltzmann constant. The previous equations consider the diffusion coefficient in the limiting case of infinite dilution, in the absence of intermolecular interactions. In practice this quantity is dependent on the concentration of the molecules and in non-ideal conditions is indicated as collective diffusion coefficient D_c . The two diffusion coefficients are related by the following equation⁸⁹:

$$D_c = D_0(1 + k_D \phi) \quad 2.16$$

Where ϕ is the volume fraction of the macromolecule, and k_D the interaction parameter, analogous to B_{22} that defines protein-protein interactions (Section 1.2.3). Measuring protein samples at different concentrations (volume fractions) yields a set of values for D_c . D_0 can be extrapolated from $D_c(c \rightarrow 0)$ and the interaction parameter can be easily quantified from equation 2.16.

2.4.4.2 Sample preparation and analysis

Samples of HGD and its mutants for Dynamic Light Scattering were prepared and analysed as follows: a stock solution of approximately 35 mg/mL was prepared by dialysing against 0.1M sodium phosphate buffer pH 7. Solutions at lower concentrations were prepared by dilution of the stock. 20 mM DTT was added to HGD solution and incubated overnight. Samples were centrifuged at 13000 RPM for 1 hour and filtered through Whatman Anotop 0.02 μm syringe driven filters (Sigma-Aldrich, UK) transferred into custom made 5mm NMR glass tube (Hilgenberg GmbH). Each sample was equilibrated for 5 minutes and measured at 20 °C for 30 minutes at scattering angle of 90°. Samples concentrations were determined after the analysis by UV spectroscopy for C110M, P23VC110M and C110S or by SE-HPLC for HGD.

VLP samples were prepared in 50mM Tris pH 7.5, centrifuged at 13000 RPM for 30 minutes and filtered twice through 0.22µm syringe driven filters (Sigma-Aldrich, UK). VLP sample concentrations were determined by UV spectroscopy. The filtered samples were transferred into custom made 5 mm NMR glass tubes (Hilgenberg GmbH) and measured over a temperature range between 20 and 55°C for 5 minutes at a scattering angle of 90°. Temperature ramps were performed manually, allowing the sample to equilibrate for 5 minutes before each measurement. Although it was not possible to accurately monitor the heating rate, each experiment was consistently carried out over approximately 3.5 - 4 hours. Hydrodynamic radii and polydispersity indices were obtained by cumulant fit of the correlation function applying a 10% cut-off dust filter and correcting viscosity and refractive index of the solution according to the temperature.

The instrument was composed of a ALV/CGS-3 goniometer, a HeNe laser operating at a wavelength of 632.8 nm, an optical fiber based detector and a ALV/LSE-5004 Light Scattering Electronics and Multiple Tau Digital Correlator. The temperature was kept constant during the course of each measurement using a Thermo Scientific DC30-K20 water bath connected to the instrument and measured with a Pt-100 probe immersed into the index matching fluid vat.

2.4.5 X-ray scattering

2.4.5.1 Theory

X-Ray scattering techniques are classified as small-angle or wide-angle X-Ray scattering (SAXS and WAXS). At small angles (0.1 – 10 °), X-Ray scattering provides information on the 1-100 nm length scale while at wider angles it probes sub-nanometer lengths. SAXS is extensively used to study biological macromolecules allowing information on overall size and shape, flexibility and assembly state to be determined¹⁰. WAXS is sensitive to small structural changes and is used as a complementary technique to X-Ray diffraction and Nuclear magnetic resonance (NMR) to validate protein structural models^{176,177}. More often, WAXS has been used to study the degree of crystallinity of crystalline or semi-crystalline materials¹⁷⁸.

The scattering of X-Rays arises from electron density inhomogeneities within the sample; the intensity of the scattered beam is therefore proportional to the difference in electron density between the solute and the solvent. However, the scattered intensity also decreases as the detection angle increases and it is usually expressed as a function of the scattering vector q :

$$q = \frac{4\pi}{\lambda} \sin\left(\frac{2\theta}{2}\right) \quad 2.17$$

Where λ and 2θ are respectively the wavelength of the incident beam and the detection angle of the scattered one; the scattering profile $I(q)$ vs q is dependent on the molecular shape. If the sample is monodisperse and dilute, i.e. there are no interactions between the solute molecules, the Guinier approximation allows the radius of gyration of the solute molecule, R_g to be determined:

$$I(q) = I_0 \exp\left(\frac{-R_g^2 q^2}{3}\right) \quad 2.18$$

where I_0 is the scattered intensity at $\theta=0$. The radius of gyration is defined as the average squared distance of any point within the object from its centre of mass; it is therefore different from the hydrodynamic radius R_H , which also includes hydration layers around the molecule. The two quantities are roughly related by $R_H=1.3R_g$. Since the Guinier approximation is only valid at small q ($R_g \cdot q < 1.3$), a more accurate and reliable method to obtain R_g is from the pair distance distribution function, $P(r)$. The intensity $I(q)$ and $P(r)$ (respectively in reciprocal and real space) are related by a Fourier transform:

$$P(r) = \frac{r}{2\pi^2} \int_0^\infty q \cdot I(q) \cdot \sin(qr) dq \quad 2.19$$

$P(r)$ is an autocorrelation function of the electron density and represents how often two electrons are at a specific distance from each other within the solute molecule; it can therefore be used to calculate R_g and gives information on molecular shape and flexibility^{10,176,179}.

2.4.5.2 Sample preparation and analysis

P23VC110M samples for SAXS and WAXS were prepared by dialysing protein solutions against 0.1 M sodium phosphate pH 7 with Amicon Ultra-4 centrifugal filters with a 10kDa cut-off membrane (Millipore, Ireland). Samples required to be monodisperse were also centrifuged at 13000 RPM for 15 minutes and filtered through Whatman Anotop 0.02 μm syringe driven filters (Sigma-Aldrich, UK). When spherical aggregates were required, P23VC110M solutions at approximately 30 mg/mL were incubated at 30-37 °C for few hours. Samples were placed in 1mm glass X-Ray capillaries (Capillary Tube Supplies Ltd, UK); spherical aggregates were also delicately pelleted to the bottom of the tube unless otherwise stated. SAXS and WAXS experiments were performed at room temperature on a SAXS Lab Ganesha instrument with a X-Ray wavelength of 1.5408 Å by Dr. Annela Seddon.

SASview software v5.0 was used to fit the experimental data and to perform pair distribution function analysis using the Levenberg-Marquardt fitting algorithm. For the P23VC110M 5

mg/mL sample, a spherical form factor and a sticky hard sphere structure factor were used as a model according to the equations^{180,181}:

$$I(q) = \frac{S}{V} \left[\frac{3V\Delta\rho(\sin(qr) - qr\cos(qr))}{(qr)^3} \right]^2 + B \quad 2.20$$

$$\varepsilon = \frac{1}{12\tau} \exp(u_0/k_B T) \quad 2.21$$

Where S is a scale factor (volume fraction), V is the volume of the scatterer, r is the radius of the sphere, q the scattering vector, $\Delta\rho$ the difference in scattering length density between solute and solvent, τ is the perturbation parameter, ε the stickiness, u_0 is the depth of the potential well, T the temperature and k_B the Boltzmann constant. The following fitting parameters were used: volume fraction 0.00360 (corresponding to 5mg/mL), scattering length density of solute and solvent respectively 12.2 and $9.47 \cdot 10^{-6} \text{ \AA}^{-2}$, polydispersity was included in the model using a Schulz distribution. For the $P(r)$ analysis, the parameters used were: number of terms 11, regularization constant $1.7 \cdot 10^{10}$, maximum distance 102 \AA . SAXS profiles arising from the aggregation experiment over time of a P23VC110M 50 mg/mL solution were fitted by Dr. Annela Seddon using a spherical form factor and the Ornstein-Zernike model according to equation 2.22¹⁸² fixing the minimum radius of gyration to 1.7 nm.

$$I(q) = \frac{S}{1 + (qL)^2} + B \quad 2.22$$

Where S is a scale factor, B the background, q the scattering vector and L the correlation length.

2.5 Microscopy

2.5.1 Cross-polarised light microscopy

Cross-polarised light microscopy allows objects with an isotropic internal structure, such as amorphous solids, and objects with an ordered but anisotropic internal structure, i.e. crystals to be distinguished from each other. In a cross-polarization experiment, an anisotropic crystalline object subdivides and refracts the incident light into two perpendicularly vibrating waves depending on its orientation (birefringence). The specimen is placed between a polariser and an analyser set perpendicular to each other; these are optical filters that let light waves with only a specific polarisation through. Amorphous (isotropic) materials refract light with the same polarisation as the polarised incident beam; this is totally blocked by the analyser (perpendicular to the oscillating direction of the incident beam) generating a black image. On the other hand,

crystalline solids refract the incident light with a different polarization hence appearing bright over a black background (Figure 2.7)¹⁸³.

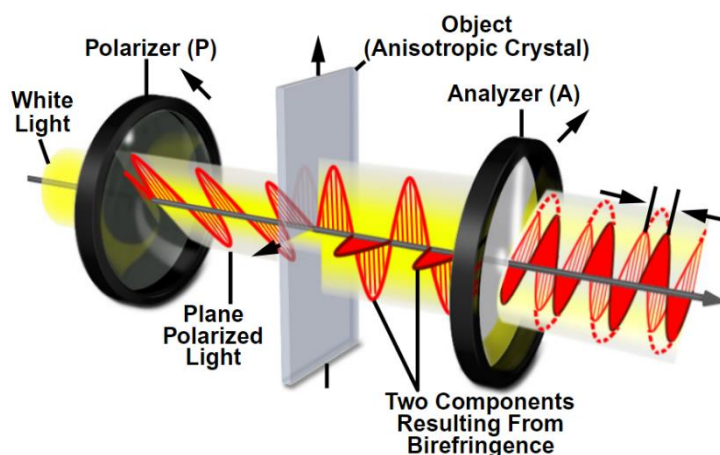


Figure 2.7: Set up of a cross-polarization light microscopy experiment. The image is adapted from ref. 183 (MicroscopyU.com).

2.5.2 Sample preparation and analysis

Protein crystals and spherical aggregates were grown in either NMR tubes or perfusion chambers (Grace Bio-labs) attached between two standard microscope slides. Images were captured by an Olympus BX61 microscope equipped with 10, 20, 60 or 100X magnification oil-immersion lens and Cell^F software. Cross-polarization experiments were performed inserting the polariser (direction NE-SW) and analyser between the specimen stage; the analyser was rotated until its direction was perpendicular to the polariser (NW-SE), the sample placed on the specimen stand and observed by the eye pieces.

2.6 Protein crystallization and solubility measurements

2.6.1 Protein crystals/assemblies growth

HGD, C110M, C110S and P23VC110M protein solutions at concentrations between 110 and 180 mg/mL were prepared by dialysing against 0.1M sodium phosphate buffer, pH 7. They were placed in custom made 5 mm NMR glass tubes (Hilgenberg GmbH) and sealed with Teflon tape to prevent evaporation. To induce protein crystallization, the glass tubes were cooled in an ice/acetone bath to induce liquid-liquid phase separation (LLPS). The temperature of the bath was maintained between -5 and -10 °C for 2-5 hours. The solutions were then stored at 4°C overnight. P23VC110M aggregates were growth from protein solutions of about 30 mg/mL in 0.1 M sodium phosphate, pH 7. To yield visible aggregates, the solutions were placed in NMR tubes and incubated overnight at 37 °C for a few hours.

2.6.2 Protein solubility curve measurements

Protein crystals/aggregates were allowed to settle to the bottom of the NMR glass tube. The mother liquor was carefully removed and replaced with fresh 0.1 M sodium phosphate buffer, pH 7. The NMR tube was then incubated at a specific temperature in either a Innova 42 Incubator, a Thermo Scientific MaxQ4000 refrigerated incubator (Thermo Scientific, UK) or a polystyrene box with an internal water coil connected to a Thermo Scientific DC30-K20 water bath (Thermo Scientific, UK). Constant mixing of crystals/aggregates was performed by fixing the glass tube to a small rotating motor. The temperature was checked and registered using a Kane May KM340 differential temperature thermometer with a K type thermocouple. After some hours, the crystals/aggregates were allowed to settle and a small aliquot of the supernatant was removed to measure protein concentration. To ensure thermodynamic equilibrium between the two phases was reached, the concentration was monitored over a period of 24-48 hours and, only when the concentration reached saturation, the concentration value was used as a point of the solubility curve. The tube was then replenished with fresh buffer if needed, the temperature was changed and the procedure repeated.

2.7 Protein crystallography

2.7.1 Theory

X-Ray diffraction is one of the most widely used methods to determine the three-dimensional structure of macromolecules with atomic resolution. The physical phenomenon on the basis of this technique is the diffraction of X-Rays from crystalline material. Electrons in an atom scatter incident X-Ray waves and the constructive and destructive interference between them generates the phenomenon called diffraction. The diffracted waves arising from the atom can be represented by a vector, the scattering atomic vector. The same process repeats in larger scale within a molecule: the scattered waves from atoms sum up generating constructive and destructive interference. Similarly, the structure factor is the sum of all the atomic scattering vector of a lattice plane, and it is composed of an amplitude and a phase. A diffraction experiment yields a diffraction pattern of spots with different intensities. Each spot corresponds to the structure factor of a lattice plane but, while the amplitude (intensity) information is retained, the phase information is lost. The geometry of the diffraction pattern allows to obtain size and shape of the unit cell; however to deduce the distribution of atoms in the lattice the phase information is needed¹⁸⁴. This is known as 'the phase problem'. The electron density of the unit cell is related to the summation of all the structure factors of the unit cell from a Fourier transform as described in equation 2.23¹⁸⁵.

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp[-2\pi i(hx + ky + lz) + i\varphi_{hkl}] \quad 2.23$$

Where $\rho(xyz)$ is the electron density, V is the volume of the unit cell, h, k, l is a set of Miller indices that identifies a single plane, $|F_{hkl}|$ and φ_{hkl} are respectively the intensity and the phase of the structure factor of that plane. Knowing the phases and intensities and applying a Fourier transform will allow the electron density and so the crystal structure of the protein to be determined. The phase problem can be solved in different ways depending on the system under study. The simplest method, molecular replacement, uses an already known structure of a similar molecule, and it is widely used to solve mutant protein structures starting from the wild-type coordinates. This method builds an initial model performing rotations and translations to appropriately place the model protein in the unit cell of the target protein. Refinements rounds are then run with the aim of minimizing the difference between the observed intensities of the target protein and the intensities calculated from the model protein¹⁸⁵.

2.7.2 X-ray diffraction data collection and analysis

Data from a single crystal of C110M mutant protein were obtained at Proxima 2 beamline at the Soleil Synchrotron, Saclay, France. Initial data were scaled and merged using *pointless* and *aimless* programs from the CCP4 suit. The X-ray crystal structure of HGD (PDB entry 1hk0) was used to solve the structure of the mutant C110M by molecular replacement using *Phenix*¹⁸⁶. The initial structure was then refined through several cycles of *Phenix* refinement and manual model building using *Coot*¹⁸⁷. Crystal contacts and interfaces were analysed using *PISA*¹⁸⁸ and *Coot*.

Chapter 3:

Surface exposed free cysteine suppresses crystallization of human γ D-crystallin*

* The content of this chapter has been published in *Journal of Molecular Biology* (Strofaldi A., Khan A.R., McManus J.J. *Surface Exposed Free Cysteine Suppresses Crystallization of Human γ D-Crystallin*. *J. Mol. Biol.* 2021; **433(22)**:167252. doi: 10.1016/j.jmb.2021.167252).

3.1 Background

Human γ D-crystallin (HGD) is a perfect model protein to study how anisotropic protein-protein interactions are affected by mutagenesis: HGD evolved to be highly stable in the eye lens over the entire life span and it is therefore recalcitrant to crystallization. However, over the last decades several single-point mutations of this protein have been thoroughly investigated due to their involvement in juvenile cataract onset. Many examples of pathogenic mutations in HGD show high propensity to crystallize causing cataract in young patients. A list of cataractogenic mutations for HGD is detailed in ref. 189 and include P23T, R14S, R36S, R58H^{56,57}.

Each single-point mutation confers specific behaviours to the protein; for instance, mutants R58H and R36S crystallize promptly⁵⁷ and P23T forms aggregates with an inverted temperature dependence of the solubility curve (i.e. they melt upon cooling)⁵⁶. In many cases, mutagenesis does not influence either net protein-protein interactions in the liquid phase or protein structure; for this reason, it has been inferred that anisotropic protein interactions due to small surface modifications are only engaged in the dense liquid or solid phase⁹⁶. Therefore, protein phase diagrams have been used to probe the effect of surface mutations on protein self-assembly behaviour.

In this context, one specific amino acid of HGD requires particular attention: Cysteine-110 (Cys). This is one of the six cysteine residues present in the molecule, but the only one surface exposed and hence able to form disulphide bridges. This amino acid is the only one known to lead to covalent dimerization of the protein, a phenomenon that contributes to triggering protein aggregation in the human eye lens causing age-related cataract^{190,191,192}. This raises the question: does Cys-110 provide any biological advantage? Previous studies suggested that cysteines in crystallin proteins may serve as free-radical scavengers^{193,194} or may be involved in UV-light protective processes^{195,196}, but in HGD, only Cys-18 seems to have this role¹⁹⁷. Also, it has been proposed that the high cysteine content in this class of proteins strongly contributes to the high refractive index increment (dn/dc) needed from eye-proteins (compared to low dn/dc residues such as Leucine or Serine); however, this does not fully explain the surface-exposure of Cys-110 as dn/dc values do not depend on amino acid positions in the chain^{198,199}. On the other hand, it has been shown that chemically linking hydrophobic or hydrophilic moieties on Cys-110 increases the net protein attraction more than other positions on the protein surface, such as Lysine-2 (Lys-2)^{120,121}. This suggests Cys-110 may have an important role in maintaining the solubility of the whole protein.

3.2 Aim of the study

The phase behaviour of HGD has been shown to be substantially affected by single-point mutations on its surface. To study how the amino acid at the 110th position of HGD affects anisotropic interactions, we chose to replace cysteine with its most similar amino acids, Methionine (Met) and Serine (Ser), creating the mutants C110M and C110S respectively (Figure 3.1). These mutations are not known to be pathogenic, but studying their self-assembly could shine more light on the biological role of Cys-110.

Since either Met and Ser are non-oxidizable residues, both mutants should lose the ability to form covalent dimers making the protein less prone to aggregation; however, since HGD is already a hugely stable protein and it has been shown to be sensitive to mutagenesis, the effects of those mutations on protein phase behaviour may be dramatic and hence should be assessed.

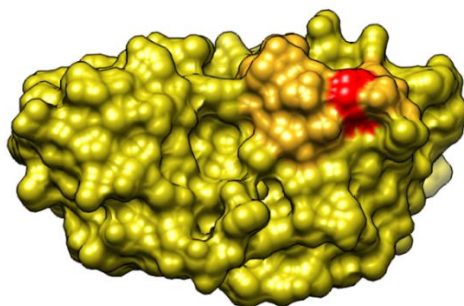


Figure 3.1: Surface model of HGD. Residue 110th is shown in red.

3.3 Results

3.3.1 Production and characterization of Human γ D-crystallin

Human γ D-crystallin (HGD) was expressed in *E. Coli* (Section 2.2.5) and initially purified by Size-exclusion chromatography (SEC) (Figure 3.2 A).

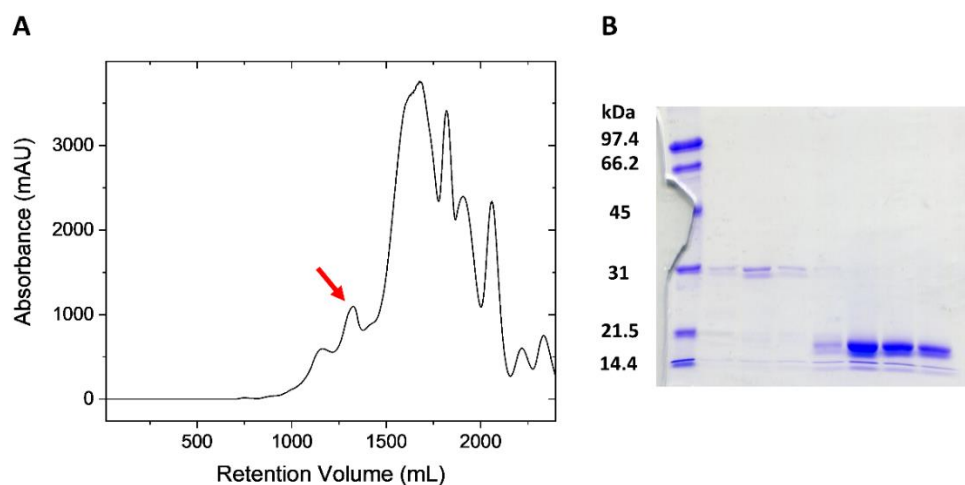


Figure 3.2: A) Size-exclusion chromatogram of HGD purification. B) SDS-PAGE gel of the fractions from SEC step. The red arrow in A) indicates the appropriate pooled fractions as indicated by SDS-PAGE analysis.

The collected fractions around the indicated peak of the chromatogram were analysed by SDS-PAGE (Figure 3.2 B); since the molecular weight of HGD is approximately 20.6 kDa, fractions showing a band corresponding to the 21.5kDa marker were pooled.

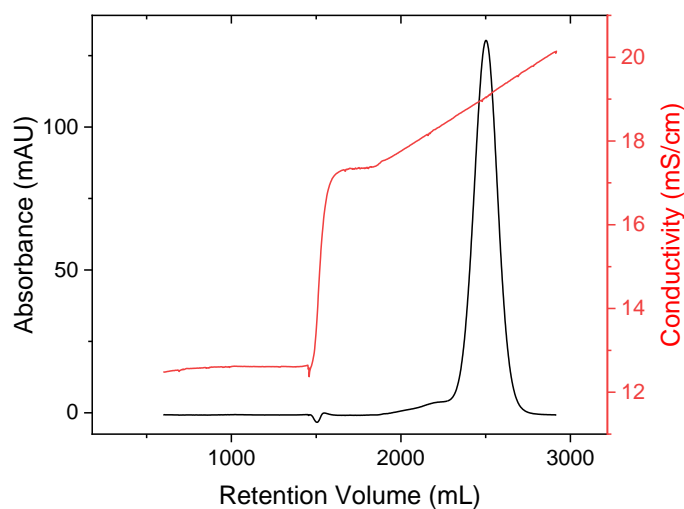


Figure 3.3: Ion-exchange chromatogram obtained for the pooled fractions from SEC step of HGD purification.

Further purification was carried out by Ion exchange chromatography (IEX) (Figure 3.3); the whole purification process led to approximately 100 mg of protein per 7.2 L of bacterial cell culture. Each protein batch was characterised by SDS-PAGE and SE-HPLC confirming purity was $\geq 98\%$ (Figure 3.4: Purity assessment of the purified HGD batch performed by A) SE-HPLC in the presence (red line) or absence (black line) of 20 mM DTT and B) by SDS-PAGE in reducing conditions. A and B). Since at pH 7 HGD forms covalent dimers spontaneously, SE-HPLC was

performed either in the presence or absence of a reducing agent, dithiothreitol (DTT); DTT completely reduces the dimeric protein giving rise to a single monomeric peak (Figure 3.4A)

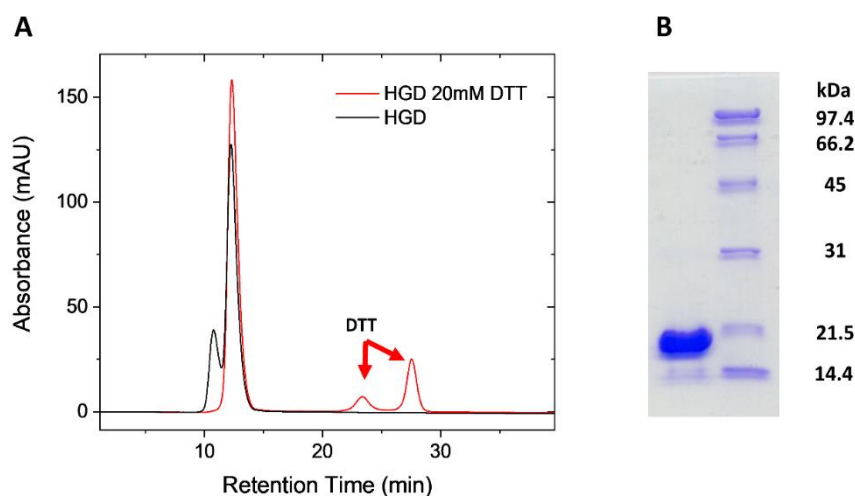


Figure 3.4: Purity assessment of the purified HGD batch performed by A) SE-HPLC in the presence (red line) or absence (black line) of 20 mM DTT and B) by SDS-PAGE in reducing conditions.

Ratios of $A_{260\text{nm}}/A_{280\text{nm}} < 0.5$ measured by UV-Vis spectroscopy indicated protein batches were free of any residual DNA fragments. The identity of the protein was confirmed by Electrospray Ionization Mass Spectrometry (ESI-MS) with a molecular weight of 20605 ± 1 Da in good agreement with previously obtained results²⁰⁰.

3.3.2 Production and characterization of the mutant C110M

The C110M mutant plasmid was modified by site-directed mutagenesis of the native HGD plasmid and successfully purified (Sections 2.2.1 and 2.2.3). DNA purity was confirmed by UV-Vis spectroscopy yielding the following results:

$A_{260\text{ nm}}/A_{280\text{ nm}}$	2.03
$A_{260\text{ nm}}/A_{230\text{ nm}}$	2.3
DNA concentration	41.3 ng/ μL

DNA sequencing with T7 promoter confirmed that the desired mutation occurred by the replacement of codon TGT with ATG. The plasmid was then transformed in *E. Coli* and then purified by size-exclusion (Figure 3.5A) and ion-exchange chromatography (Figure 3.6).

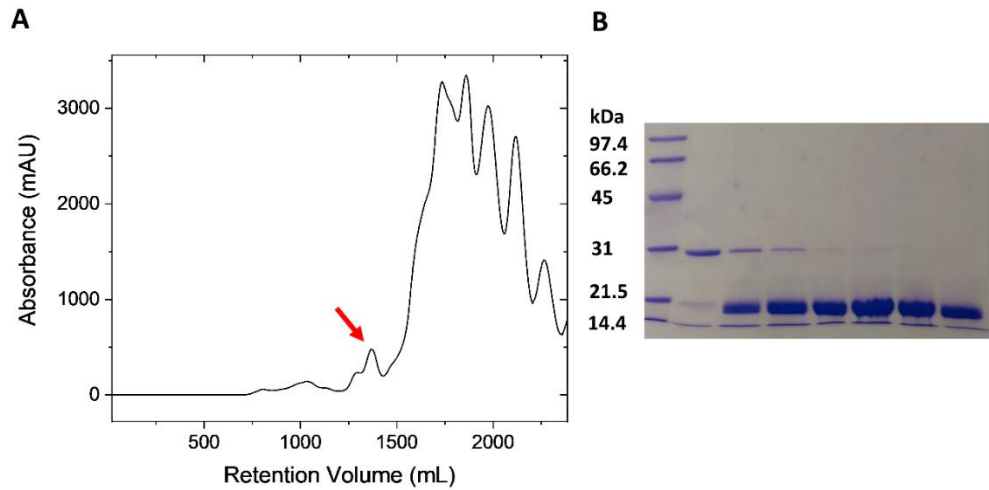


Figure 3.5: A) Size-exclusion chromatogram of C110M and B) SDS-PAGE of fractions obtained by the size-exclusion step. The red arrow in A) indicates the pooled fractions as indicated by SDS-PAGE gel results.

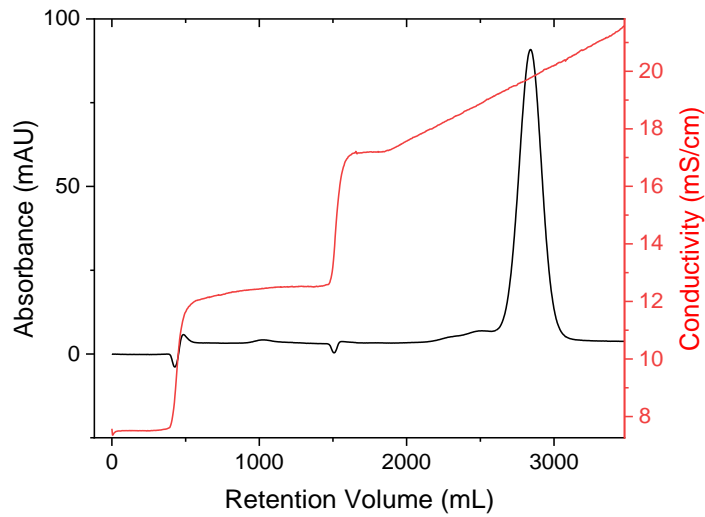


Figure 3.6: Ion-exchange chromatogram for C110M purification.

After the IEX step, fractions were pooled and the purity of the batch was assessed to be $\geq 98\%$ by SE-HPLC and SDS-PAGE (Figure 3.7). Typical purifications yielded approximately 120 mg per 7.2 L of bacterial culture.

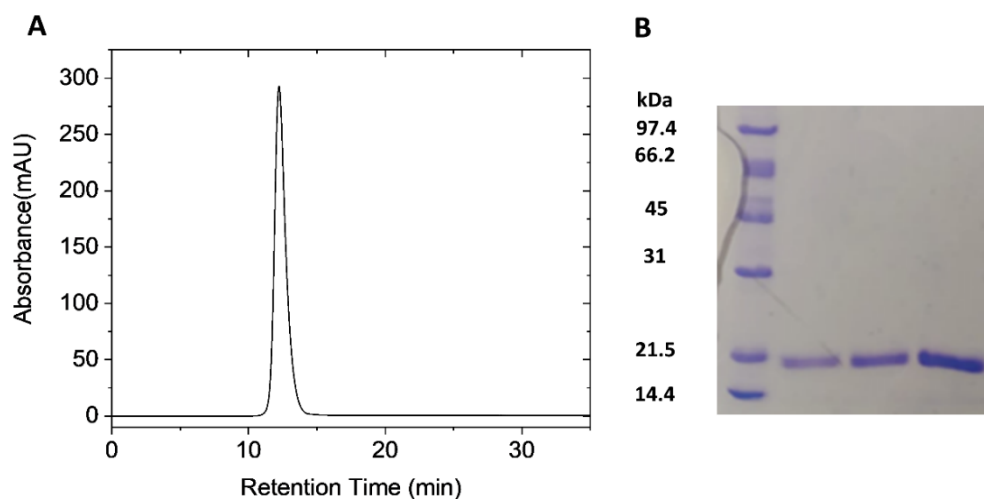


Figure 3.7: Purity assessment of C110M by A) SE-HPLC and B) SDS-PAGE.

SE-HPLC also confirms C110M does not form covalent dimers, as the only surface-exposed cysteine residue has been replaced by a non-oxidizable amino acid (methionine). Protein batches were confirmed to be free from any residual DNA fragment by UV-Vis spectroscopy yielding $A_{260\text{nm}}/A_{280\text{nm}} < 0.5$. Finally, the mass of the mutant C110M was confirmed to be 20634 ± 1 by ESI-MS.

3.3.3 Production and characterization of the mutant C110S

The C110S plasmid was produced as previously described starting from HGD plasmid. DNA purity was confirmed by UV-Vis spectroscopy leading to the following results:

$A_{260\text{ nm}}/A_{280\text{ nm}}$	1.87
$A_{260\text{ nm}}/A_{230\text{ nm}}$	2.4
DNA concentration	57 ng/ μL

DNA sequencing with T7 promoter confirmed the appropriate mutation occurred showing the original codon TGT replaced with AGT. After plasmid transformation in *E. Coli* and protein expression, purification was carried out by size-exclusion (Figure 3.8) and ion-exchange chromatography (Figure 3.9) as previously described.

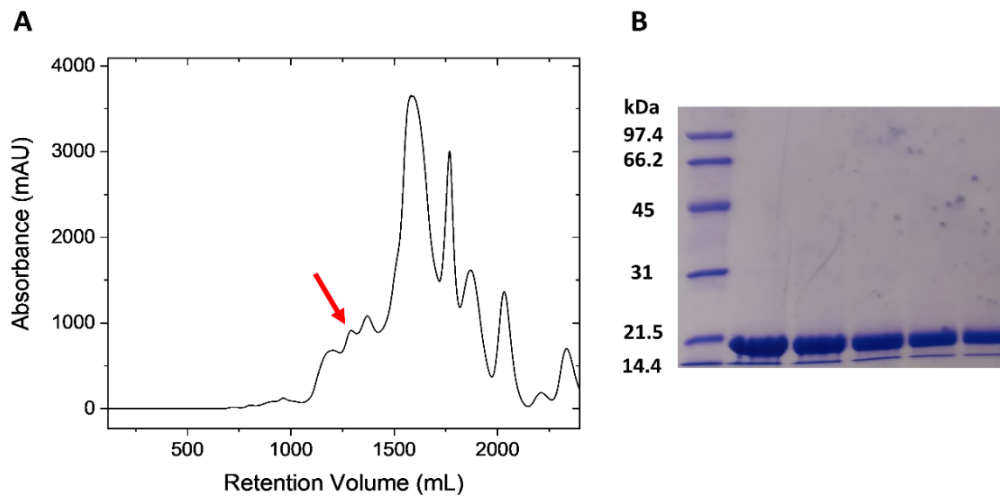


Figure 3.8: A) Size-exclusion chromatogram of C110S mutant and B) SDS-PAGE of the resultant fractions. The red arrow in A) indicates the fractions pooled after identification by SDS-PAGE gel.

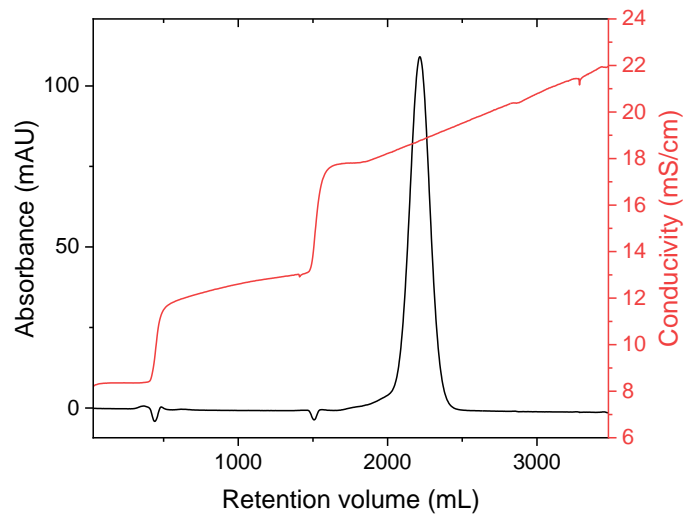


Figure 3.9: Ion-exchange chromatogram for C110S mutant.

C110S purifications yielded typically 100 mg of protein from a 7.2 L of bacterial culture. Purity was confirmed to be $\geq 98\%$ by SE-HPLC and SDS-PAGE (Figure 3.10)

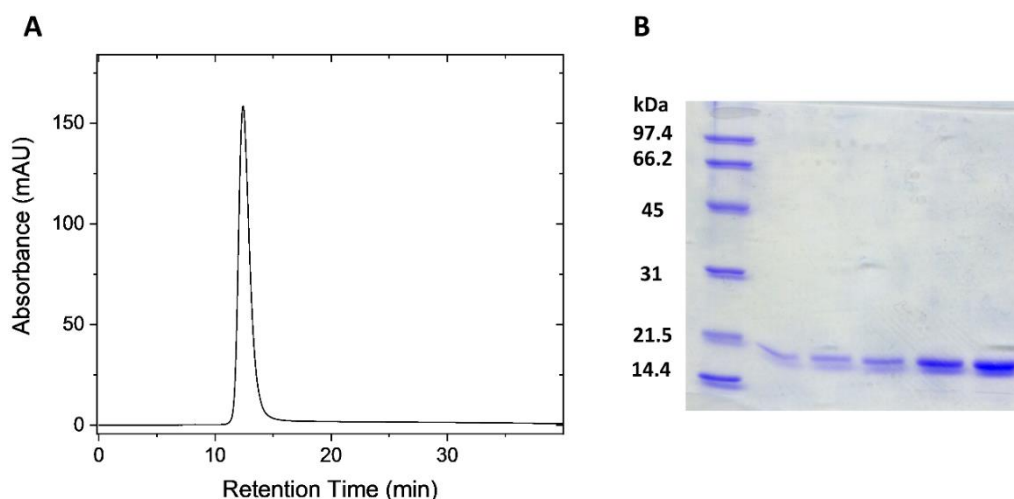


Figure 3.10: Purity assessment of C110S batch performed by A) SE-HPLC and B) SDS-PAGE.

SE-HPLC shows the C110S mutant does not form covalent dimers as a result of the Cys to Ser replacement. Purity of the DNA fragments was confirmed by UV-Vis spectroscopy yielding $A_{260\text{nm}}/A_{280\text{nm}} < 0.5$. The molecular weight of the mutant protein was confirmed to be $20588 \pm 1\text{Da}$ by ESI-MS.

3.3.4 Structural characterization of HGD, C110S and C110M

Protein structure is dictated by amino acids side-chains and it is stabilized by the hydrophobic effect¹⁷⁵. The ability of aromatic residues to absorb in the UV region can be used to probe their solvent exposure and hence in protein structure; this can be achieved by second derivative UV (Section 2.4.1) and Fluorescence emission (Section 2.4.2) spectroscopies. Fluorescence spectroscopy was carried out with excitation wavelengths of 280 nm, to excite all types of aromatic residues, and 295 nm, to selectively excite tryptophan residues¹⁶⁵; fluorescence emission spectra, normalized by the value of maximum intensity, and second-derivative UV absorbance spectra for HGD, C110M and C110S are shown respectively in Figure 3.11 and Figure 3.12.

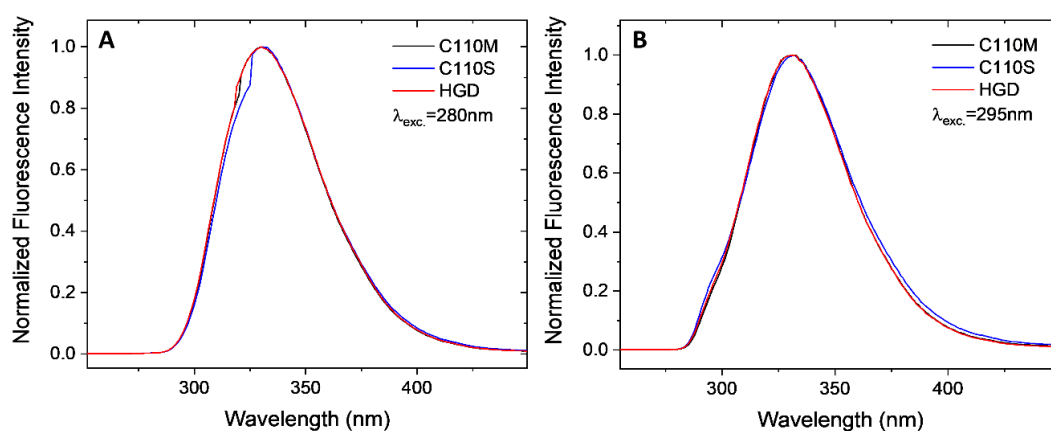


Figure 3.11: Fluorescence emission spectra of HGD and its mutants C110S and C110M obtained using an excitation wavelength of A) 280 and B) 295 nm at a protein concentration of 0.3 mg/mL.

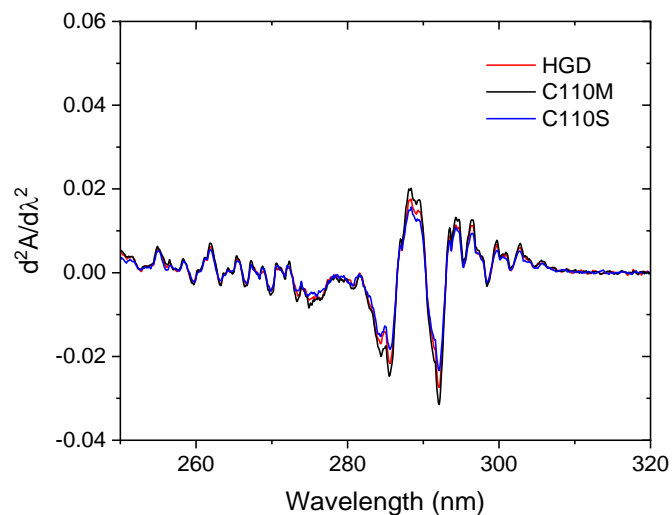


Figure 3.12: Second-derivative UV spectra of HGD and its mutants C110M and C110S.

Both techniques show no relevant blue or red-shift indicating none of the mutations alter protein tertiary structure.

Protein structure can be also assessed by a technique able to probe the absorption of peptide bond (180-240 nm) such as far-UV Circular Dichroism (Section 2.4.3). Far-UV CD spectra of the mutants C110S and C110M are compared with the one of HGD from ref. 57.

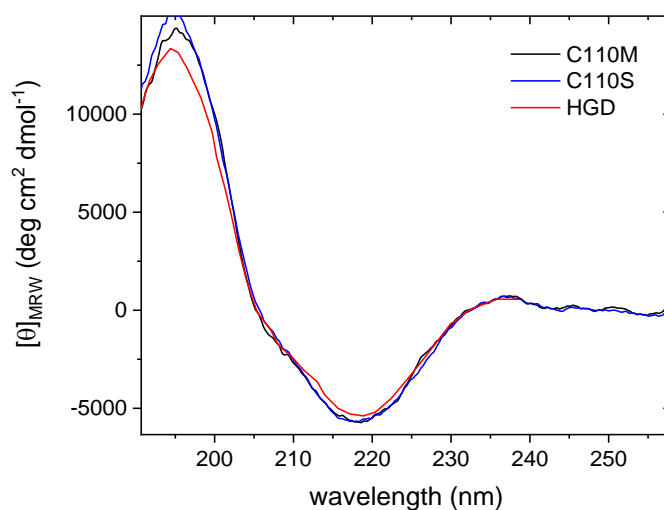


Figure 3.13: Circular Dichroism spectra in the far-UV region of HGD (from ref. 57), C110M and C110S. Spectra were recorded from samples at a concentration of 0.7 mg/mL.

Also in this case, the spectra of the three proteins are nearly identical suggesting no conformational change occurred due to mutagenesis.

3.3.5 Protein-protein interactions in the liquid phase of HGD, C110M and C110S

The strength of protein-protein interactions in solution can be assessed by the interaction parameter k_D obtained by Dynamic Light Scattering (DLS) (Section 1.2.3 and 2.4.4.1). This technique allows the protein diffusion coefficient at different concentrations (collective diffusion coefficient, D_C) to be measured. The diffusion coefficient of the ideal solution at a protein concentration equal to zero, D_0 , can be obtained by extrapolation, while the interaction parameter k_D can be obtained from the slope of D_C/D_0 versus volume fraction (or concentration) (Figure 3.14) according to equation 2.16 (Section 2.4.4.1). Also the hydrodynamic radius R_H can be calculated from D_0 values using Stokes-Einstein equation (2.4.4.1).

$$D_C = D_0(1 + k_D \phi) \quad 2.16$$

$$R_H = \frac{k_B T}{6\pi\eta D_0} \quad 2.15$$

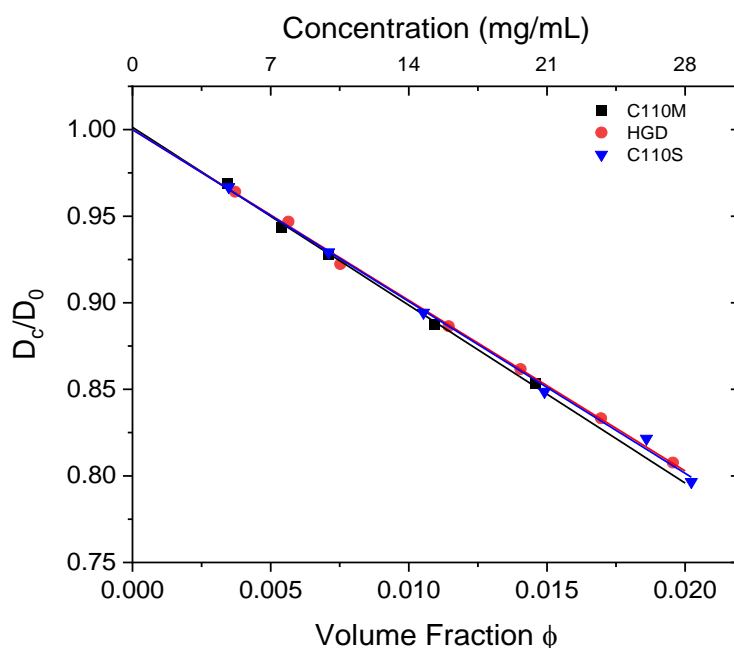


Figure 3.14: Viscosity corrected diffusivity for HGD, C110S and C110M. The slope indicates a negative value for k_D and therefore, net-attractive protein-protein interactions.

Table 3.1: k_D , D_0 and R_H values (with corresponding errors) for the proteins HGD, C110M and C110S as obtained from DLS experiments.

Protein	k_D	D_0 ($\times 10^{-7}$) cm^2/s	R_H (nm)
HGD	-9.86 ± 0.35	9.31 ± 0.02	2.30
C110M	-10.27 ± 0.28	8.97 ± 0.06	2.39 ± 0.02
C110S	-9.92 ± 0.26	8.81 ± 0.03	2.43 ± 0.01

All the samples used in the analysis had polydispersity indices (PDI) less than 0.1 indicating that protein samples were monomeric and free of oligomers/aggregates. k_D and R_H values are in good agreement with those for HGD present in the literature²⁰¹. k_D , D_0 and R_H for all proteins are compared in Table 3.1 and show no significant differences; usually, this type of experiment allows protein-protein interactions to be quantified with an estimated uncertainty in k_D of ± 1 . Hence, in our case, the mutations cause no significant changes in net protein-protein interactions and hence in the chemical potential of the liquid phase μ_L .

3.3.6 Crystallization of HGD, C110M and C110S

HGD crystals have been obtained in the past by inducing liquid-liquid phase separation (LLPS)^{201,202}. However, repeated attempts to crystallize HGD by LLPS within the McManus group were unsuccessful and it was only crystallized using the precipitant (RS)-2-methyl-2,4-pentanediol (MPD)²⁰³.

Here, we attempted to crystallize HGD in batch from 140-160 mg/mL solutions inducing LLPS for 3-5 hours at approximately -8 °C and then incubating the samples at 4 °C for several days. The crystallization buffer and method used were the same as those used for crystallizing HGD in ref. 57, from which thermodynamic data in Figure 3.17 and Table 3.2 are also taken. Attempts to crystallize the mutant C110S were made with the same LLPS method described above but inducing LLPS for up to two weeks in 150-300 mg/mL solutions. None of them ever produced any type of crystals suggesting the nucleation behaviour of HGD and C110S may be similar. On the other hand, C110M readily crystallizes if subjected to the same crystallization method described above. Specifically, C110M crystals were obtained from 140-160 mg/mL samples after inducing LLPS at approximately -8 °C for about 3 hours and incubating overnight at 4 °C (Figure 3.15A). C110M crystals grew over 3 days of incubation leading to very large, diffraction-quality crystals (Figure 3.15B)²⁰⁴.

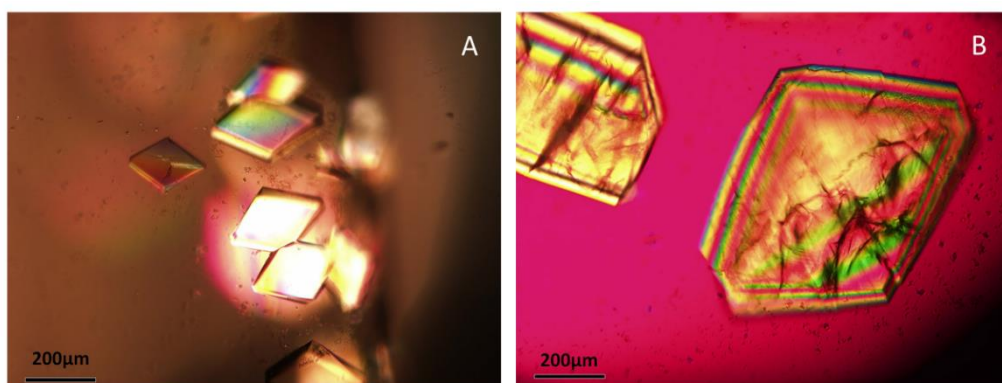


Figure 3.15: C110M crystals after A) one day and B) three days of incubation at 4 °C.

Those observations suggest that C110M has high propensity to crystallize compared to the wild-type protein while apparently maintaining the same rhombic crystalline form. Since C110S did not crystallize, we can exclude the possibility that HGD crystal nucleation is hindered by undetectable or transient amounts of covalent dimer that may form even in the presence of DTT used as a reducing agent.

3.3.7 Solubility curve of C110M crystals and thermodynamic analysis of crystallization

As shown by the DLS results (Section 3.3.5), the mutation C110M does not change net protein-protein interactions in the solution phase; this behaviour is common among HGD mutants. However, even single-point mutations in HGD can influence the position of the solid-liquid equilibrium boundary of crystal/aggregates as shown previously^{56,57}. It has been suggested that while anisotropic interactions due to mutagenesis are averaged out in the liquid phase, they are instead fully engaged in the solid phase (or dense liquid phase)⁹⁶. We therefore measured the solubility curve of C110M crystals and compared it with the phase diagram of HGD present in the literature⁵⁷ (Figure 3.16):

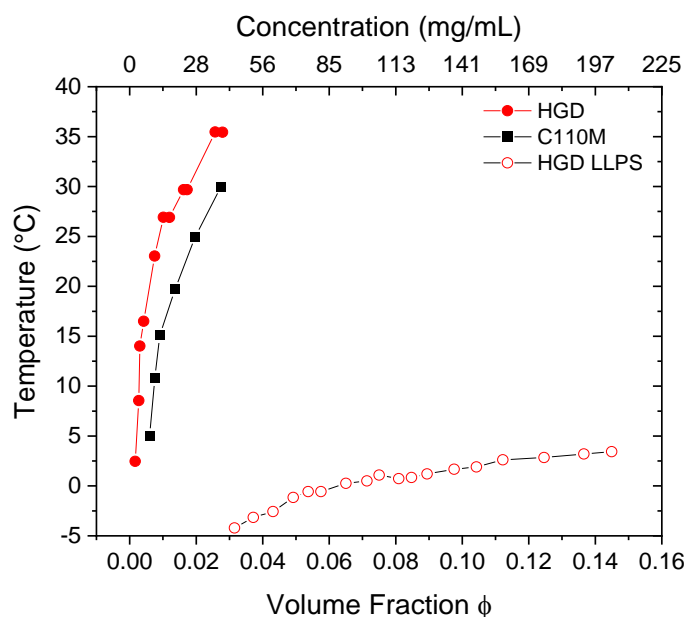


Figure 3.16: Equilibrium solubility curve of C110M crystals compared with the one of HGD from ref. 57. The LLPS boundary of HGD⁵⁷ (empty symbols) is also included for reference.

As shown by Figure 3.16, the solubility curve captures differences in anisotropic interactions due to site-directed mutagenesis. To understand the origins of this shift, we obtained the thermodynamic parameters of crystallization from Van't Hoff equation⁵⁷:

$$\ln\phi = \frac{\Delta G_C}{RT} = \frac{\Delta H_C}{RT} - \frac{\Delta S_C}{R} \quad 3.1$$

Where ΔG_C , ΔH_C and ΔS_C are respectively the Gibbs free energy, enthalpy and entropy of crystallization, ϕ is the protein volume fraction, R the gas constant, T the temperature and T_0 the reference temperature chosen as 298 K. The fits to the solubility data are shown in Figure 3.17 and the results reported in Table 3.2

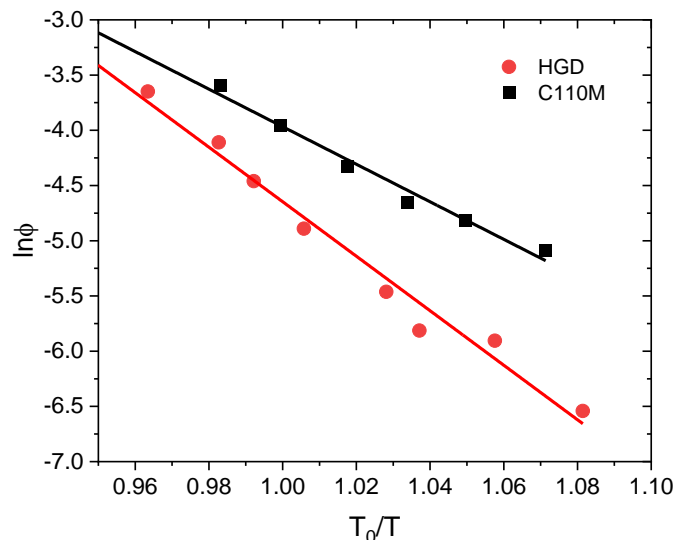


Figure 3.17: Temperature dependence of HGD and C110M crystals solubility. T_0 is the reference temperature taken as 298 K. Data for HGD are taken from ref. 57.

Table 3.2: Thermodynamic parameters for HGD and C110M crystallization. T_0 is the reference temperature taken as 298 K. For a more accurate comparison, data for HGD are taken from Table 1 of ref. 57 and not from the fitting shown in Figure 3.17.

	$\Delta G_C(T_0)$ (KJ/mol)	ΔH_C (KJ/mol)	ΔS_C (J/mol·K)	$-T_0\Delta S_C$ (KJ/mol)
HGD	-11.7	-61.5	-167.6	49.8
C110M	-9.8	-42.1	-108.4	32.3

From DLS results, we know that the diffusivities of the two proteins are the same, hence we can assume $\mu_L^{C110M} = \mu_L^{HGD}$. The difference in ΔG_C between the mutants can then be reduced to the difference in the Gibbs free energy of the two crystalline phases. This corresponds to $G_C^{C110M} - G_C^{HGD} = 1.9 \text{ KJ/mol}$ or $\mu_c^{C110M} - \mu_c^{HGD} = 0.77 k_B T_0$ at room temperature. Hence, C110M crystals are slightly less stable than those of HGD, having a slightly lower binding energy between molecules in the crystal phase - the difference is however tiny and is about the thermal energy.

The lower binding energy could be related to the loss of specific intermolecular bonds in the crystal (as also indicated by a higher ΔH_C for C110M). As an example, the P23T mutant has a higher binding energy than HGD (corresponding to a difference of $2.4 k_B T_0$) and this has been explained by the gain of an intermolecular hydrogen bond as observed in the protein crystal

structure^{96,119}. To see if also in our case the difference in binding energy could be explained by differences in molecular contacts, we performed single-crystal X-Ray diffraction on C110M.

3.3.8 C110M crystal structure solution from single-crystal X-ray diffraction data

C110M rhombic-shaped crystals were analysed by single-crystal X-Ray diffraction and they were found to diffract at 1.57 Å resolution. The final refined structure has been deposited to the PDB with ID 7P53. Details of the data collection, method used to solve the structure and refinement are listed in Table 3.3.

Table 3.3: X-Ray diffraction data collection and analysis of C110M crystals (PDB ID: 7P53)

Resolution range	45.511-1.573 (1.629-1.573)
Wavelength (Å)	0.97918
Space Group	P 2 ₁ 2 ₁ 2 ₁
Unit cell (Å; °)	33.70, 52.84, 89.55; 90, 90, 90
Total Reflections	139173
Unique reflections	22817
Multiplicity	6.1
Completeness (%)	97.78 (81.83)
Mean I/σ(I)	6.0
Wilson B-factor (Å ²)	26.1
R-merge	0.188
R-meas	0.206
R-pim	0.083
Mn(I) half-set CC(1/2)	0.989
Reflections used for R-free	1126
R-work	0.2147 (0.4452)
R-free	0.2738 (0.4395)
Number of non-hydrogen Atoms	1575
Macromolecules	1467
Solvent	108
Protein residues	173
RMS bonds (Å)	0.01
RMS angles (°)	1.07
Ramachandran favoured (%)	97.66
Ramachandran allowed (%)	2.34
Ramachandran outliers (%)	0.00
Rotamers outliers	0.00

Average B-factor	32.55
Macromolecules	32.18
Solvent	37.62
Clashscore	1.05

Initial inspection of the data reveal that crystals belong to $P 2_1 2_1 2_1$ orthorhombic space group with cell parameters $a = 33.70$, $b = 52.84$, $c = 89.55 \text{ \AA}$ and $\alpha = \beta = \gamma = 90^\circ$. C110M crystallises with one molecule in the asymmetric unit and Matthew's coefficient $V_M = 2.05 \text{ \AA}^3/Da$, corresponding to a solvent content of 40%. All these parameters are in perfect agreement with the those of the wild-type protein HGD already present in the PDB (1HK0)²⁰².

The B-factors (atomic displacement parameters containing information on atomic fluctuations in the crystal) of the C110M refined crystal structure are consistently higher than those of the HGD structure 1HK0; although this is most likely due to the lower resolution of our structure, we observe that the average B-factor of the N-terminal domain (residues 1-81) is slightly higher than the one of the C-terminal domain (residues 82-173) for both proteins, suggesting that the mutation is not influencing this parameter. The final refined structure was superposed with HGD using the function MatchMaker available in Chimera software, the results are shown in Figure 3.18:

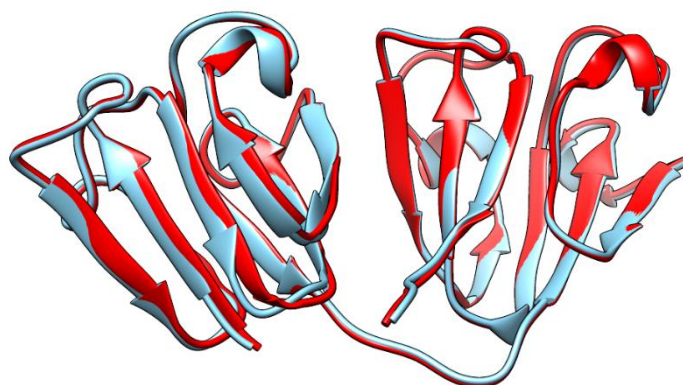


Figure 3.18: Superposition of C110M (red, PDB 7P53) and HGD (light blue, PDB 1HK0) performed with the tool MatchMaker of Chimera software.

The structures superpose nicely with a low root mean square deviation (RMSD) of 0.337 \AA . C110M coordinates also capture HGD secondary structure elements with good agreement.

3.3.9 Crystal contact analysis of C110M structure

In HGD, Cys-110 is part of one crystal contact, i.e. an intermolecular contact occurring in the crystal phase. Inspection of the C110M crystal structure reveals that Met-110 is also part of the same crystal contact; this result was expected since the two proteins crystallise in the same space group. Figure 3.19 shows a symmetry map of the C110M crystal: the molecule in the unit cell in the origin (light blue, molecule A) is close to a second molecule of a different unit cell (molecule

B, yellow) that exposes Met-110. According to the properties of the $P 2_1 2_1 2_1$ spacegroup, the two molecules are related by the symmetry operation $(-x, y-1/2, -z+1/2)$. A close look at the structure suggests that Met-110 of molecule B interacts with Arg-59 of molecule A. The structure suggests that Met-110 of molecule B interacts with Arg-59 of molecule A.

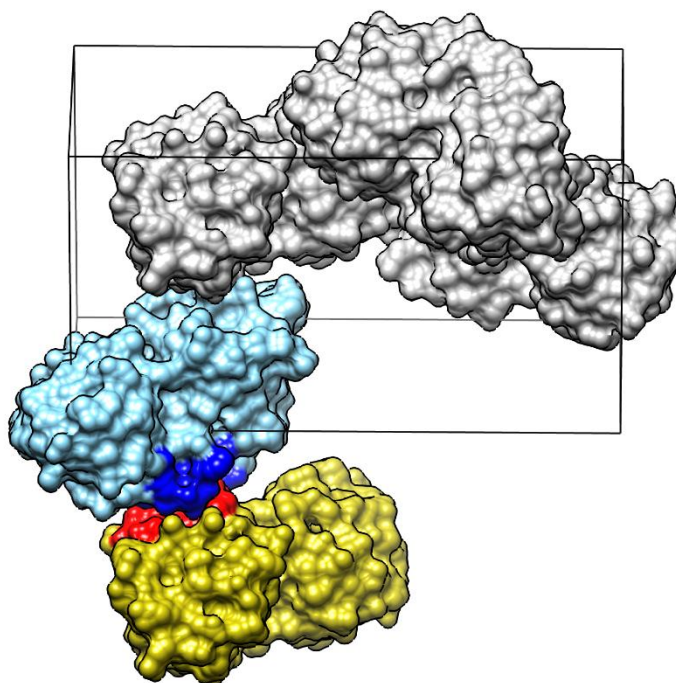


Figure 3.19: Symmetry map of C110M: Met-110 in molecule B [yellow, unit cell origin $(-1, -1, 0)$] is part of a crystal contact formed with molecule A [light blue, unit cell origin $(0, 0, 0)$]. Molecule B is related to molecule A by the symmetry operation $(-x, y-1/2, -z+1/2)$. The two interacting surfaces containing Met-110 and Arg-59 are shown respectively in red and blue. The other molecules that part of the unit cell with origin $(0, 0, 0)$ are shown in grey.

To understand if the mutation C110M caused any variation in the bond network among the two molecules, we carried out interface analysis by PISA webserver and compared with the values for HGD. As shown in Figure 3.20: Interface analysis performed with PISA on A) C110M (7P53) and B) HGD (1HK0) crystal structures., the two proteins generate the same sets of interfaces.

Structure 1			x	Structure 2			interface		ΔG	ΔG	N_{HB}	N_{SS}	N_{DS}	
N_{st}	N_{res}	Surface \AA^2	Range	Symmetry op-n	Sym.ID	N_{st}	N_{res}	Surface \AA^2	area, \AA^2	kcal/mol	P-value			
44	13	8558	x X	$x-1,y,z$	1_455	43	10	8558	397.0	2.8	0.481	6	3	0
34	13	8558	x X	$-x,y-1/2,-z+1/2$	4_545	35	12	8558	296.3	0.7	0.416	8	1	0
34	13	8558	x X	$-x+1,y-1/2,-z+1/2$	4_645	26	11	8558	274.8	2.1	0.449	5	2	0
22	8	8558	x X	$-x+1/2,-y,z-1/2$	2_554	30	9	8558	225.6	-0.8	0.270	3	0	0
5	2	8558	x X	$x-1/2,-y+1/2,-z$	3_455	4	4	8558	30.0	0.7	0.669	0	0	0

Structure 1			x	Structure 2			interface		ΔG	ΔG	N_{HB}	N_{SS}	N_{DS}	
N_{st}	N_{res}	Surface \AA^2	Range	Symmetry op-n	Sym.ID	N_{st}	N_{res}	Surface \AA^2	area, \AA^2	kcal/mol	P-value			
58	19	8653	x X	$x-1,y,z$	1_455	51	12	8653	504.7	6.1	0.625	8	9	0
41	13	8653	x X	$-x,y-1/2,-z+1/2$	3_545	39	13	8653	356.9	2.3	0.543	9	3	0
28	12	8653	x X	$-x+1,y-1/2,-z+1/2$	3_645	26	10	8653	260.6	2.0	0.448	3	2	0
22	7	8653	x X	$-x+1/2,-y,z-1/2$	2_554	25	7	8653	222.5	-0.9	0.241	3	0	0
4	2	8653	x X	$x-1/2,-y+1/2,-z$	4_455	3	2	8653	20.1	1.1	0.842	0	0	0

Figure 3.20: Interface analysis performed with PISA on A) C110M (7P53) and B) HGD (1HK0) crystal structures.

Slight differences in the interface properties are most likely due to a different crystal structure resolution and refinement. As previously stated, residue 110 only takes part in the second listed interface, due to symmetry operation $(-x, y-1/2, -z+1/2)$. A list of all the intermolecular bonds engaged in this crystal contact is reported in Table 3.4:

Table 3.4: Intermolecular bonds predicted by PISA between molecule A and molecule B (related by symmetry operation $(-x, y-1/2, -z+1/2)$ of C110M (7P53). The letters D, G, E, and X stand for greek letters $\delta, \gamma, \epsilon, \chi$ with χ indicating the terminal carboxylic oxygen of the protein. HB and SB stand for hydrogen bond and salt bridge respectively.

Bond (Mol. A-Mol.B)	Bond Distance (Å)		Bond type
	HGD	C110M	
Ser39(OG)-Asp113(OD1)	2.62	2.76	HB
Ser173(O)-Ser136(OG)	2.61	2.64	HB
Asp171(O)-Arg139(NH1)	2.88	2.76	HB
Asp171(O)-Arg139(NH2)	2.98	2.76	HB
Ser173(O)-Arg139(NH2)	3.04	3.08	HB/SB
Ser173(OXT)-Ser136(N)	2.89	2.83	HB
Arg59(NH1)-Gln112(OE1)	3.50	-	HB
Arg59(NH2)-Sulfur110	3.74	3.37	HB
Asp171(OD2)-Ser136(OG)	4.12	3.85	HB
Asp38(OD1)-Arg116(NH1)	3.65	-	SB
Asp38(OD2)-Arg116(NH2)	3.05	-	SB/HB
Asp38(OD2)-Arg116(NH1)	3.50	-	SB

We therefore compared the intermolecular bonds of C110M with the ones of HGD. The first six bonds listed are not significantly different, showing only subtle changes in bonds distances. The most marked differences in the interface bonds between the two proteins are listed below:

- Ser-136 of molecule B, adjacent to the mutated residue, forms a hydrogen bond with Asp-171 with bond length of 4.12 and 3.85 Å respectively in HGD and C110M.
- In HGD, Arg-116 forms three salt bridges, not present in the C110M crystal structure; however, it should be noted that the electron density for this amino acid in our structure was not available and its positioning and conformation were hence assigned arbitrarily. The lack of Arg side-chain electron density is not uncommon, especially if they are surface exposed, and is due to high side-chain flexibility. This could in theory indicate that the protein molecules are less well-bound but it may also be due to experimental

differences. Specifically, finding undefined electron density for flexible side-chains on protein surface was expected for our structure since its resolution is lower than the HGD structure.

- Cys/Met-110 forms a hydrogen bond with Arg-59(NH2) of molecule A. In C110M the bond is shorter (3.37 instead of 3.74 Å) and it causes a second hydrogen bond formed by Arg(NH1) with Gln112(OE1) to be disengaged. The superimposed electron density maps of HGD and C110M around residue 110th are shown in Figure 3.21.



Figure 3.21: *Electron density maps of C110M (black) and HGD (light blue) superposed. Positive and negative blobs of the difference electron density map of C110M are also shown, respectively in red green and red.*

Despite the variations above discussed, PISA analysis suggests that all the interfaces between the two proteins (Figure 3.20) are slightly different in hydrogen bonds, salt bridges, interfacial residue content as well as interfacial area. Those differences in interface properties are most likely due to different crystal structure resolution and refinement. We can then conclude that Cys to Met mutation of residue 110th maintains the crystal contact mostly unaltered causing possibly only subtle changes that slightly influence the enthalpy and entropy of crystallization

3.4 Discussion

Overall, the most prominent difference in C110M phase behaviour compared to that of HGD, is its ability to promptly nucleate and crystallise; this is not an uncommon result since other single-point mutations on HGD surface, such as R36S and R58H, dramatically increase the nucleation ability of the protein but giving rise to crystals with different shapes⁵⁷. Furthermore, site-directed mutagenesis is a widely used method to enhance protein crystallization; the most common approach is the reduction of protein surface entropy (SER) by the substitution of surface amino acids with high conformational entropy, such as Arg, Gln or Lys, with shorter side-chains amino

acids such as Ala^{137,155}; SER was inspired by the “evolutionary negative design hypothesis”, stating that amino acids with high conformational entropy are more frequent on protein surfaces, to hinder protein crystallization *in vivo*⁵⁹. However, this strategy does not usually involve the replacement of surface-exposed cysteines that are generally mutated only to avoid covalent dimerization and hence uncontrolled protein aggregation.

Higher crystal nucleation could be induced by stronger protein-protein interactions in the liquid phase or differences in intermolecular contacts in the crystalline phase. However, we demonstrate that this is not the case and hence we speculate that the cause of the different behaviour of C110M lies in the formation of pre-nucleation clusters. Recently, it was demonstrated that protein crystal nucleation begins with the formation of liquid clusters composed of 2-10 molecules with different characteristics compared to the final crystalline phase^{205,206}. Hence enhanced nucleation could be explained in two ways that do not exclude each other: 1) higher ΔH_C and ΔS_C (obtained by measuring the crystal solubility curve) suggests that molecules in the crystals are less tightly-bound and hence that they may more easily rearrange to find the correct configuration necessary for crystallization⁵⁷ or 2) the mutation may cause differences in the hydration shell in the pre-nucleation stages where more water molecules/more structured water-molecule network would increase the crystallization entropy^{155,207}. The second hypothesis finds support in the fact that, as shown by C110M and HGD refined crystal structures, the solvent accessible surface area (SASA) of residue 110 increases from 30 to 60 Å² with Met being more hydrophobic than Cys; indeed, when replacing Cys-110 by a less hydrophobic amino acid (C110S) we were not able to obtain protein crystals. Recent literature highlights how hydrophobic amino acids on protein surface can induce a better structuring of the water network^{208,209,210,211}. Furthermore, Cys to Met mutagenesis can be considered as analogous to sulphur-methylation; this chemical modification, when applied to surface lysines, has been proved to increase protein nucleation propensity due to a better structuring of the hydration layer around the altered site²¹².

Usually, higher nucleation rates are associated with lower protein solubility (higher crystal binding energy). In our case however, the mutation C110M makes the protein more soluble while increasing its propensity to crystallize. The change in binding energy is only minor (0.77 $K_B T$), hence it is reasonable to assume that the pre-nucleation clusters mechanism described above acts in the early stages of crystallization while the thermodynamic parameters obtained are properties of the final crystalline phase only. A similar behaviour has been described for polymorphs of glycine: while β -glycine is more soluble, it also nucleates faster than the α form. In this case the phenomenon is also explained by different properties of the initial crystallization nuclei compared to the final crystalline phase: their smaller size lowers the interfacial free energy decreasing the nucleation free energy barrier²¹³.

Furthermore, in an analogy to our study, similar work was carried out on bovine γ B-crystallin: the mutation Cys-18 to Ser (C18S) increases protein nucleation propensity while also increasing the binding energy of the crystal phase (lower solubility) by roughly the same amount as in our case. However, a major difference between the described study and our work is that Cys-18 in bovine γ B-crystallin is not a surface exposed residue¹²⁸.

Finally, our results suggest that Cys mutations can have a high impact on protein crystallization and they should be considered not only to suppress/enhance disulphide bridge formation. In crystallins, surface exposed cysteine residues have the drawback of inducing age-related cataract; since this side-effect is engaged only in late age, Cys-110 provides the advantage of suppressing HGD crystallization compared to Met-110, which enhances nucleation. In contrast, Ser does not influence protein solution stability; however, this residue is a low-refractive index amino acid and, along with other primary sequence modifications during evolution, would have decreased the protein refractive power hindering its primary biological function.

3.5 Conclusions

To further address the role of anisotropic protein-protein interactions involved in protein self-assembly, we expressed and purified two mutants of HGD, C110M and C110S. The mutant proteins show no differences in protein structure or net protein-protein interactions in the liquid phase when compared to the wild-type protein. However, while C110S is recalcitrant to crystallization like HGD, C110M promptly crystallizes. We further showed that protein phase diagrams capture differences in anisotropic interactions due to surface single-point mutations: the solubility curve of C110M crystals allowed us to quantify that the crystal binding energy is lower compared to HGD by $0.77 K_B T$.

The crystal structure of C110M reveals the mutant protein crystallizes with the same parameters (space group, unit cell dimensions, etc.) as HGD; however, no major differences in the intermolecular contacts were found and hence the lower binding energy of the mutant must be due to only subtle changes, such as differences in the residue hydrophobicity and hydration layer.

We therefore conclude that, while Cys-110 triggers aggregation causing cataract in late age, it is also essential to maintain protein stability against crystallization without decreasing protein refractive ability, the primary function of eye-lens proteins. We hence demonstrated that the investigation of anisotropic protein-protein interactions due to surface mutagenesis is also useful for speculating about the biological role of specific amino acids within a protein.

Chapter 4:

Polymorphic self-assembly of human γ D-crystallin driven by protein surface anisotropy

4.1 Background

Human γ D-crystallin (HGD) and its cataractogenic mutants provide a robust framework to study the contribution of molecular anisotropy to protein-protein interactions. By increasing the complexity of protein-protein interactions in this way, we can probe the degree to which anisotropy contributes to the phase behaviour of the protein. In this chapter, we create the double mutant P23VC110M, which is a protein with complex phase behaviour, which we have explored using phase diagrams, Dynamic Light Scattering (DLS), light microscopy and Small and Wide X-Ray Scattering (SAXS /WAXS).

The P23V mutant of HGD has similar phase behaviour to the P23T variant, which has been found in cataractous eye lenses of young patients²¹⁴; for this reason, the phase behaviour of those mutants have been studied previously⁵⁶. Both P23V and P23T are less soluble than the wild-type protein and form amorphous assemblies/crystals with a solubility curve that has inverted temperature dependence, i.e. they melt when cooled^{56,215} while the liquid-liquid phase separation boundary has normal temperature dependence. Inverted solubility has been reported for a few globular proteins such as carbomonoxy-haemoglobin C²¹⁶, deoxy haemoglobin S⁹⁸ and bovine pancreatic trypsin inhibitor (BPTI)⁹⁹. In HGD, the 23rd site is not a crystal contact and mutating proline to threonine does not change protein structure²¹⁷. In the double mutant of HGD, P23TR36S, solubility inversion is explained by the formation of a new hydrogen bond in the crystal by Thr-23 that is only activated as the temperature increases¹¹⁹.

Interestingly, the inverted solubility assemblies formed by P23V and P23VR58H are perfectly spherical with a radius of several micrometers^{97,203}. Protein spherical assemblies, have been reported before for a number of proteins and have been classified as spherulites or particulates²¹⁸. Spherulites are widely known in polymer science and are defined as spherical bundles of radially aligned crystallites²¹⁹. However, spherulitic assembly has also been observed for proteins, which can form quasi-spherical bundles of crystalline needles also known as sheaf-like spherulites or sea urchins²²⁰. Furthermore, protein spherulites can also be formed by radially aligned amyloid fibrils and can grow up to a few millimeters in size (amyloid-like spherulites)²²¹. Protein particulates are instead smaller spherical structures (up to a few micrometers) capable of forming networks and to induce gelation²²²; although it is still not clear if their internal morphology is amorphous, recent studies on a range of different proteins showed that particulates are rich in β -sheets resulting from partial structural changes induced by temperature^{218,221}. Several proteins were reported to form either spherulites and/or particulates depending on solution conditions, including bovine β -lactoglobulin²²³, bovine insulin²²⁴, lysozyme²²⁵, bovine serum albumin²²⁶ and human -synuclein²²¹. Although spherical assemblies formed by HGD have morphology that looks

somewhat similar to amyloid-like spherulites or particulates, a morphological study is lacking and differences or similarities have yet to be addressed.

Other than forming spherical assemblies with retrograde solubility, double mutants have a more complex self-assembly behaviour than single mutants; this has been used to explore how anisotropic protein-protein interactions contribute to protein phase behaviour⁹⁷. The phase diagrams of HGD double mutants P23TR36S, P23VR36S and P23VR58H are defined by the properties of their respective single mutants. P23TR36S forms two crystal types, one with inverted solubility due to mutation P23T, the other with normal solubility due to mutation R36S. P23VR36S forms instead only crystals due to mutation R36S since they are more stable (less soluble) than P23V assemblies over the whole temperature range explored. P23VR58H also forms spherical assemblies due to mutation P23V; however, the mutation R58H causes crystallization on their surface making the P23VR58H solubility curve dependent on crystallization kinetics and hence a combination of the solubility curves of the two single mutants⁹⁷. Although the phase behaviour of double mutants have been demonstrated to be complex, it is possible to design new double mutants incorporating the desired properties given by known single-point mutations.

4.2 Aim of the study

A new double mutant of HGD, P23VC110M, has been designed to include the properties of two single-point mutations: 1) P23V, that forms spherical aggregates with inverted solubility and 2) C110M (Chapter 3), that leads to higher crystal nucleation rates while maintaining the same structural properties and protein-protein interactions in the liquid phase. A surface model of HGD indicating the two mutation sites is shown in (Figure 4.1). The P23V mutation was chosen over P23T as it is soluble over a larger concentration range, allowing a greater phase space to be explored. Given the work previously done on double mutants of HGD⁹⁷, our aim is to probe if in this case, the phase diagram of the new double mutant can be rationalized given those of the two single mutants and hence if this approach can be generalized. Furthermore, the internal structure of the spherical assemblies formed due to the 23rd proline mutations in HGD require further investigation.

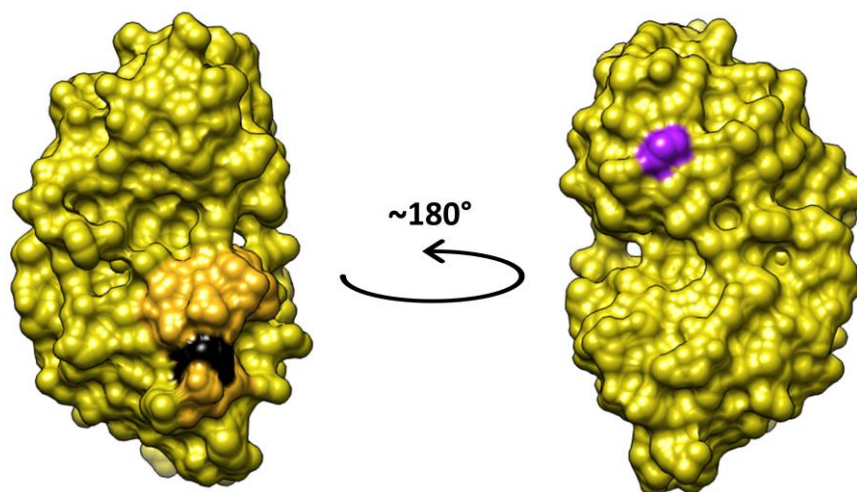


Figure 4.1: Surface model of HGD; the 23rd and 110th positions are highlighted in purple and black respectively.

4.3 Results and discussion

4.3.1 Production and characterisation of the double mutant P23VC110M

The P23VC110M double mutant plasmid was produced by site-directed mutagenesis starting from the P23V single mutant plasmid⁹⁷. The desired mutation, C110M, was confirmed by DNA sequencing with a T7 promoter showing the replacement of codon TGT with ATG. DNA purity was confirmed by UV-Vis spectroscopy providing the following results:

$A_{260\text{ nm}}/A_{280\text{ nm}}$	1.9
$A_{260\text{ nm}}/A_{230\text{ nm}}$	2.2
DNA concentration	16 ng/mL

The plasmid was then transformed into *E.Coli* BL21 (DE3), the protein expressed and purified by size-exclusion (Figure 4.2) and ion-exchange chromatography (Figure 4.3).

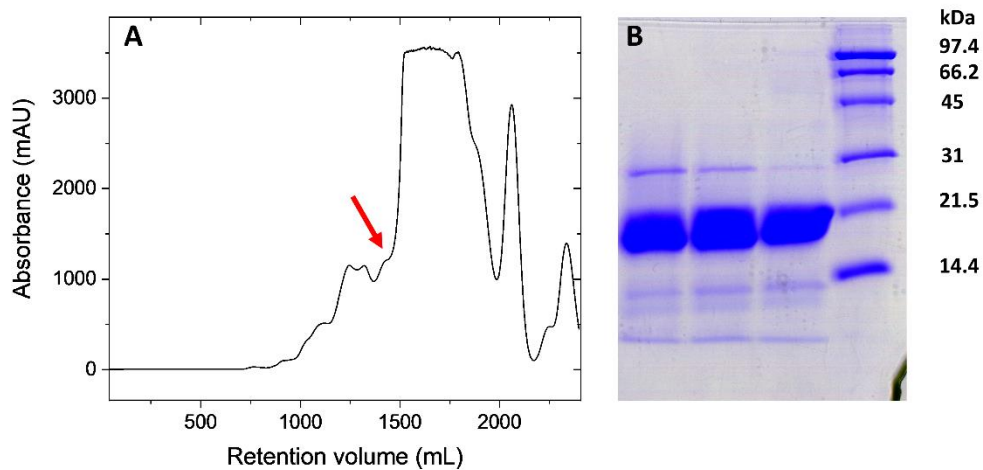


Figure 4.2: A) Size-exclusion chromatogram of P23VC110M and B) SDS-PAGE of fractions from size-exclusion. The red arrow in A) indicates the fractions pooled as indicated by SDS-PAGE.

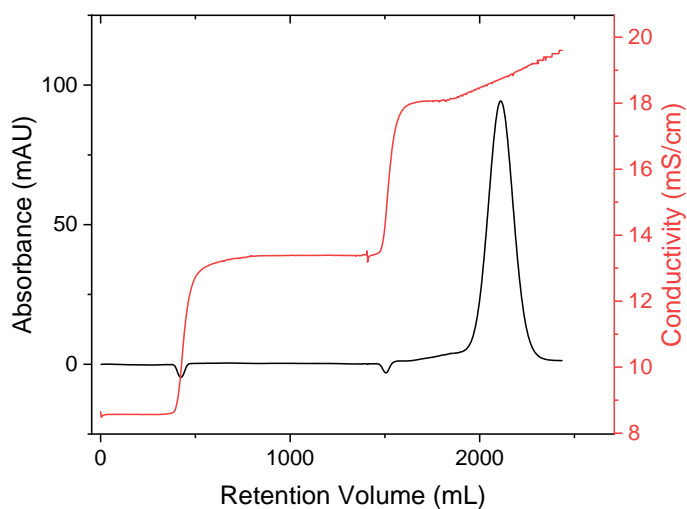


Figure 4.3: Ion-exchange chromatogram of P23VC110M purification.

P23VC110M purifications yielded typically 80 mg of purified protein from 7.2L of bacterial cell culture. The purity of each batch was assessed to be $\geq 98\%$ by SE-HPLC and SDS-PAGE (Figure 4.4)

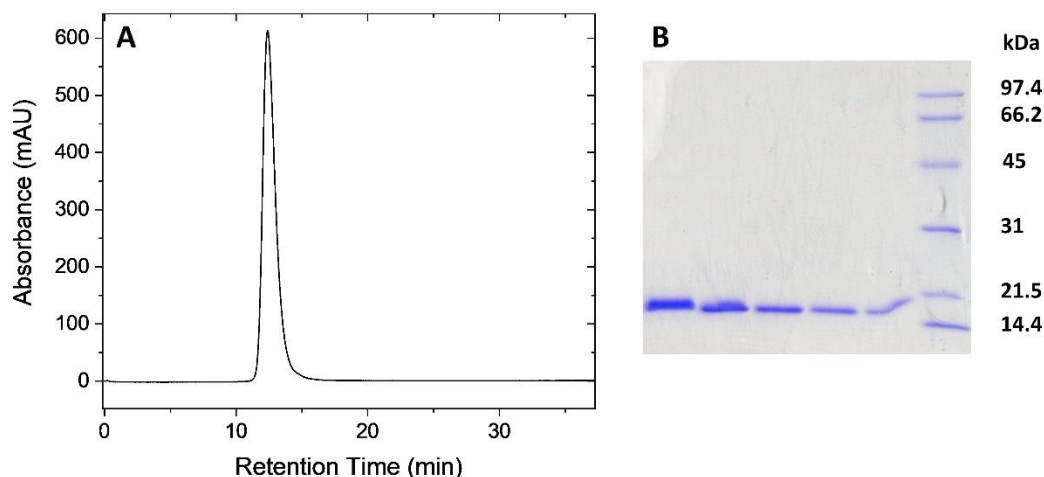


Figure 4.4: Purity assessment of P23VC110M by A) SE-HPLC and B) SDS-PAGE.

As shown in the SE-HPLC chromatogram (Figure 4.4A), the P23VC110M double mutant does not form covalent dimers due to replacement of Cys-110 with Met. Hence the use of DTT in the buffer is unnecessary. The ratio $A_{260\text{nm}}/A_{280\text{nm}} < 0.5$ obtained by UV-Vis spectroscopy indicates that P23VC110M batches were free of any residual DNA. The accurate molecular weight of the mutant was confirmed to be 20637 ± 1 by electrospray ionization mass spectrometry (ESI-MS).

4.3.2 Structural characterisation of P23VC110M

Fluorescence emission (Section 2.4.2) and second-derivative UV absorption spectroscopies (Section 2.4.1) can be used to investigate the exposure of aromatic amino acids of the protein and hence to compare the structures of the native and mutant proteins to determine if structural differences exist. To assess whether the two mutations in P23VC110M caused any significant change to the structure of HGD, we compared their fluorescence emission (Figure 4.5) and second-derivative UV spectra (Figure 4.6). Fluorescence emission spectra are normalized by the value of maximum intensity:

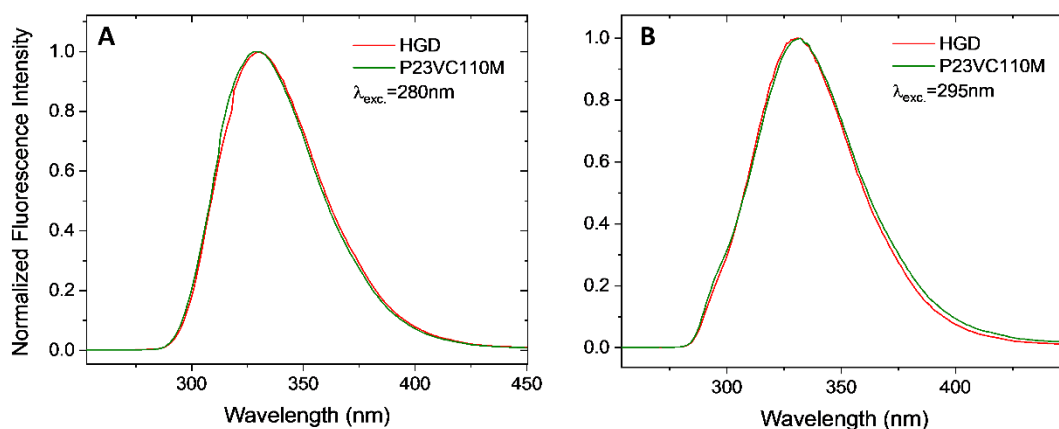


Figure 4.5: Fluorescence emission spectra of HGD and P23VC110M using excitation wavelengths of A) 280 nm and B) 295 nm at a concentration of 0.3 mg/mL.

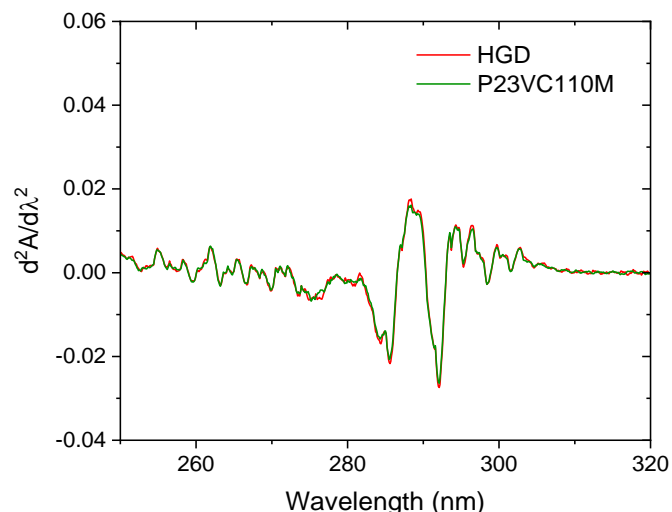


Figure 4.6: Second-derivative UV spectra of HGD and P23VC110M.

Fluorescence emission and second-derivative UV spectroscopy reveal that the two mutations induce no significant changes to the structure of the protein. CD spectroscopy in the far-UV region (Section 2.4.3) on HGD and the double mutant P23VC110M (Figure 4.7) also indicate no structural change to the protein:

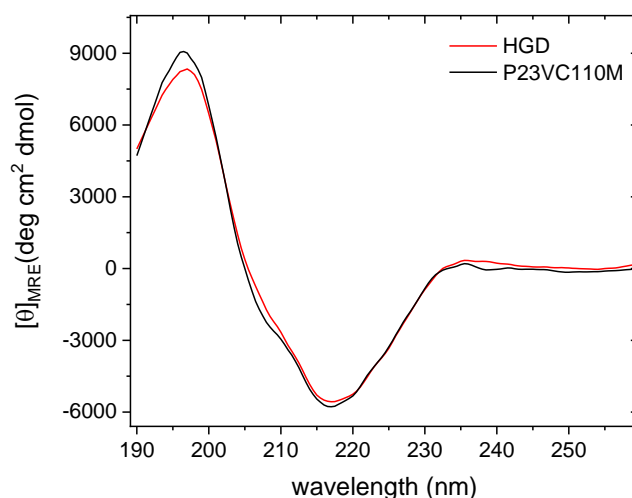


Figure 4.7: Circular Dichroism spectra in the far-UV region for HGD and its P23VC110M mutant. Spectra were recorded from samples at a concentration of 0.7 mg/mL.

4.3.3 Protein-protein interactions in the liquid phase of P23VC110M

The strength of net protein-protein interactions in the liquid phase of the double mutant P23VC110M were quantified by Dynamic Light Scattering (DLS) as described earlier (Sections 1.2.3 and 2.4.4). The slope of D_c/D_0 plotted against protein volume fraction (Figure 4.8) provides the net-interaction parameter k_D according to the following equation (Section 2.4.4) :

$$D_c = D_0(1 + k_D\phi) \quad 2.16$$

$$R_H = \frac{k_B T}{6\pi\eta D_0} \quad 2.15$$

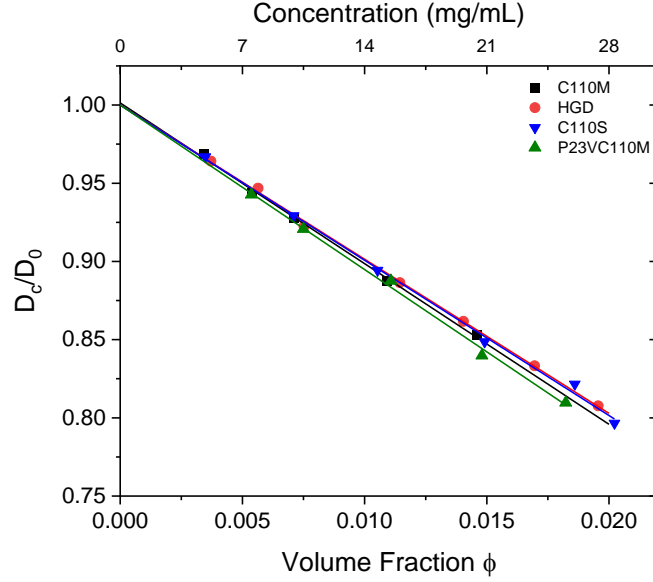


Figure 4.8: Viscosity corrected diffusivity for P23VC110M and HGD, C110S and C110M mutants (Section 3.3.5). The slope indicates a negative value for k_D and therefore, net-attractive protein-protein interactions, which are unchanged by mutagenesis.

The net-interaction parameter is compared with the those of the previously studied proteins HGD, C110M and C110S (Section 3.3.5) in Table 4.1

Table 4.1: k_D , D_0 and R_H (with errors) of P23VC110M and the previously studied proteins HGD, C110M and C110S (Section 3.3.5).

Protein	k_D	$D_0 (x10^{-7}) \text{ cm}^2/\text{s}$	$R_H \text{ (nm)}$
HGD	-9.86 ± 0.13	9.31 ± 0.02	2.30
C110M	-10.27 ± 0.28	8.97 ± 0.06	2.39 ± 0.02
C110S	-9.92 ± 0.26	8.81 ± 0.03	2.43 ± 0.01
P23VC110M	-10.52 ± 0.35	8.89 ± 0.04	2.41 ± 0.01

The analysis of the DLS data indicates that the polydispersity indices (PDIs) are less than 0.1, which for protein solutions indicates that the protein is in its monomeric form and the samples are free of aggregates; R_H and k_D are in good agreement with literature values for HGD²⁰¹. Also, the comparison of k_D values shown in Table 4.1 indicates that mutagenesis does not produce a protein

in which the net protein-protein interactions have changed, at least in the low concentration regime.

4.3.4 Solubility curves of P23VC110M polymorphs and thermodynamic analysis of crystallization

The single mutant P23V forms spherical assemblies with retrograde solubility if incubated at 30 °C or higher temperatures. These assemblies are reversible with temperature⁹⁶. The mutation C110M results in spontaneous crystallization, producing orthorhombic crystals following liquid-liquid phase separation. Previous work has shown that both characteristics should be retained in the double mutant protein containing both of these mutations⁹⁷. When two distinct mutations result in the protein forming two different solid phases, the possible scenarios are: 1) the first phase is more stable than the second (i.e. less soluble) at all temperatures, hence the second one will not form or 2) the first phase is more stable than the second only in a specific temperature range, hence, the desired phase can be selectively obtained by choosing the incubation temperature where it is most stable - in this case their solubility curves will intersect. This has been demonstrated for HGD double mutants P23VR36S and P23TR36S in which both 1) and 2) were observed (Figure 1 and 2 of ref. 97). In addition to thermodynamic restraints on the phase behaviour, nucleation kinetics also contributes and less stable phases are obtainable (at least temporarily) in some cases (e.g. P23VR58H). It is therefore possible to design double mutants with these hybrid phase behaviours by overlapping the phase diagrams of the two single mutants. As shown in Figure 4.9, the solubility curves of P23V and C110M intersect (orange shaded area) in a region of the phase diagram. Based on the work done on earlier mutants, for P23VC110M, behaviour 2) is expected.

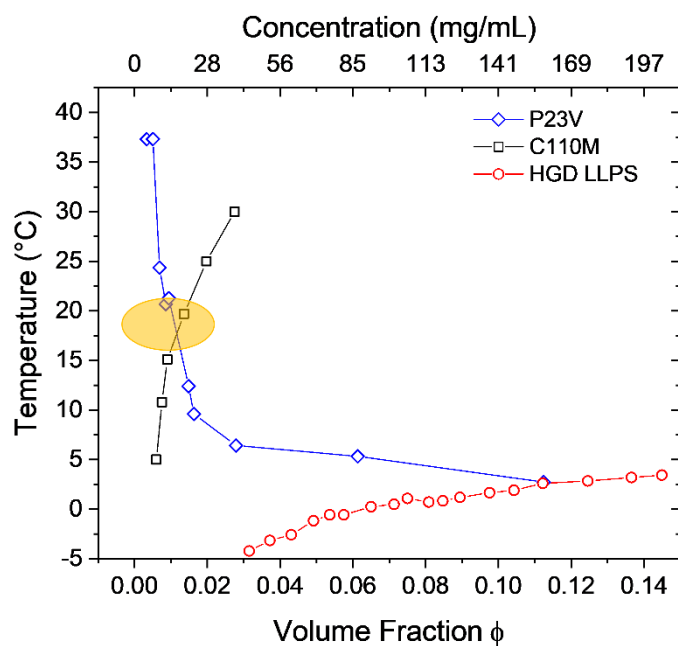


Figure 4.9: The equilibrium solubility of P23V spherical assemblies (ref.203, 227) and C110M orthorhombic crystals (Section 3.3.7) at different temperatures are shown along with HGD liquid-liquid phase separation boundary (ref. 57). The two solubility curves intersect indicating that the double mutant P23VC110M is expected to form two different polymorphs at different temperatures. The shaded orange area highlights the intersection between P23V and C110M solubility curves.

The P23VC110M double mutant of HGD has very complex phase behaviour. We observe spherical assemblies with inverted solubility when incubated at 30-37 °C (concentration 15-30 mg/mL [corresponding to volume fractions of 0.0106 – 0.0213], above the intersection point, Figure 4.10), which is consistent with the behaviour of the P23V single mutant and expected behaviour for the double mutant. Attempts to measure the solubility of these spherical structures with decreasing temperatures (near the intersection) produced orthorhombic crystals at about 18°C, without complete dissolution of the spherical assemblies; these crystals are of a similar shape to the crystals formed by the C110M single mutant and were present simultaneously with the spherical assemblies (Figure 4.10, left middle picture). This suggests that two polymorphs exist within a narrow region of the phase diagram (orange shaded area of Figure 4.9 and Figure 4.10). To further confirm this result, more experiments may be directed to observing by microscopy how the shapes of orthorhombic crystals and spherical assemblies are affected upon reversible temperature changes (i.e. rounding of crystals edges or shrinking of spherical assemblies upon dissolution).

The P23VC110M spherical assemblies have a solubility curve that overlaps with that for P23V²²⁷ (Figure 4.10). If our predictions about the phase behaviour are correct, the solubility curves of orthorhombic crystals formed by P23VC110M and C110M should overlap. To confirm this, we attempted to crystallize them selectively, with no coexisting spherical aggregates. Therefore, we

used the same method as for the single mutant C110M (LLPS induction, Section 2.6.1) that involves incubation at low temperatures, which are below the solubility curve for the reversible retrograde spherical assemblies, and will therefore suppress their formation. In short, this was done by setting the initial solution conditions in the narrow region between P23V solubility curve and HGD liquid-liquid phase separation boundary, where the single mutant C110M is supersaturated and P23V undersaturated (Figure 4.9). Surprisingly, we obtained a third polymorph, i.e. crystals resembling needles or elongated plates with a different solubility curve than the orthorhombic crystals (Figure 4.10).

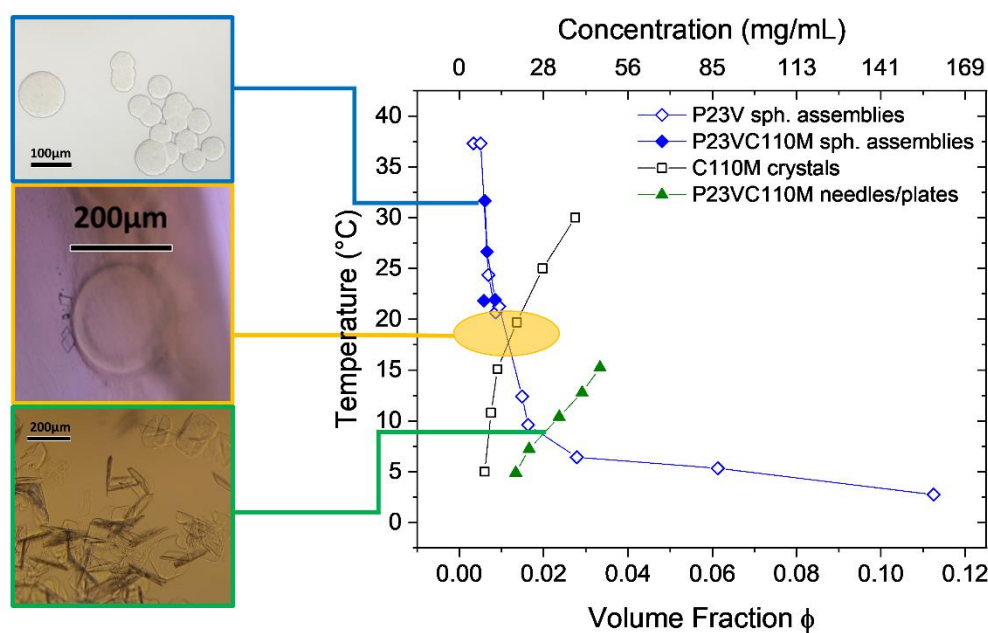


Figure 4.10: The equilibrium solubilities of P23VC110M polymorphs at different temperatures (full symbols) are shown with the ones of the single mutants P23V and C110M (empty symbols, taken respectively from ref.56 and Section 3.3.7). The orange shaded area indicates the intersection between P23V and C110M solubility curves as per Figure 4.9. The microscopy images countered in blue, orange and green show respectively 1) P23VC110M spherical assemblies, 2) the co-existence of P23VC110M spherical assemblies and orthorhombic crystals and 3) P23VC110M elongated plates/needles. Each microscopy image is linked to the respective phase diagram region.

Since when using the C110M crystallization method only plates/needles form, we were not able to measure the solubility curve of P23VC110M orthorhombic crystals; however, it is likely that they would have the same solubility curve as C110M orthorhombic crystals.

Thermodynamic analysis of needles/plates solubility curve was carried out as outlined in Section 3.3.7 using Van't Hoff equation:

$$\ln\phi = \frac{\Delta G_C}{RT} = \frac{\Delta H_C}{RT} - \frac{\Delta S_C}{R} \quad 3.1$$

ΔG_C , ΔH_C and ΔS_C are the Gibbs free energy, enthalpy and entropy of crystallization respectively, T the temperature, ϕ the protein volume fraction, R the gas constant and T_0 the reference temperature. Since experimental data for P23VC110M needles/plates were not available at 298K (e.g. the reference temperature used for C110M, Section 3.3.7), here we used $T_0=283K$; for comparison, the same analysis was repeated for C110M orthorhombic crystals at 283K (Figure 4.11)

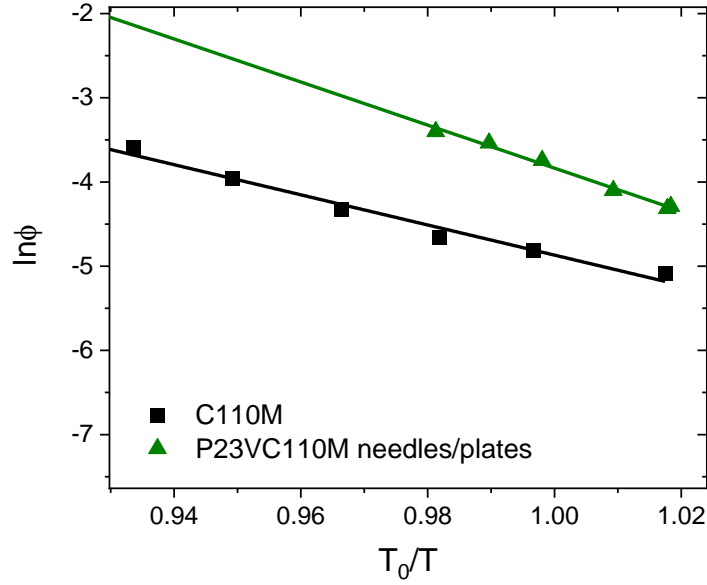


Figure 4.11: Temperature dependence of C110M orthorhombic crystals and P23VC110M needles/plates solubility; T_0 is the reference temperature taken as 283K. The slope and the intercept of the linear fits provide respectively ΔH_C and ΔS_C .

Protein	$\Delta G_C(T_0)(KJ/mol)$	$\Delta H_C(KJ/mol)$	$\Delta S_C(J/mol \cdot K)$	$-T_0\Delta S_C(KJ/mol)$
P23VC110M	-9.0	-60.1	-180.7	51.1
C110M	-11.4	-42.1	-108.4	30.7

Table 4.2: Thermodynamic parameters for P23VC110M and C110M crystallization. T_0 is the reference temperature taken as 283K.

As indicated by the thermodynamic parameters shown in Table 4.2, the P23VC110M needles/plates are less stable than C110M orthorhombic crystals since, at 283K, $G_C^{P23VC110M} - G_C^{C110M} = 2.4 KJ/mol$ that corresponds to $\mu_C^{P23VC110M} - \mu_C^{C110M} = 1.03 k_B T_0$. The enthalpy of crystallization of P23VC110M needles/plates is more negative than C110M, indicating that more/stronger intermolecular contacts are engaged in the crystal phase. In addition, the entropy of crystallization is more negative: this suggests that the hydration shell of the double mutant has less water molecules or a less structured water molecule network than C110M. Unlike C110M in

Section 3.3.7, it is likely here that the high nucleation rate of P23VC110M needles/plates is related to stronger intermolecular contacts in the crystal phase.

P23VC110M needles/plates are hence less stable (i.e. more soluble) than C110M orthorhombic crystals, but they still form in apparent contradiction to the expected phase behaviour 2); this can be explained considering P23VC110M plates/needles to be kinetically favoured over the orthorhombic crystals and hence allowed to form even if they are not the most stable thermodynamic phase. To confirm this, we induced crystallization by LLPS at lower concentrations, i.e. 75 mg/ml instead of 120-150 mg/mL. At lower concentrations the nucleation rate decreases and may allow the stable orthorhombic crystals to form or, at least, needles/plates may nucleate more slowly allowing the formation of diffraction-quality single crystals for structure determination. However, solutions at lower concentrations still produced poor-quality needles/plates. Furthermore, needles/plates were also obtained storing 35mg/mL solutions at 4°C for about two weeks with no LLPS induction, mild crystallization conditions where C110M orthorhombic crystals were never observed. This suggests P23VC110M needles/plates nucleate promptly, supporting the kinetically favoured crystallization hypothesis.

Unfortunately, none of our low-concentration samples produced diffraction-quality needles/plates. Nevertheless, a different crystal habit most likely indicates another space group and/or crystal contacts which would also result in a different solubility curve and nucleation rate; since P23V crystals with normal solubility have never been reported, it is then likely that P23VC110M needles/plates emerge by the formation of new crystal contacts, possibly involving both mutation sites. One way to approach this may be by measure the mass change of crystals upon drying to estimate their water content²²⁸. If compared with the water content from the diffraction data for C110M crystals (40%, Section 3.3.8), this may allow different space-groups to be identified and/or determine the reason for the lower stability of needles/plates.

4.3.5 Characterisation of P23VC110M spherical assemblies by Light Microscopy

Mutations on the 23rd site allow HGD to form assemblies with retrograde solubility and a spherical shape. The molecular mechanism for retrograde solubility has been already investigated for mutant P23T^{119,229} and reasonably this may be similar to the one affecting other HGD mutants such as P23V; however, less is known about the morphology of HGD mutant spherical assemblies, which are observed for P23T, P23S, P23V, P23VR58H^{56,203} and now for P23VC110M mutants. Regardless of the mutant they arise from, the spherical assemblies share similar properties. From previous investigations on P23V and P23VR58H, it has been shown that the spherical assembly growth rate is linearly dependent on the initial protein concentration which indicates first-order kinetics²⁰³; this finding is consistent with a diffusion-limited nucleation and

growth mechanism which consists of the formation of an initial nucleus where monomers from the solution consequently add to the particle surface²⁰³. Their maximum diameter reaches 40-50 μm in approximately 8 hours; when the growth phase of P23VR58H spherical assemblies is complete, crystals due to mutation R58H start growing on their surface while no crystallization is observed for P23V²⁰³. It has been shown that the growth of spherical assemblies and surface crystallization are two uncoupled phenomena, i.e. the latter occurs only when spherical assemblies reach their maximum size²⁰³. Furthermore, when nucleation is controlled the assemblies all grow at the same rate and hence show very low polydispersity²³⁰. Interestingly, the HGD mutant W42E (tryptophan to glutamic acid) also forms assemblies with a spherical shape¹⁹⁶ but unfortunately a thorough morphological study is lacking.

Spherical superstructures formed by other proteins (Section 4.1) can be either classified as amyloid-like spherulites or particulate depending on their characteristics; to address if P23VC110M assemblies lie in either group, we studied their external and internal morphology by microscopy, Thioflavin-T (ThT) binding assays and small and wide-angle X-Ray scattering (SAXS and WAXS).

Cross-polarized light microscopy can be used to obtain information on the internal morphology of the spherical assemblies: optically birefringent objects generally indicate ordered structures and hence appear bright in a cross-polarization experiment (Section 2.5.1). Therefore, this technique might indicate if the spherical assemblies have any internal order; furthermore, it may indicate amyloid-like spherulites since other studies have suggested that a characteristic Maltese cross pattern is due to the radial alignment of amyloid fibrils in the interior²³¹. Birefringence has been observed for P23VR58H spherical assemblies (after their growth phase) but not for P23V; this has been interpreted as crystallization on their surface due to the high crystal nucleation rate provided by mutation R58H²⁰³. Freshly grown P23VC110M spherical assemblies (Figure 4.12A) are not birefringent, suggesting the lack of any internal order. However, some aged samples of P23VC110M spherical assemblies show rough surfaces and faint birefringence (Figure 4.12B and C), showing that surface crystallization is possible also for P23VC110M.

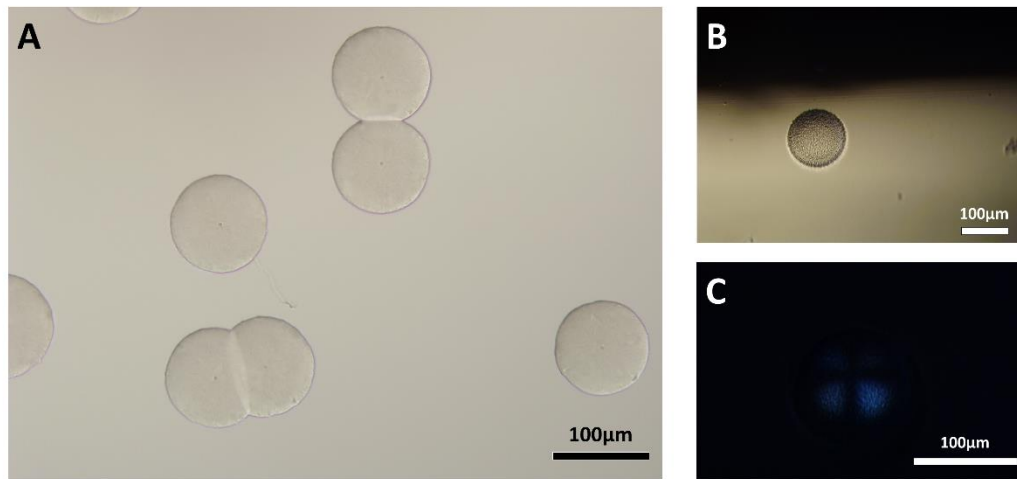


Figure 4.12: A) Freshly prepared spherical aggregates of P23VC110M. They show no birefringence if seen by cross-polarizers (black image not shown). B) Aged P23VC110M spherical aggregates often show a rough surface that produces C) faint birefringence in a cross-polarization experiment.

The absence of a Maltese cross pattern in freshly prepared spherical assemblies indicates the internal material has no radial ordering; however, this does not exclude the presence of amyloid fibrils with random orientation. Fibrillar material has never been observed in the interior or surroundings of spherical assemblies of the double mutant, even when they were mechanically broken or dried (Figure 4.13A and B) suggesting amyloid fibrillation is not involved at all in spherical assembly formation.

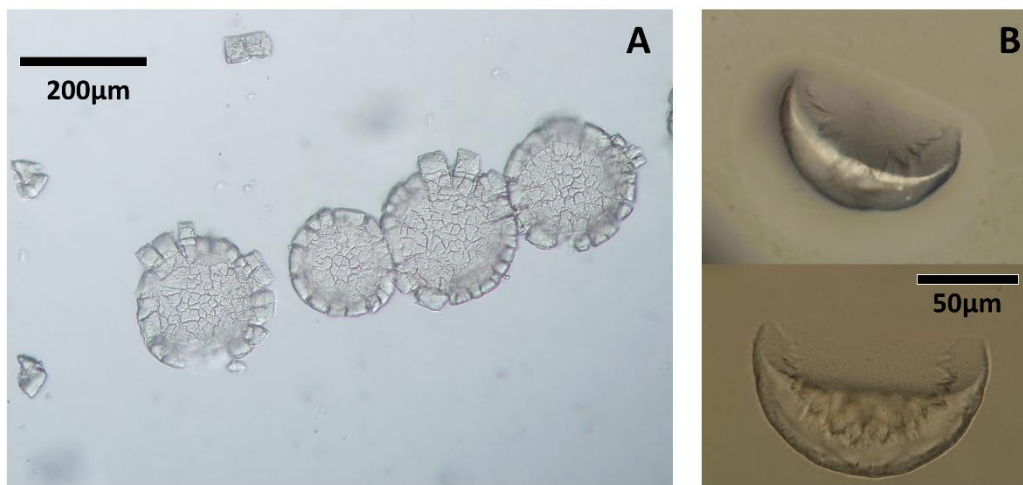


Figure 4.13: A) Mechanically broken and B) dried P23VC110M spherical assemblies; fibrillar material is not present either in the vicinity or in the interior of the spherical assemblies.

4.3.6 Characterisation of P23VC110M spherical assemblies by ThT binding assay

Light microscopy experiments show no evidence of amyloid fibrils in the interior or vicinity of P23VC110M spherical assemblies. As a further confirmation, we employed a ThT binding assay,

a fluorescence technique that is often used to detect the formation of amyloid fibrils¹⁶⁹ ; the intensity of fluorescence emission due to ThT in P23VC110M samples with and without spherical assemblies is shown in Figure 4.14 and indicates that no significant increase in ThT fluorescence is observed in the spherical particles and therefore excludes the possibility that amyloid-like material is present in the interior.

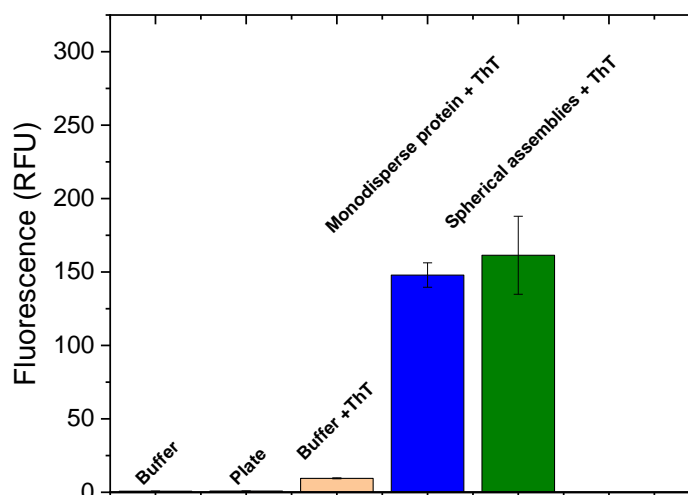


Figure 4.14: Fluorescence emission intensity of ThT from a monodisperse P23VC110M solution at 30mg/mL (blue bar) and from a suspension of spherical assemblies obtained incubating the monodisperse solution at 30°C (green bar). The fluorescence emission intensity of the buffer, the fluorescence plate and Buffer + ThT alone (pink bars) are shown as a control.

4.3.7 Characterisation of P23VC110M spherical assemblies by SAXS and WAXS

To gain more insights into the morphology of P23VC110M spherical assemblies, we employed SAXS and WAXS to study their growth. However, these techniques necessitate an initial characterisation of the protein sample in a monodisperse and diluted form; this is usually considered good practice since the main difficulty of SAXS lies in data analysis that can easily lead to over-interpretation¹⁷⁹. Initial characterisation of P23VC110M by SAXS was performed on a 5 mg/mL sample; the experimental $I(q)$ profile, Guinier and pair distribution function analysis (Section 2.4.5) are shown in Figure 4.15:

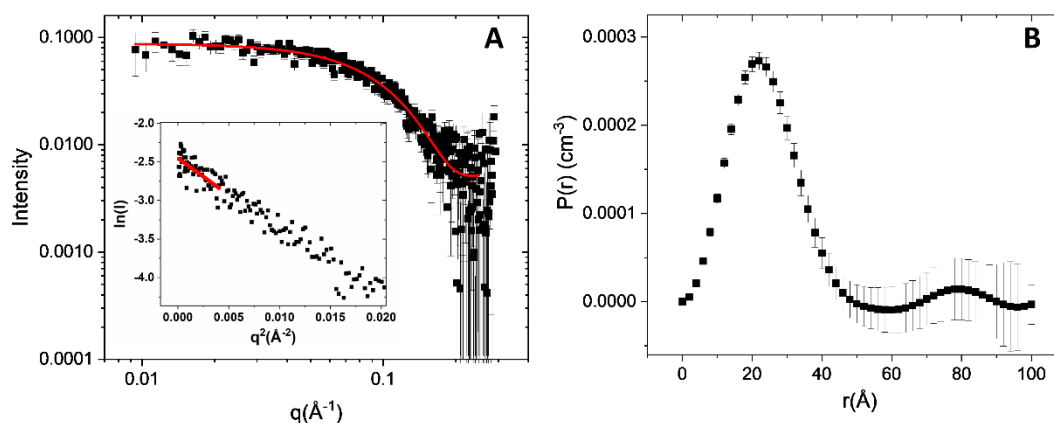


Figure 4.15: A) X-Ray Scattering intensity of a P23VC110M sample at 5mg/mL; initial analysis carried out through Guinier equation leads to $R_g=1.69\text{nm}$ ($R^2=0.80$) (inset). The dataset was then fitted using a spherical form factor and a sticky hard sphere structure factor (red fit line in A) leading to $R_g=1.94\text{ nm}$ ($\chi^2=1.0276$). B) Pair distribution function analysis giving $R_g=1.82\text{ nm}$ ($\chi^2/\text{dof}=235$).

Initial inspection of SAXS data (Figure 4.15 A) shows the curve is flat in the low- q region, indicating the sample is essentially monomeric²³². The Guinier equation gives a radius of gyration $R_g=1.69\text{ nm}$ (inset of Figure 4.15 A); however, this type of analysis is usually performed only as an initial inspection and it is well known to be inaccurate causing under or over estimations of the radius of gyration¹⁷⁹. Better results are usually provided by the pair distribution function analysis (Figure 4.15 B) that gives $R_g=1.82\text{ nm}$; the bell-shape and sharp decay of the $P(r)$ function also confirm that the protein is approximately spherical and in a folded state. The scattering intensity was also fitted using a spherical form factor and a sticky hard-sphere model as a structure factor; the fitting gives $R_g=1.94 \pm 0.11\text{ nm}$, polydispersity 0.13 (indicating the sample is essentially monomeric), stickiness parameter $\varepsilon=0.1157$ and fitting error $\chi^2=1.0276$. The radii of gyration obtained both from fitting the experimental data and from the $P(r)$ function are in good agreement with the hydrodynamic radius measured by light scattering (as a rule of thumb $R_H=1.3R_g$ [Section 4.3.3]).

To address if the interior of spherical assemblies is amorphous or crystalline, we monitored the initial steps of aggregation over time by WAXS; the presence of sharp peaks in WAXS profiles commonly indicates a certain degree of crystallinity^{178,233}. Since this technique monitors only sub-nanometer length scales, SAXS was also employed simultaneously to verify that aggregation was occurring. Although spherical assemblies were usually obtained at 30-37 °C, the experiment was carried out at room temperature to better capture the initial features of aggregation. According to P23VC110M phase diagram (Figure 4.10) the concentration chosen for the experiment (50 mg/mL) was well above the spherical assembly solubility limit. The sample was monitored for approximately 11 hours starting immediately after preparation; SAXS and WAXS profiles at different time points are shown in Figure 4.16 A and B:

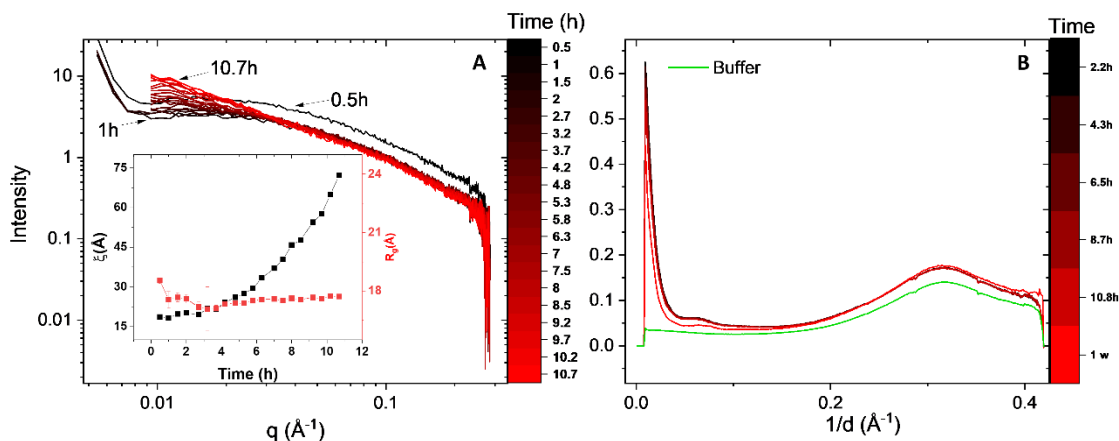


Figure 4.16: Early stages of aggregation monitored over time by A) SAXS and B) WAXS in a P23VC110M 50mg/mL sample. SAXS profiles were fitted with spherical form factor and the Ornstein–Zernike model fixing the minimum R_g to 1.7 nm (highest error $\chi^2=3.02$), fitting results are shown in the inset of A): R_g (red) remains constant while the correlation length ξ (black) increases. Unsubtracted WAXS profiles (shown along with the buffer, green) are reported as a function of $1/d$, where d is the interplanar spacing as per Bragg’s law ($d = 2\pi/q$).

SAXS profiles are flat in the low- q region at the start of the experiment, suggesting the sample is mostly monomeric, but develop a shoulder at later time points, which is indicative of aggregation. A spherical form factor and Ornstein–Zernike model were used to fit the experimental datasets; the results in terms of correlation length ξ and radius of gyration R_g are shown in the inset of Figure 4.16A. The radius of gyration, R_g remains constant while the correlation length ξ , which is the maximum distance of correlation between two solute molecules in solution, sharply increases suggesting aggregation. At the same time, WAXS profiles show a small broad feature at low- q and a band due to water structure at high- q ; they do not show any sharp peak(s) due to crystallinity or any minimal change during the course of the experiment. However, since we chose to mildly trigger aggregation avoiding high temperatures, spherical aggregates were not visible by eye at the end of the experiment; the sample was then remeasured by WAXS after one week when visible aggregation occurred (Figure 4.16 B, bright red line) giving the same result. These findings show that the interior of spherical assemblies is composed of amorphous material.

Furthermore, to check if crystalline material can emerge after aggregation, we measured two more samples of spherical assemblies prepared by incubation at 37 °C, the first fresh and the second two months old; the solid material was pelleted down to the bottom of the tube to increase the scattering intensity. The WAXS profiles of the fresh and old pelleted samples are shown in Figure 4.17:

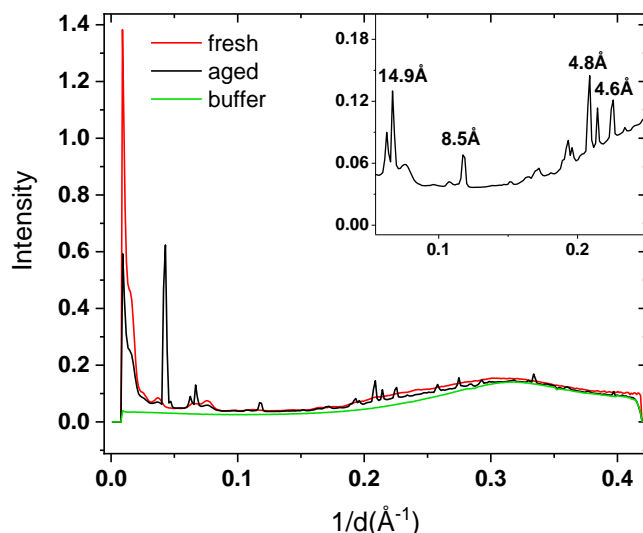


Figure 4.17: Un-subtracted WAXS profiles of P23VC110M spherical assemblies pelleted to the bottom of X-Ray capillaries shown along with the buffer. The fresh and aged samples were used immediately and two months after preparation. The inset shows a zoom in of the WAXS profile of the aged sample with the interplanar distances of some peaks in Å. The absence of sharp peak in the fresh sample indicates the spherical assemblies are not formed by crystalline material.

The WAXS profile of the fresh sample shows only few small and broad features at low- q suggesting the aggregates are mostly composed of amorphous material. On the other hand, the aged sample clearly shows several sharp peaks, indicating a certain degree of crystallinity; while crystallinity may also be due to aging of the internal amorphous material, it is likely that the sharp peaks are a consequence of crystallization on the spherical assembly surface as observed by microscopy (Section 4.3.5). The WAXS pattern of the aged sample shows reflections at 8.5 and 4.8-4.6 Å (Figure 4.17, inset); these values may correspond respectively to β -strand and β -sheet distances found in amyloid fibrils (Section 2.4.2.1). Along with other techniques, X-Ray diffraction is indeed routinely used to identify amyloid fibrils from these specific reflections. Although in our case these reflections are present, they have low intensity and appear among several other sharp peaks; no other technique employed suggests the presence of amyloid fibrils indicating that they are likely not the cause of such reflections. More importantly, we observe those features after aging, suggesting they are not related to the intrinsic properties of the spherical assemblies but more likely to aging/crystallization.

4.3.8 Comparison of P23VC110M spherical assemblies with other protein spherical superstructures

Given the discussion of P23VC110M spherical assemblies properties provided in the previous sections, we aim here to summarize and compare their features with those of amyloid-like spherulites and particulates; Table 4.3 shows a comparison of their main properties.

	Size	Conditions of formation	Degree of unfolding/ Secondary structure	Internal structure	Other
P23VC110M	1-200 μ m	At protein pI. T >25 °C	Native	Amorphous	Reversible
Amyloid-like Spherulites	μ m-mm	Away from protein pI. T >37 °C	Extensively unfolded/ Cross β -Sheet	Radially aligned amyloid fibrils	Maltose-cross. Irreversible
Particulate	nm-few μ m	At protein pI. High Temperatures (~ 90 °C) and ionic strength.	Mildly unfolded. Depends on protein	Irregular/fractal-like.	Gels

Table 4.3: Comparison of P23VC110M spherical assemblies with amyloid-like spherulites and particulate. Properties of amyloid-like spherulites and particulate are summed up according to ref. 218, 221, 234.

It is important to note the relationship between the conditions of formation for amyloid-like spherulites/particulates and their final morphology; an explanation based on DLVO theory is proposed in ref. 221 and summed up below. Protein particulates form at the isoelectric point (pI) and at high ionic strengths, conditions where long-range electrostatic repulsions are suppressed. Incubation at high temperatures partially unfolds protein molecules, allowing aggregation through the hydrophobic effect; since there is no net charge on the protein surface to counterbalance this attraction, the aggregation is fast and results in irregular/fractal-like structures until a spherical shape is reached. The structural changes that trigger this type of aggregation are minimal, and protein molecules mostly retain their native structure. On the other hand, amyloid fibrils form away from the pI, when electrostatic repulsions are strong; the fine balance between repulsion and attraction due to the hydrophobic effect induced following unfolding leads protein molecules to aggregate less quickly, allowing them to arrange in a cross β -sheet structure forming amyloid fibrils and, finally, amyloid-like spherulites.

We have shown that P23VC110M spherical assemblies are not formed by amyloid fibrils, either randomly or radially oriented, and cannot hence be classified as amyloid spherulites. Amyloid fibrillation is an irreversible process²²⁴ In contrast P23V and P23VC110M spherical assembly formation, which occurs at the protein pI, is reversible⁹⁶. The reflections seen by WAXS at 4.8 Å and 8.5 Å (Figure 4.17) which are suggestive of amyloid fibrillation only emerge after spherical assemblies age and are hence not related to their internal morphology.

We see instead that P23VC110M spherical assemblies form at the protein pI, similar to protein particulates. This may suggest a similar formation mechanism. However, there are still significant differences between these two types of structures, such as their very different size ranges and the ability of particulates to form gels, which does not happen for the HGD mutants. Protein particulates are likely to have an irregular/fractal-like internal morphology²²¹. For equine lysozyme, these are composed of isotropically oriented β -sheet structures (with no amyloid fibrillation involved)^{218,225} while for β -lactoglobulin, randomly oriented short stretches of amyloid-like fibrils have been proposed²²³. In any case, these structural features give rise to reflections at 4.8 Å and 8.5 Å and strong ThT binding, characteristics that we have not found for P23VC110M spherical assemblies.

Although thermal stability data for the P23V mutant is not available, experiments on the wild-type protein HGD and mutants P23T and P23S indicate that those proteins retain their native structure at the temperatures used to grow spherical assemblies^{57,235,236}; hence it is highly unlikely that the P23V structure is perturbed either. This is a further difference between HGD mutant spherical assemblies and amyloid spherulites and particulates, which are obtained either by extensive or mild protein unfolding induced by temperature.

Given the differences and similarities discussed, P23VC110M spherical assemblies cannot be classified either as amyloid-like spherulites or particulates and they are to our knowledge a new and previously unreported protein assembly type.

4.4 Conclusions

It is well-established that anisotropic protein-protein interactions can dramatically alter protein phase behaviour. The double mutant P23VC110M of HGD has been studied in detail. After successfully expressing and purifying P23VC110M, we showed that the new double mutant has no differences in secondary and tertiary structure compared to the wild-type protein. In addition, protein-protein interactions in the liquid phase are not affected by the mutations.

We observed that P23VC110M forms three solid polymorphs: spherical assemblies with an inverted solubility curve due to mutation P23V, orthorhombic crystals due to mutation C110M and needle/plate-shaped crystals; the latter are more soluble (less stable) than orthorhombic crystals formed by C110M ($1k_B T$ difference at 283K) which suggests their formation is kinetically favoured. Due to their high nucleation rate, it was not possible to obtain X-Ray diffraction-quality crystals of P23VC110M needles/plates. However, it is possible that this new polymorph nucleates due to the formation of new crystal contacts that may involve the two mutation sites. We therefore further showed that the phase diagram of double mutants can be predicted starting from the ones

of the single mutants; however, as in our case, increasing the number of mutations can increase complexity, giving rise to richer phase behaviours.

Secondly, we investigated the morphological properties of P23VC110M spherical assemblies by microscopy, SAXS, WAXS and ThT binding assays. We showed that spherical superstructures are composed of amorphous material and no evidence of amyloid fibrils has been found. Our experiments on P23VC110M also suggest that, similarly to mutant P23VR58H, crystals due to mutation C110M likely nucleate on spherical assembly surface. Although these spherical particles resemble other generic forms of protein aggregates (i.e amyloid spherulites and particulates), we show they do not fit in any of these classes, and are therefore a new, unique type of protein superstructure.

Chapter 5:

Stability of ADDomer virus-like particles

5.1 Background

Virus-like particles (VLPs) are capsid protein-based self-assembled particles that can be exploited as biopharmaceuticals (Section 1.1.2.4). Viral capsid assembly is a highly specific process, normally involving the self-assembly of individual protein units into structures with icosahedral symmetry, containing exactly 60 subunits. The assembly process can be dramatically impacted by slight variations in protein primary structure: single-point mutations can in some cases improve viral capsid stability but also partially or completely hinder its formation^{79,160}. During assembly, subunits normally assemble initially into pentons, which then further assemble into higher order units. Protein-protein interactions are important therefore both between individual subunits, driving penton assembly, but also between pentons to form dodecahedra. Given the number of specific contacts required for successful assembly and the observation that minor variations in amino acid sequence can disturb the assembly pathway, it is clear that anisotropic protein-protein interactions play an important role in the assembly process and the stability of the assembled capsid.

Surface protein amino acid modifications can be included to provide therapeutic functions: the exposure of specific epitopes on VLP surfaces can induce immune-response against viral infections (Chikunguya, Dengue, Zika, etc.) making VLP formulations valuable candidates for vaccine development²⁴. However, vaccine formulations often need to be stored at extremely low temperatures (e.g. -80 or -20°C). This makes their distribution impractical and expensive²³⁷. While these difficulties can be overcome in developed countries, the dependence on the cold chain has more drawbacks in remote and poorer areas of the world²³⁸. On the other hand, storage at milder temperatures may lead to undesired phenomena such as VLP disassembly in addition to protein unfolding and aggregation²³⁹. By examining the solution and structural thermostability of VLPs, insight into the fundamental drivers for particle assembly and the potential for stability can be assessed.

ADDomers are a new type of VLP developed by the joint efforts of the Max Planck-Bristol Centre for Minimal Biology and BrisSynBio, the Centre National de la Recherche Scientifique (CNRS, France) and Imophoron Ltd²⁴⁰. ADDomer VLPs are dodecahedral assemblies based on Human Adenovirus serotype 3 (HAdV-3) penton base protein (Pb). Penton base proteins are one of the three components of Human Adenoviruses capsids along with hexon and fiber proteins (Figure 5.1, left). HAdV-3 Pb self-assembles *in vitro* into pentamers (pentons, Figure 5.1, middle) that in turn form dodecahedral VLPs with a radius of approximately 15 nm (called either ADDomer or HAdV-3 dodecahedrons, Figure 5.1, right)²⁴⁰. Hence, each ADDomer VLP is formed by a total of 60 Pbs sub-assembled into 12 pentons²⁴¹. Similarly to other VLP vaccines, ADDomer triggers strong immune-response against a specific infection by the exposure of several copies of the

relevant pathogenic epitope on its surface; in particular, ADDomers have been designed as a scaffold to rapidly and efficiently display epitopes from a pool of human and livestock pathogens in a plug-and-play manner^{240,242}. Furthermore, given its internal cavity of approximately 8 nm, it could be also used to encapsulate small molecules for other applications in biomedicine and biotechnology²⁴⁰.

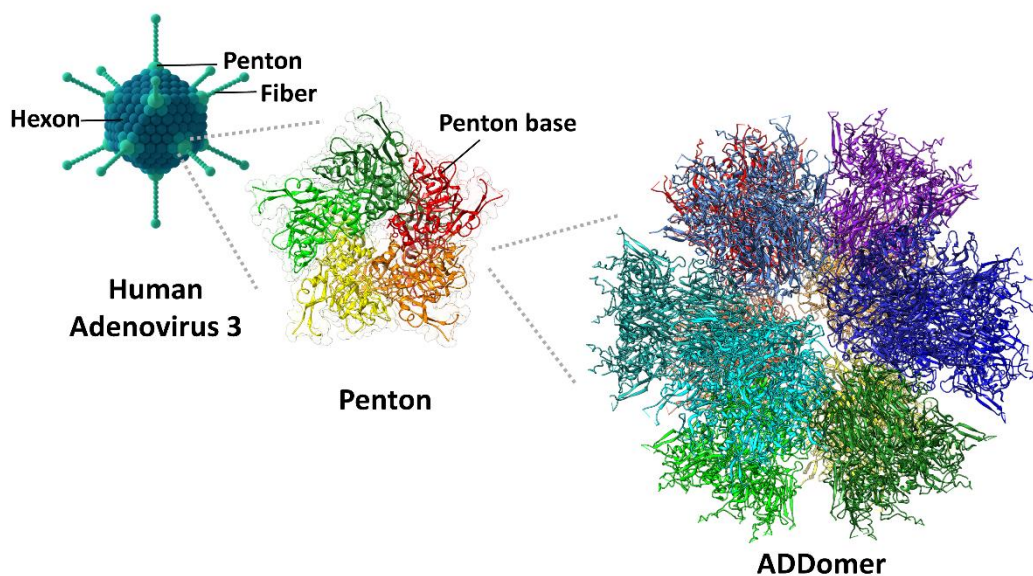


Figure 5.1: The icosahedral capsid of human adenoviruses is a self-assembled particle formed by three different types of proteins: the hexon, that self-assembles in trimers forming the triangular facets, the penton base (Pb), that self-assemble into pentamers (pentons) at the apexes, and the fiber, also present on the apexes (left picture). In the absence of hexon and fiber proteins, Pbs (shown in different colours, middle picture) still self-assemble into pentons that in turn form 12-penton dodecahedral VLPs (right picture). ADDomer is a dodecahedral VLP derived from HAdV-3pentons.

5.2 Aim of the study

In a previous study, ADDomers were shown to be a promising vaccine candidate with good thermotolerance²⁴⁰. Given these results, in collaboration with researchers in BrisSynBio and the Bristol Max Planck Centre for Minimal Biology (Georgia Balchin, Prof. Christiane Schaffitzel-Berger and Prof. Imre Berger), our aim was to test a range of protein surface modifications of ADDomer particles to understand what influences VLP stability and thereby further improve the properties of ADDomer VLPs. We employed Dynamic Light Scattering (DLS) to assess temperature induced changes in particle size (e.g. from aggregation or disaggregation) and Intrinsic Fluorescence Emission Spectroscopy and Far-UV Circular Dichroism Spectroscopy to address to what extent those changes were related to structural changes within subunit proteins.

In a further step, we compared the thermal resistance of ADDomer particles to a similar construct, but derived from Chimpanzee Adenovirus serotype-3 Pb (ChAdV-3), named ChADDomer. Since thermal stress can also induce VLP disassembly into free pentons, ChADDomer mutants L56C

and S57C (leucine and serine to cysteine, respectively) were designed to include the formation of inter-penton disulphide bridges to stabilize the VLP higher order assembly. Furthermore, a chimeric Pb, namely Chimera, was engineered using parts of the primary structures of ChAdv-3 and HAdV-3 Pbs. The wild-type Chimera and its mutant S57C (analogous to mutant S57C of ChADDomer) were also investigated.

5.3 Results and discussion

5.3.1 Comparison of the structural features of ADDomer, ChADDomer and Chimera penton base proteins

Design, expression, purification and Cryo Electron Microscopy (cryo EM) structures of all the VLPs were performed by our collaborators in the Max Planck-Bristol Centre for Minimal Biology and BrisSynBio (Georgia Balchin, Prof. Christiane Berger-Schaffitzel and Prof. Imre Berger). We performed Dynamic Light Scattering, Circular Dichroism, Intrinsic Fluorescence measurements and primary structure comparison on particles that were supplied by them as part of this collaboration.

ADDomer Pb is a 60kDa protein with an elongated shape divided into two domains, the jellyroll, containing mostly β -sheets, and the crown, formed by α -helices and random coil regions. Between the two domains there is a tetradentate site composed of four sulphur-containing residues which coordinate to metal ions, which may stabilize the protein fold (Figure 5.2A)²⁴³. When Pbs are assembled into VLPs, the jellyroll is oriented towards the inside of the particle while the crown is exposed to the solvent. The crown contains two variable regions that show high variability in length and primary sequence among adenovirus serotypes; they are known as RGD (i.e. contains the Arginine-Glycine-Aspartic Acid motif) and variable loop (VL) (Figure 5.2A). These regions can be exploited to display up to three different epitopes on the ADDomer surface²⁴². The electron density of those regions is not present in the cryo EM and X-Ray Diffraction structures of ADDomers, most likely due to high flexibility or proteolysis. Similarly, the electron density of the region 1-47 (protein N-terminus) is not available; although this portion of the protein is part of the jellyroll and should therefore point towards the VLP cavity, it has been shown to protrude outward. It was also suggested that residues 1-47 are indirectly involved in strand swapping between N-terminal domains of Pbs from adjacent pentons, with the site of strand swapping being between residues 58-61 (⁵⁸SELS⁶¹) (Figure 5.2B, black vertical line). Unlike HAdV-3 Pbs, HAdV-2 Pbs self-assemble into full dodecahedra only in specific conditions where strand swapping is not observed. It is therefore suggested that, in physiological conditions, strand swapping may be an important stabilizing factor allowing HAdV-3 dodecahedra to assemble (Figure 5.2C)²⁴³.

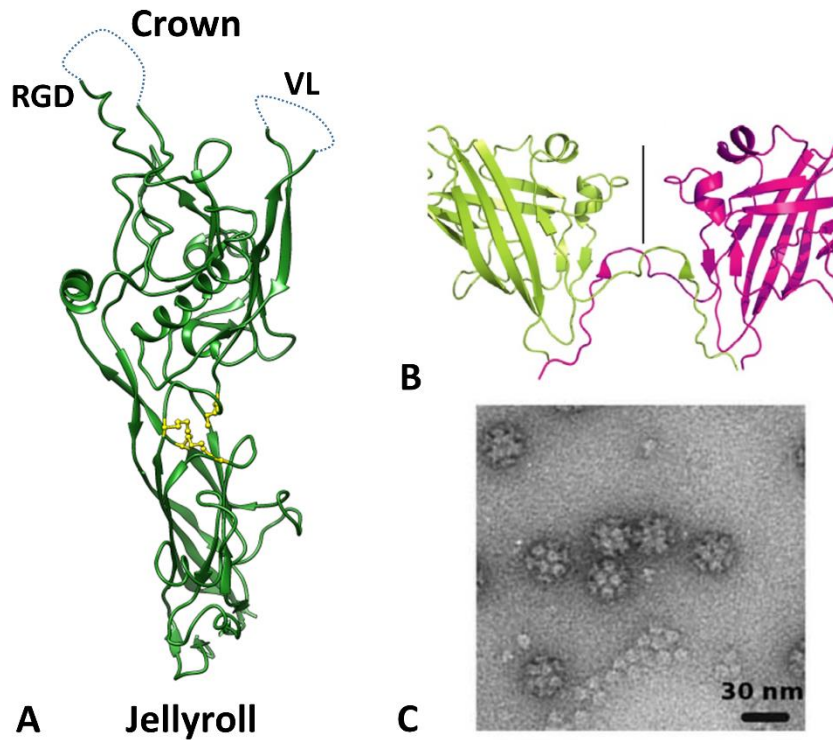


Figure 5.2: A) Lateral view of ADDomer penton base protein structure as obtained by Cryo-EM (PDB ID 6HCR): the protein is divided into the jellyroll (bottom) and the crown domains (top); residues 1-47 are not shown due to the lack of electron density in this region. The variable regions RGD and VL are shown in blue dotted lines while the four sulphur-containing residues between the two domains are shown in yellow. B) Illustration of strand swapping between two penton base proteins belonging to different pentons taken from ref. 243; the approximate position of the ⁵⁸SELS⁶¹ is highlighted by the black vertical line. C) Fully formed ADDomer VLPs (HAdV-3 dodechedra) seen by negative-stain EM as shown in ref. 240; the micrograph also shows the presence of non-dodecahedral structures which may result from aggregation or Cryo-EM sample preparation.

ChAdV-3 and HAdV-3 Pbs have 83.5% sequence identity and share most of the structural features described above. ChAdV-3 Pbs spontaneously self-assemble into dodecahedral VLPs (unpublished results, personal communication). Although there is no experimental confirmation of strand swapping in ChADDomer, the SELS region (⁵⁴SELS⁵⁷) is located at the interface between two pentons in a fully assembled VLP. To further stabilise the assembly, cysteine mutations were made to facilitate inter-penton disulphide bridges (Figure 5.3). Two mutants, L56C and S57C have been created and shown to form dodecahedral VLPs (unpublished results, personal communication).

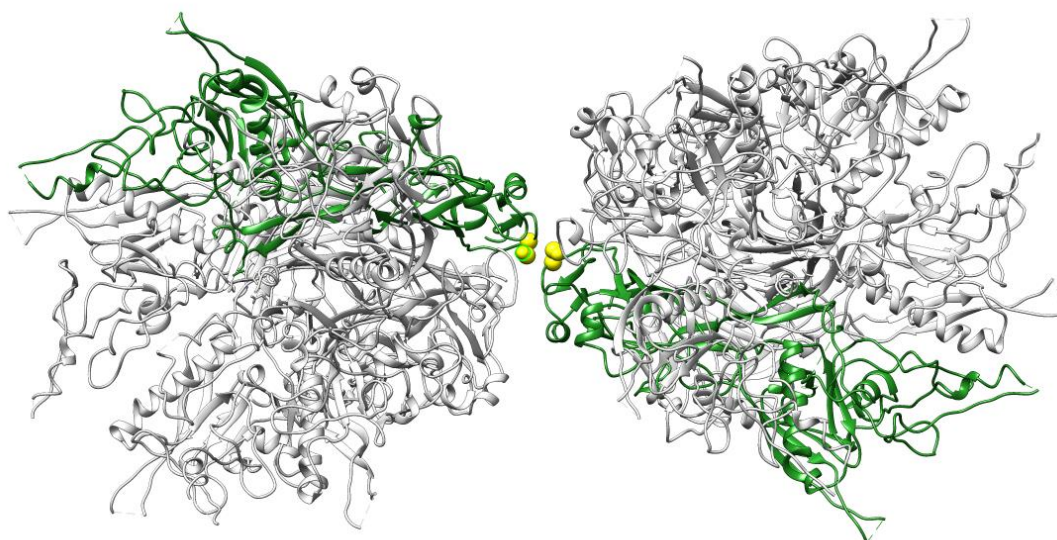


Figure 5.3: Cryo EM structure of ChADDDomer provided by our collaborators at Max Planck-Bristol Centre for Minimal Biology and BrisSynBio; only two pentons are shown. Cysteine mutations (S57C, shown in yellow) can be introduced in the SELS region to link penton base proteins belonging to different pentons (green) by disulphide bridge formation.

Finally, to test if either of the two protein domains (jellyroll and crown) have any discernible impact on VLP disassembly or aggregation propensity, a chimeric construct composed of HAdV-3 crown and ChAdV-3 jellyroll have been created along with its disulphide bridge-forming mutant S57C. In contrast to the wild-type proteins, all the cysteine variants have been shown to form disulphide bridges by SDS-PAGE (unpublished result, personal communication) even though it was not possible to quantify their number per VLP by this technique.

5.3.2 Preliminary characterisation of VLPs by DLS

The successful formation of VLPs for all the proteins of interest was verified by our collaborators with negative-stain EM. However, this technique is time-consuming and does not necessarily reflect the native solution environment²⁴⁴. DLS is a powerful technique for particle size determination, particularly for samples with low polydispersity, that is complementary to EM, but also allows solution behaviour to be assessed. In addition, since in DLS the scattering intensity increases as the 6th power of the particle radius, it is exceptionally sensitive to even low levels of particle aggregation. We initially measured the ChADDDomer with the L56C mutation in two different buffers, 50mM Tris, pH 7.5 with and without 150mM NaCl (Figure 5.4A and B).

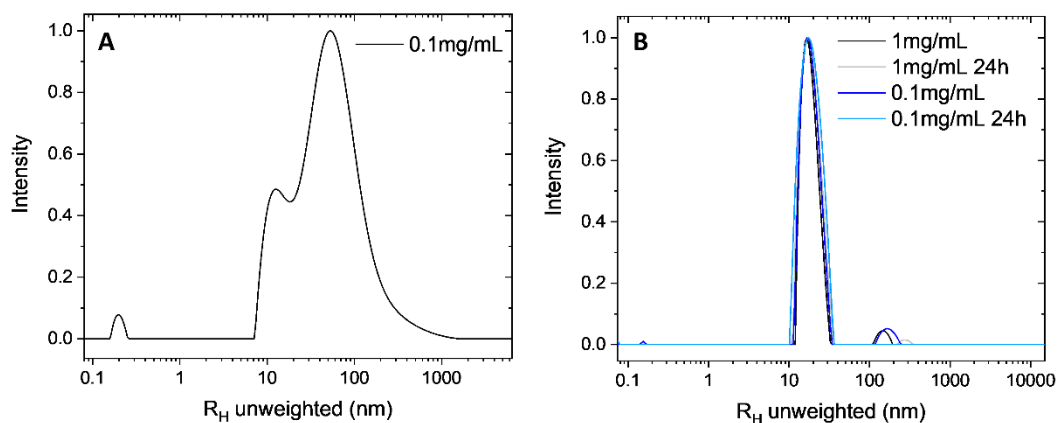


Figure 5.4: Size distribution of ChADDomer L56C VLPs measured by DLS in A) 50mM Tris pH 7.5 + 150mM NaCl and in B) 50mM Tris pH 7.5. The samples at low ionic strength were also monitored after 24 hours to assess longer-term stability against aggregation.

Either in the presence or absence of 150 mM NaCl, the DLS size distribution shows a population of particles with a hydrodynamic radius of approximately 15-18 nm, in agreement with the VLP particle size measured by negative-stain EM. However, in the presence of 150mM NaCl, VLPs aggregate, forming particle sizes over a large range of sizes, up to 1 μ m. Hence, the presence of 150 mM NaCl, which is usually used to replicate physiological conditions, is not necessary for the successful assembly of this type of VLP and instead leads to VLP aggregation. Low ionic strength samples showed only a minimal amount of aggregation and, when kept at room temperature for one day, maintained the same particle size, with no evidence of further aggregation.

5.3.3 Temperature stability of ChADDomer VLPs probed by DLS

Given the sensitivity of the VLPs to higher ionic strengths, all the experiments were carried out in 50 mM Tris, pH 7.5. DLS experiments on wild-type (WT) ChADDomer with increasing temperature indicate that the VLPs are stable against temperature induced aggregation until approximately 40 °C (Figure 5.5). From 20 – 40 °C there is no significant change in either the particle hydrodynamic radius or the polydispersity index.

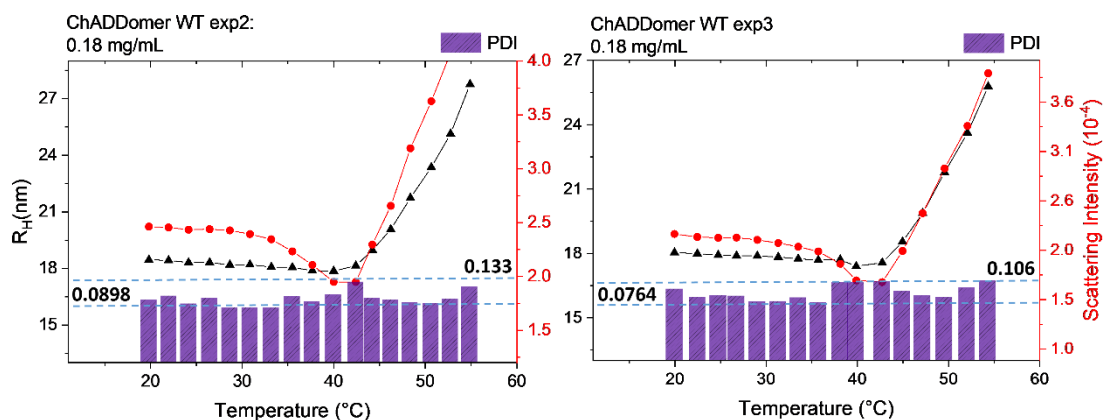


Figure 5.5: Hydrodynamic radius (R_H), scattering intensity and polydispersity index (PDI, for which the minimum and maximum values are indicated in bold and by the light blue dashed lines) as a function of temperature for ChADDomer wild-type from two independent experiments.

Interestingly, the scattering intensity of the samples starts decreasing a few degrees before an increase in hydrodynamic radius is observed. Here, the scattering intensity is the ratio between the detector count rate and the intensity of the incident beam recorded for each datapoint. Decreases in light scattering intensity have been linked to protein unfolding²⁴⁵ or, in the case of viral assembly, to particle swelling^{246,247}. We have recorded the onset temperature for this scattering intensity decrease as it may relate to the VLP temperature instability. The temperature at which the scattering intensity trend changes slope is defined as T_{SID} and for the ChADDomer WT is $33.4 \pm 0.1^\circ\text{C}$. The VLPs start to aggregate at higher temperatures, indicated by the concomitant increase of hydrodynamic radius and scattering intensity. The temperature at which the scattering intensity starts to increase again along with the hydrodynamic radius is defined as the temperature of aggregation onset, T_{AGG} , which for ChADDomer WT is $44.5 \pm 0.5^\circ\text{C}$.

ChADDomer mutant L56C was analysed with the same method and the results are shown in Figure 5.6.

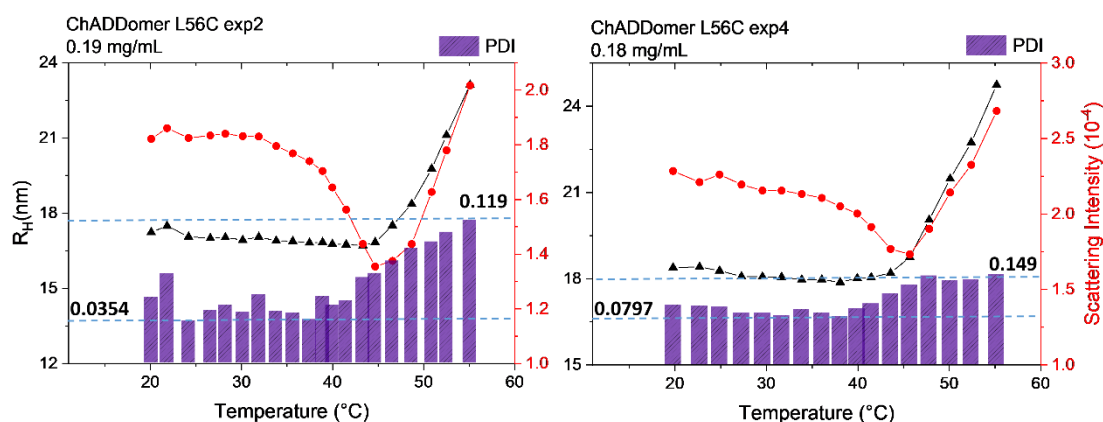


Figure 5.6: Hydrodynamic radius (R_H), scattering intensity and polydispersity index (PDI, for which the minimum and maximum values are indicated in bold and by the light blue dashed lines) as a function of temperature for ChADDomer L56C from two independent experiments.

For ChADDomer L56C the T_{SID} is 38.5 ± 0.4 °C while T_{AGG} is 47.2 ± 1.2 °C. Hence, mutation L56C has the effect of increasing both T_{SID} and T_{AGG} compared to the wild-type protein. Furthermore, the increase in hydrodynamic radius, R_H for the L56C mutant, is less than for the wild-type protein.

The DLS results for ChADDomer S57C are shown in Figure 5.7.

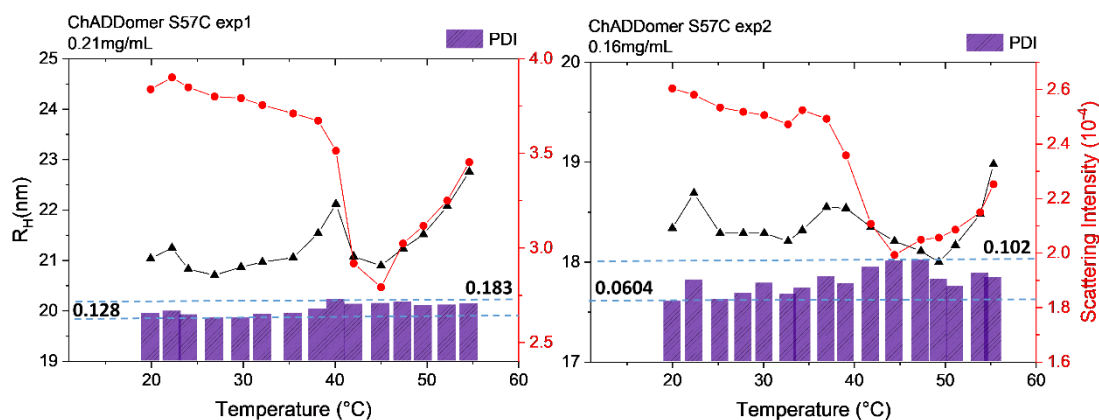


Figure 5.7: Hydrodynamic radius (R_H), scattering intensity and polydispersity index (PDI, for which the minimum and maximum values are indicated in bold and by the light blue dashed lines) as a function of temperature for ChADDomer S57C from two independent experiments.

The higher R_H observed in the first experiment is accompanied by a higher initial polydispersity index, which reflects poor sample quality. However, as we will show later, this does not affect the scattering intensity and hydrodynamic radius trends, which remain almost identical for the two experiments. The DLS experiments on ChADDomer S57C show that for this mutant T_{SID} is 37.5 ± 0.7 °C while T_{AGG} is 49 ± 2 °C. The onset temperatures for those two phenomena are very similar to those for the mutant L56C and higher than for the wild-type protein. In addition, the mutant S57C seems to be the least prone to aggregation which is reflected by the very small increase in hydrodynamic radius as the temperature is increased.

5.3.4 Temperature stability of ADDomer and Chimera VLPs probed by DLS

We also probed the thermostability of ADDomer, Chimera WT and Chimera S57C VLPs by DLS. The results for ADDomer are shown in Figure 5.8.

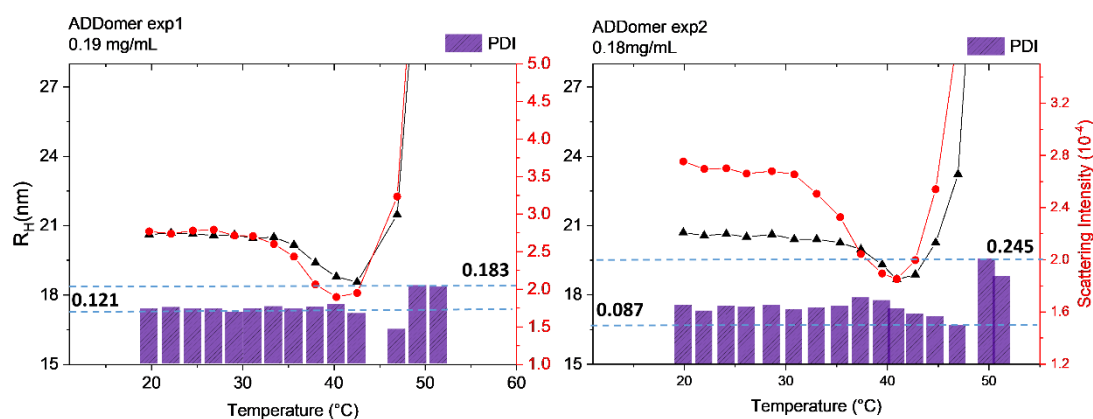


Figure 5.8: Hydrodynamic radius (R_H), scattering intensity and polydispersity index (PDI, for which the minimum and maximum values are indicated in bold and by the light blue dashed lines) as a function of temperature for ADDomer from two independent experiments.

ADDomer VLPs aggregate abruptly, reaching R_H values of approximately 60 nm at 51°C (data not shown); this phenomenon was not observed for ChADDomer VLPs. For ADDomer, T_{SID} is 32.2 ± 1.3 °C while T_{AGG} is 44.7 °C (evaluated only from experiment 2 given the low number of datapoints at about 45 °C in exp1, where R_H and the scattering intensity most likely to re-increase). These onset temperatures are similar to those obtained for ChADDomer WT. Furthermore, ADDomer VLPs undergo a substantial decrease in R_H , at T_{SID} prior to aggregation.

Chimera WT thermostability was also assessed by DLS, the results are shown in Figure 5.9:

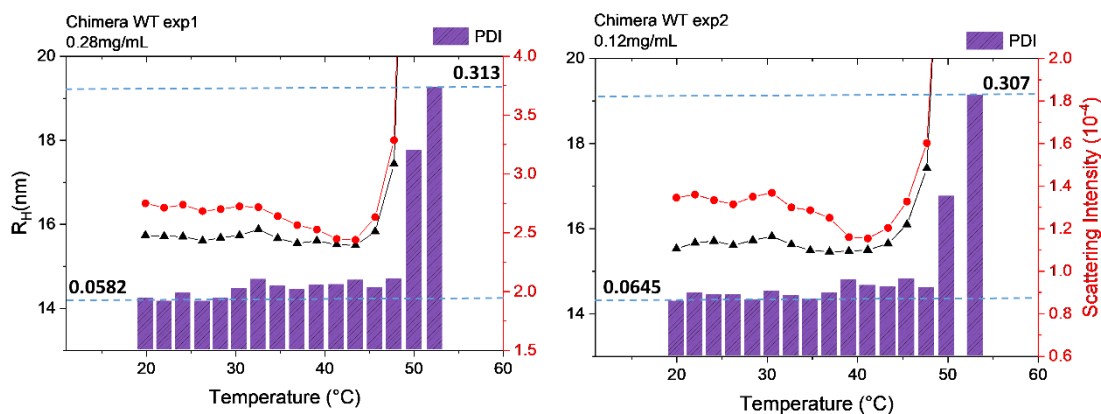


Figure 5.9: Hydrodynamic radius (R_H), scattering intensity and polydispersity index (PDI, for which the minimum and maximum values are indicated in bold and by the light blue dashed lines) as a function of temperature for Chimera WT from two independent experiments.

Chimera WT shows a small decrease in scattering intensity as the temperature is increased. Its aggregation behaviour is very similar to that observed for ADDomer, showing abrupt aggregation once T_{AGG} is reached. The hydrodynamic radius reaches approximately 130 nm at 53 °C (data not shown). T_{SID} is 33.7 ± 1.1 °C and T_{AGG} is 45.5 ± 0.2 °C; similar T_{SID} and T_{AGG} values were obtained for ChADDomer WT and ADDomer.

Finally, the aggregation behaviour of Chimera S57C was also assessed (Figure 5.10).

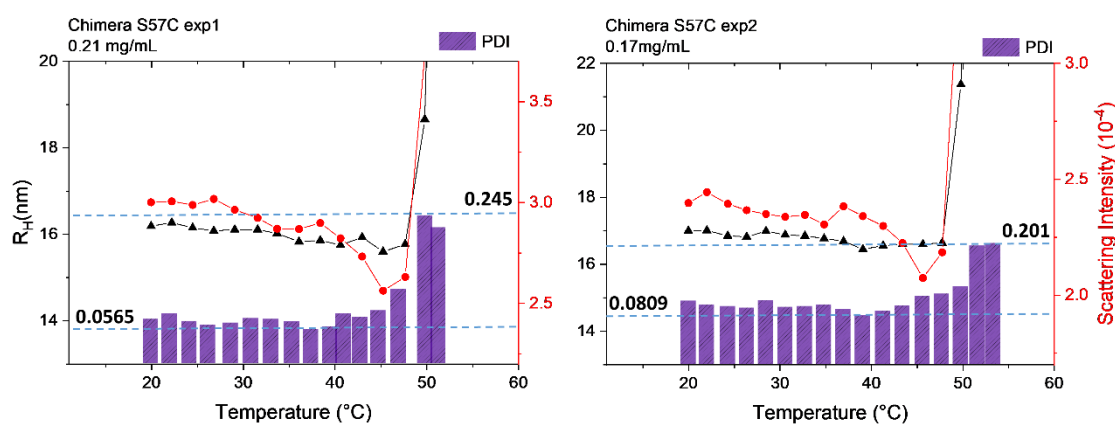


Figure 5.10: Hydrodynamic radius (R_H), scattering intensity and polydispersity index (PDI, for which the minimum and maximum values are indicated in bold and by the light blue dashed lines) as a function of temperature for Chimera S57C from two independent experiments.

Chimera S57C aggregates abruptly, similarly to ADDomer and Chimera WT reaching a R_H of approximately 85 nm at 54 °C (data not shown). T_{SID} and T_{AGG} are 37.7 ± 0.7 and 49.8 °C respectively, similar to ChADDomer L56C and S57C.

5.3.5 Comparison of VLPs thermal stabilities

To better address the differences observed for each type of VLP, all the DLS results are shown in Figure 5.11. It is clear that the experiments were remarkably reproducible and show near-quantitative consistency in independent experiments. The initial average hydrodynamic radius of the VLPs in solution at room temperature in each experiment is slightly different depending on small variations in concentration and on the polydispersity of the samples (sections 5.3.3 and 5.3.4); Figure 5.11 shows that this has no significant effect on the trends observed and hence, for comparison, we plot the difference between the initial R_H and the R_H measured at each temperature T , ΔR_H , which is shown in figure 5.11A.

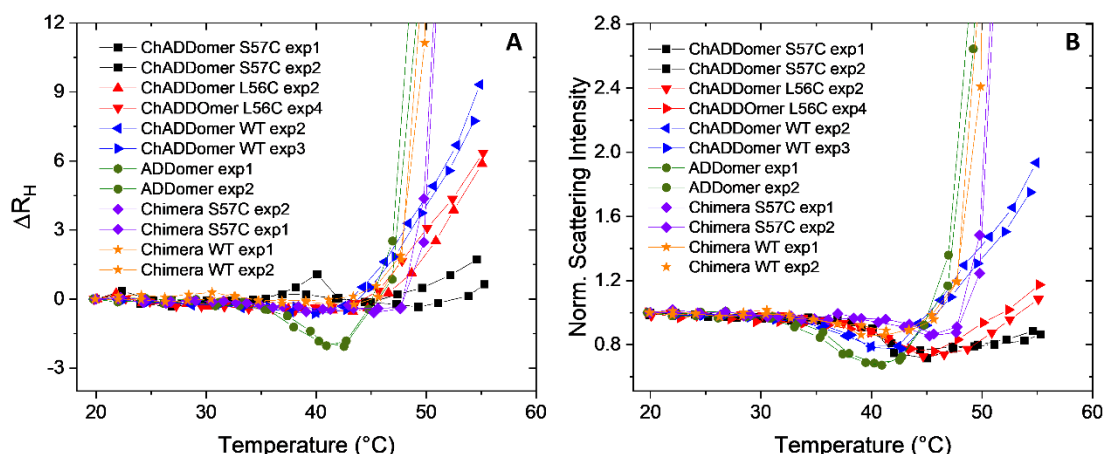


Figure 5.11: Comparison of all VLP experiments showing A) the hydrodynamic radii shift between the end and the beginning of the experiment and B) the normalized scattering intensity.

As shown in Figure 5.11A, ADDomer is the only VLP that has a significant decrease in hydrodynamic radius prior to aggregation. This may be due to partial disassembly of the dodecahedron into pentons. However, regularized fits of the correlation functions did not reliably show peaks corresponding on particles with lower R_H values for any of the VLPs investigated. Given that the majority of the sample is composed of larger dodecamers, DLS may not resolve a small numbers of pentons if partial disassembly occurs. In contrast, negative stain EM at 50°C indicates the presence of pentons for all the VLP types (unpublished results, personal communication). This suggests that increasing temperature may trigger some disassembly of the penton for ADDomer VLPs, but not in sufficient quantity to alter either the average hydrodynamic radius or the polydispersity index. It may also be that the pentons observed in negative stain EM are artifacts of the sample preparation. Future static light scattering studies might address this issue: the R_H of the VLPs is approximately 15-20 nm, more than one order of magnitude larger than the inverse of the scattering vector (q^{-1}) making these VLPs Rayleigh scatterers⁸⁹; furthermore, the product of the estimated radius of gyration and the scattering vector, $R_g \cdot q$, is $\ll 1.3$; these conditions would allow the average molecular weight to be estimated as a function of temperature, which might help in understanding if disassembly is involved and, if the process is reversible at these temperatures, in estimating the chemical potential of dodecamers formation/disassembly.

ADDomer and Chimera VLPs undergo a steep increase in hydrodynamic radius as the onset temperature of aggregation is reached; the same is not observed for ChADDomer VLPs, which aggregate to a lesser extent. This behaviour may be due to different aggregation kinetics and or mechanisms. To compare the extent of this phenomenon among the VLP types, we normalized the R_H at 51 °C, the temperature at which all the VLP types have aggregated to some extent, by the R_H at 20 °C; the ratios $R_H(51\text{ °C})/R_H(20\text{ °C})$ are shown in Table 5.1 and provide a qualitative estimation of the extent of aggregation for each VLP type. Interestingly, ADDomer and Chimera

VLPs, which abruptly aggregate, are composed of the same crown domain that is exposed on the VLP surface and may be the source of such different aggregation behaviour compared to ChADDomer VLPs.

DLS results show that both mutants of ChADDomer, L56C and S57C, aggregate less than the wild-type, with S57C being the least prone to aggregation. ChADDomer WT, Chimera WT and ADDomer have similar T_{AGG} at approximately 45 °C (Table 5.1). However, mutations L56C and S57C in ChADDomer increase T_{AGG} compared to the wild-type protein; the same effect was observed for the S57C mutant of Chimera VLPs.

The intensity of the scattered light is proportional to particle density. Hence, protein unfolding or virus swelling are usually reflected in increasing particle size with a concomitant decrease in the scattering intensity^{246,247}. While we observe the intensity of light scattered decreasing for all the VLPs probed at T_{SID} , this is not concomitant with an increase in hydrodynamic radius (Figure 5.11 B). However, a simultaneous swelling of particles and partial disassembly may result in no net-change in average hydrodynamic radius and this possibility cannot be excluded. Partial disassembly alone could also decrease the intensity of scattered light, but we would expect this to be accompanied by a decrease in R_H , which is not observed. A second possible explanation would be partial protein unfolding of protein subunits within the VLP, which we would not expect to have a significant impact on the average R_H . The temperature at which the scattering intensity starts decreasing for each VLP (T_{SID}) and the aggregation onset temperature (T_{AGG}) are shown in Figure 5.11B and Table 5.1 respectively.

Table 5.1: Comparison of the onset temperatures of aggregation and scattering intensity drop for all the VLPs.

VLP	T_{AGG} (°C)	T_{SID} (°C)	R_H (51 °C)/ R_H (20 °C)
Chimera S57C	49.8	37.7 ± 0.7	1.85
ChADDomer S57C	49 ± 2	37.5 ± 0.7	1.04
ChADDomer L56C	47.2 ± 1.2	38.5 ± 0.4	1.17
Chimera WT	45.5 ± 0.2	33.7 ± 1.1	4.39
ADDomer	44.7	32.2 ± 1.3	2.86
ChADDomer WT	44.5 ± 0.5	33.4 ± 0.1	1.27

As the ranking indicates, T_{SID} is lowest for ADDomer and highest for the ChADDomer mutants. If we examine the ChADDomer particles more carefully, while the T_{AGG} and T_{SID} are lower for the WT protein than for both mutants, all three indicate a decrease in scattering intensity before

T_{AGG} . The mutant proteins are designed to increase the strength of the inter-penton interactions, and this therefore suggests that it is unlikely that disassembly of the VLP is responsible for this observation. It is more likely that a global change in the VLP particle occurs.

5.3.6 Thermal stability of VLPs – structural aspects

Protein aggregation is often related to protein unfolding. To probe if this process contributes to either the decreases in scattering intensity observed before T_{AGG} or VLP aggregation, we employed intrinsic fluorescence emission spectroscopy to examine protein structural changes with increasing temperature (Figure 5.12).

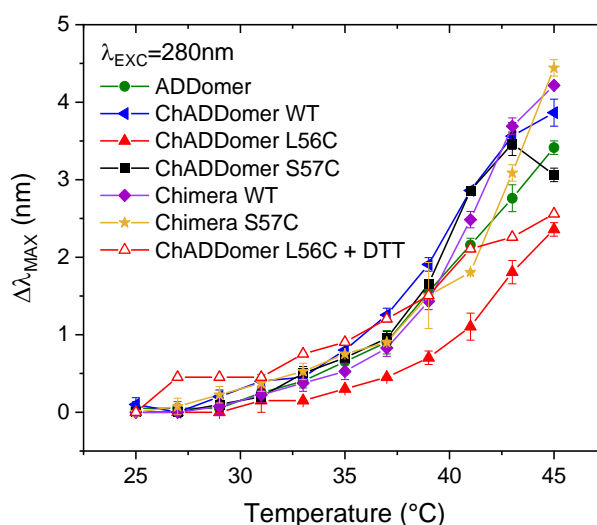


Figure 5.12: VLP temperature ramps followed by Intrinsic Fluorescence emission spectroscopy with excitation wavelength = 280nm. The wavelength shift is the difference between the maximum emission wavelength at each temperature T and maximum emission at 25 °C.

To compare all VLPs, we define the fluorescence emission wavelength shift ($\Delta\lambda_{MAX}$) as the difference between the wavelength of maximum emission (λ_{MAX}) at each temperature T and λ_{MAX} at $T = 25$ °C. The λ_{MAX} for all VLPs is $338 \text{ nm} \pm 0.8 \text{ nm}$ at 25°C indicating that all the different protein subunits have a similar fold. For all VLPs, the λ_{max} is red-shifted as the temperature increases, which indicates a higher exposure of aromatic residues to the polar solvent. Our fluorescence experiments are limited by the temperature range of our fluorimeter and no plateau is reached in the data. Also, the magnitude of $\Delta\lambda_{MAX}$ for complete protein unfolding is usually in the region of 10-15 nm with the lowest values at about 5 nm, depending on the protein^{248,249,250}. Hence, in the temperature range we explored by fluorescence, it seems unlikely that significant protein unfolding occurs and the higher exposure of aromatic amino acids is more likely only related to minor conformational changes, which may also have caused the scattering intensity drops observed by DLS. Furthermore, swelling or VLP disassembly may possibly contribute to such subtle shifts but no definitive proof for either has been found. $\Delta\lambda_{MAX}$ is lower for

ChADDomer L56C compared to other VLPs (Figure 5.12, full red triangles); hence, to address whether the formation of inter-penton disulphide bridges is responsible for this difference, we carried out the same experiment on ChADDomer L56C supplementing the sample with 0.02% DTT as a reducing agent (Figure 5.12, empty red triangles). In this case, $\Delta\lambda_{\text{MAX}}$ is similar to the wild-type protein, suggesting that the formation of disulphide bridges between pentons by mutation L56C helps to suppress the slight conformational changes induced by temperature. Although a similar behaviour was expected for ChADDomer S57C, $\Delta\lambda_{\text{MAX}}$ is similar to the wild-type protein.; Although the reason remains unclear, this could be due to a different number of inter-pentons disulphide bridges per VLP between mutants L56C and S57C.

Fluorescence spectroscopy only probes the solvent exposure of aromatic residues; to assess if the protein backbone undergoes structural changes in the temperature range studied, we use Circular Dichroism spectroscopy in the absorption region of the peptide bond (Far-UV, 190-260 nm, Figure 5.13).

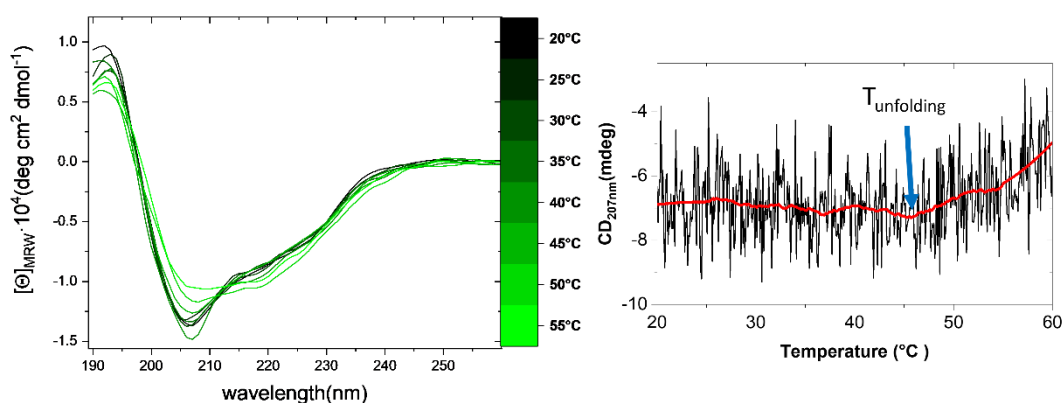


Figure 5.13: A) Far-UV CD spectra of ChADDomer WT at different temperatures showing minimal structural changes. The spectra are colour-coded for different temperatures, for which the legend is shown on the right; B) representative temperature ramp of a ChADDomer WT sample following CD at 207 nm, the onset temperature of unfolding is indicated by a blue arrow.

The spectra at different temperatures for the ChADDomer WT VLPs are shown in Figure 5.13 A. The two negative absorption bands at approximately 207 and 218 nm indicate the protein is composed of a mixture of α -helices and β -sheets²⁵¹, in agreement with the structural information available from cryo EM. As the temperature is increased, the changes in the spectrum are minimal, suggesting that in the temperature range explored, the protein is not fully unfolded. To better capture the onset of any conformational changes we performed temperature ramps monitoring the signal at 207 nm (Figure 5.13 B). We define the onset temperature of unfolding, T_{unfold} , as the temperature at which the CD signal starts to increase significantly; for ChADDomer WT this is 45.8 ± 0.4 °C. The same experiments were performed on ADDomer, ChADDomer L56C and

ChADDomer S57C indicating similar behaviour for all VLP types; the T_{unfold} obtained by CD temperature ramps are compared to T_{AGG} obtained by DLS in Table 5.2.

Table 5.2: Comparison of T_{unfold} and T_{AGG} as obtained by CD spectroscopy and DLS respectively.

VLP	T_{AGG} ($^{\circ}C$)	T_{unfold} ($^{\circ}C$)
ChADDomer S57C	49 ± 2	50.3 ± 0.3
ChADDomer L56C	47.2 ± 1.2	53.4 ± 0.5
ADDomer	44.7	48.6 ± 0.1
ChADDomer WT	44.5 ± 0.5	45.8 ± 0.4

If protein aggregation is induced by subunit unfolding, we would expect a correlation between T_{AGG} and T_{unfold} . However, as shown in Table 5.2, we found no correlation between those two temperatures; however, for ChADDomer L56C and S57C the T_{unfold} is higher compared to the wild-type protein confirming DLS results; this shows that the formation of inter-penton disulphide bridges contributes to stabilising protein secondary structure and delaying aggregation. The CD results also confirm that the maximum emission wavelength shifts observed in the fluorescence experiments (Figure 5.12) are related to only minor structural changes, while the onset of peptide backbone unfolding occurs at temperatures higher than T_{AGG} .

5.3.7 Comparison of the primary structure of ChADDomer, ADDomer and Chimera penton base proteins

DLS experiments indicate that ADDomer and Chimera VLPs undergo rapid aggregation when compared to the more stable ChADDomer WT, L56C and S57C VLPs; however, we have shown that ADDomer and ChADDomer WT T_{unfold} are very similar. Therefore, different aggregation behaviour may be due to intrinsic structural differences between ChADDomer and ADDomer/Chimera. Interestingly, we noted that Chimera VLPs are composed of the ADDomer crown domain and ChADDomer jellyroll (Section 5.3.1). We speculate that the ADDomer crown domain, which is the most solvent exposed, is responsible for the abrupt aggregation observed by DLS. A preliminary visual inspection of the electrostatic and hydrophobic surfaces of the proteins does not reveal significant differences (e.g. specific hydrophobic patches or regions with very different charge). Therefore, to better identify the protein regions responsible for this behaviour, we performed primary structure alignment of ChADDomer, ADDomer and Chimera penton base protein primary structures (Figure 5.14).

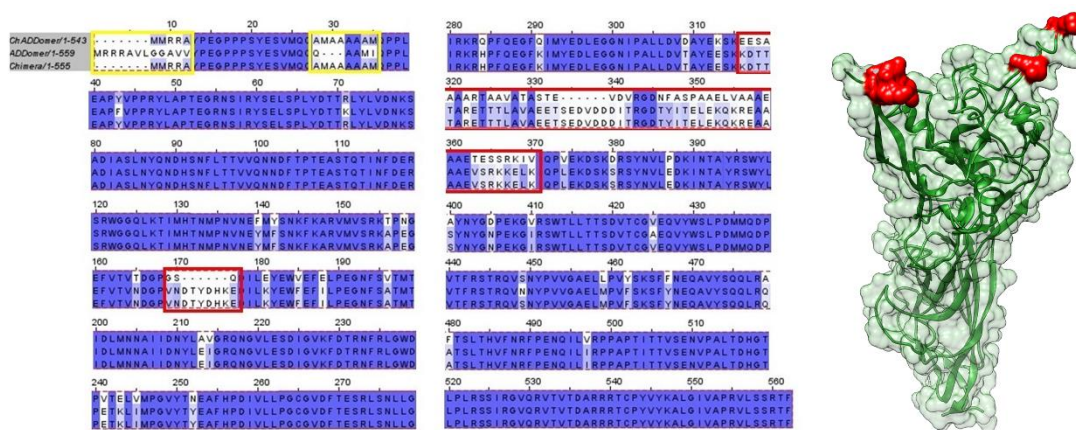


Figure 5.14: A) Primary structure alignment of ChADDomer, ADDomer and Chimera penton base proteins performed with the software LALNVIEW²⁵². Residues common to all three sequences are shown in dark purple while residues common to only two sequences are shown in light purple. The regions shared by ChADDomer and Chimera but different in ADDomer (in the jellyroll domain) are highlighted in yellow, while the regions shared by ADDomer and Chimera but different in ChADDomer (in the crown domain) are highlighted in red; B) surface representation of ADDomer penton base protein; the red surface blobs correspond to the red-highlighted primary sequence regions as shown in A).

From the primary structure alignment, we see that the stretches 1-12 and 28-35 of the consensus sequence (Figure 5.14 A, yellow) are shared by Chimera and ChADDomer but differ in ADDomer. In contrast, the stretches 169-177 and 316-370 of the consensus sequence are common to Chimera and ADDomer but different in ChADDomer (Figure 5.14 B, red); those latter stretches correspond to the VL and RGD regions (Section 5.3.1) which are situated on the top surface of a fully assembled VLPs. It may then be that the VL and RGD regions common to ADDomer and Chimera are responsible for the abrupt aggregation behaviour observed by DLS. The hydrophobicity of these regions, along with the theoretical isoelectric point of the whole protein, is compared between ChADDomer, ADDomer and Chimera penton base proteins in Table 5.3.

Table 5.3: Hydrophobicity score for regions 169-177 and 316-370 of the consensus sequence for ADDomer, ChADDomer and Chimera. The hydrophobicity score is normalized by the number of amino acids of the stretch and it was obtained by the Kyte-Doolittle scale²⁵³. The theoretical pI of the three proteins obtained by SWISSPROT is also shown.

VLP	Hydrophobicity Score		Theoretical pI
	169-177	316-370	
ADDomer	-1.89	-1.13	5.32
Chimera	-1.89	-1.13	5.27
ChADDomer	-1.57	-0.0061	5.07

Unexpectedly, the hydrophobicity score per number of amino acids indicates that regions 169-177 and 316-370 are more hydrophobic in ChADDomer than ADDomer/Chimera; hence, the reason for abrupt aggregation in ADDomer and Chimera is unlikely to be due to higher surface hydrophobicity alone and remains unclear. However, we note that the theoretical pI of ChADDomer is lower, suggesting that it may be more charged than ADDomer and Chimera at pH 7.5; hence, we speculate that stronger electrostatic repulsion due to the lower pI of ChADDomer may balance the higher hydrophobicity making ChADDomer VLPs less prone to aggregation.

5.4 Conclusions

VLPs are employed as vaccines; however, they are sensitive to thermal stress that can induce particle disassembly, protein unfolding and aggregation, which may hinder vaccine safety and efficacy. We have studied temperature induced changes on ADDomer, ChADDomer and Chimera VLPs and selected mutations promoting internal disulphide bridges in ChADDomer (S57C, L56C) and Chimera (S57C).

DLS indicates that ADDomer has a small, but significant decrease in hydrodynamic radius (~2 nanometres) with increasing temperature, possibly due to partial disassembly of the particle into pentons, prior to further aggregation at higher temperatures. Neither ChADDomer or Chimera VLPs showed such behaviour. ChADDomer, ADDomer and Chimera VLPs have a similar temperature of aggregation onset (T_{AGG}), but, disulphide bridge-promoting mutations increase the VLP thermostability to aggregation. DLS also shows that, prior to aggregation, the light scattering intensity decreases with increasing temperature. While this is often related to particle swelling or protein unfolding, in our case the reasons for this phenomenon are unclear. Although this phenomenon happens at higher temperatures for VLPs with inter-penton disulphide bridges, it occurs for all the VLPs tested suggesting that it cannot be explained by VLP disassembly alone.

Fluorescence emission spectra undergo a red-shift with increasing temperature, for all the VLP types. However, the shift is subtle compared to those expected for protein unfolding and hence, only minor conformational changes are involved, at least up to 45 °C. CD spectroscopy confirms this observation, and shows that protein unfolding onset at higher temperatures (T_{unfold}); T_{AGG} and T_{unfold} obtained by CD, are consistently higher for disulfide bridge-containing mutants, showing that the mutations help stabilize protein secondary structure.

In contrast to ChADDomer, ADDomer and Chimera VLPs undergo abrupt aggregation which is not related to major differences in thermal unfolding or aromatic residue exposure between VLPs as demonstrated by CD and Fluorescence respectively. Instead, we speculate that abrupt

aggregation is promoted by specific surface exposed regions of the primary sequence common only to ADDomer and Chimera VLPs (epitopes displaying regions, VL and RGD). Our analysis shows that hydrophobicity alone is unlikely to be the driving force of this aggregation behaviour, for which the cause remains unclear.

In conclusion, ChADDomer S57C and L56C are the most thermally stable VLPs in terms of aggregation (DLS) and have a stabilizing effect on protein secondary structure (CD).

Summary and final conclusions

This work focuses on exploring how surface modifications can impact protein phase behaviour. Given the high degree of chemical anisotropy on a protein surface, it is well known that even subtle changes can significantly influence protein self-assembly. In the first two results chapters of this thesis, the topic is investigated exploiting Human γ D-crystallin (HGD) as a model, since several mutants of this protein have been studied previously due to their role in the onset of juvenile cataract. Secondly, the effects of protein surface modifications on the thermal stability of Virus-like particles (VLPs) are probed for their potential usage as vaccine platforms.

Cys-110 of HGD is responsible for covalent dimerization of the protein which contributes to age-related cataract; however, it is unclear if it provides any biological advantage. We hence studied the phase behaviour of mutants C110M and C110S, replacing Cys-110 with two similar but non-oxidizable amino acids. Similarly to HGD, C110S is recalcitrant to crystallization, a property that maintains eye lens function during organism aging; in contrast C110M readily forms orthorhombic crystals maintaining the same properties and crystal contacts as HGD crystals. We speculate that the differences in nucleation rates observed are due to higher hydrophobicity and/or water molecule structuring around the mutation site. Therefore, Cys-110 confers an unexpected evolutionary advantage over other amino acids, such as Methionine, since it suppresses HGD crystallization. We also speculate that HGD evolved with Cys on position 110th instead of Ser since the latter would have provided a lower refractive index increment, a property necessary for an eye-lens function.

Previous work showed that phase the behaviour of double mutants of HGD can be predicted by the respective single-mutant phase diagrams. To confirm this, we created the double mutant P23VC110M incorporating the properties of 1) P23V, that forms reversible spherical assemblies with inverted solubility and 2) C110M, that enhances nucleation of orthorhombic crystals. We found that P23VC110M forms both these types of assemblies: its phase behaviour can hence be explained by the phase diagrams of the respective single mutants. However, P23VC110M also forms a third polymorph, crystal with needle/plate shape, which we speculate may arise from new crystal contacts, possibly involving the two mutation sites. The retrograde solubility assemblies due to mutation P23V also have a perfectly spherical shape, which is generally only observed for assemblies of amyloid material or amorphous particulates containing unfolded protein, which are irreversible. Using the double mutant P23VC110M, we showed they are composed of amorphous material but without any trace of amyloid and we therefore reveal that the HGD spherical assemblies are most likely a new type of protein superstructure.

Finally, the effect of protein surface modifications on ADDomer VLPs was assessed for their potential usage as vaccines. Specifically, this was done with the aim of improving VLP thermostability and hence improve their potential for distribution in remote areas of the world. The thermostability of ADDomer was compared to that of ChADDomer, a similar VLP type, and Chimera, a hybrid construct between ADDomer and ChADDomer. In addition, mutants L56C and S57C of ChADDomer and S57C of Chimera were probed for their ability to form intra-VLP disulphide bridges that may hinder VLP disassembly. We found that mutants S57C and L56C are the most thermally stable in terms of aggregation and protein secondary structure. In contrast to ChADDomer, ADDomer and Chimera VLPs abruptly aggregate, behaviour not due to minor differences in aromatic residue solvent exposure. However, by primary structure comparison, we show that the protein surface regions that most likely cause this difference are the epitopes-containing regions, VL and RGD loops.

This work both confirms that protein surface modifications, as modest as single-point mutations, have the potential to drastically affect protein phase behaviour. Furthermore, we show that new protein phase behaviours can be rationally designed in double mutations, extending earlier work and revealing that this may be a general route to designing proteins with specific phase transitions. Using this approach, we have also identified a new protein superstructure. By extending this fundamental work to VLPs, we show that weak and anisotropic interactions, when harnessed correctly, can lead to potentially life-saving advances in biotechnology, by combining tools and approaches from Soft Matter and Synthetic Biology.

Due to the pathogenic nature of the many HGD mutants that have been studied, the investigation of the effects of surface anisotropy on protein self-assembly using this eye-lens protein has been mainly experimental. A significant step forward was made by demonstrating that some non-pathogenic point mutations have the potential to drastically change the self-assembly behaviour of the whole protein^{96,97,204}. In recent years, the computational determination of phase diagrams for mutated and chemically modified HGD using patchy particle models has recapitulated the experimental work^{119,120,229}, but these models are not predictive of the impact of unknown mutations. In addition, building the models and simulating protein phase diagrams is computationally and time expensive. Careful studies such as these however, broaden our view of how anisotropic protein-protein interactions contribute to protein assembly, and may improve future modelling work. We believe that a close interplay between computational and experimental work, along with a careful design of the mutations to be attempted, is the best strategy to unravel this fundamental issue, and has the potential to relate the presence of specific surface amino acids to the biological role of the protein under investigation. The applications of this knowledge extend well beyond a single protein or assembly type. Understanding this topic is extremely useful in directing experiments to more applied projects such as for the assembly of virus-like particles. In

future, efforts will be directed to include specific epitopes on the VLP surface to respond to infectious diseases – and in evaluating their thermal stability by the methods we have used here. Furthermore, we believe it would be significant to establish new approaches for assessing to what extent VLPs disassembly is involved but also probing whether the process is reversible. Knowing the effects of some specific surface modifications or mutations on the self-assembly and solution stability of the VLPs studied may be crucial to develop highly stable vaccine formulations for safer and more efficient distribution and administration.

References

- (1) Whitford, D. *Proteins Structure and Function*; 2005.
- (2) Whitesides, G. M.; Grzybowski, B. Self-Assembly at All Scales. *Science* (80-.). **2002**, 295 (5564), 2418–2421. <https://doi.org/10.1126/science.1070821>.
- (3) Sharma, P. Self-Assembly of Colloidal Particles. *Resonance* **2018**, 23 (3), 263–275. <https://doi.org/10.1007/s12045-018-0616-0>.
- (4) Berj, J. M.; Tymoczko, J.; Stryer, L. Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media (Section 12.4). In *Biochemistry 5th Edition*; 2002.
- (5) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. *DNA Origami: Scaffolds for Creating Higher Order Structures*; 2017; Vol. 117. <https://doi.org/10.1021/acs.chemrev.6b00825>.
- (6) Yang, L.; Liu, A.; Cao, S.; Putri, R. M.; Jonkheijm, P.; Cornelissen, J. J. L. M. Self-Assembly of Proteins: Towards Supramolecular Materials. *Chem. - A Eur. J.* **2016**, 22 (44), 15570–15582. <https://doi.org/10.1002/chem.201601943>.
- (7) Whitesides, G. M.; Boncheva, M. Beyond Molecules: Self-Assembly of Mesoscopic and Macroscopic Components. *Proc. Natl. Acad. Sci.* **2002**, 99 (8), 4769–4774. <https://doi.org/doi.org/10.1073/pnas.082065899>.
- (8) Tamura, Y.; Kawamura, R.; Shikinaka, K.; Kakugo, A.; Osada, Y.; Gong, J. P.; Mayama, H. Dynamic Self-Organization and Polymorphism of Microtubule Assembly through Active Interactions with Kinesin. *Soft Matter* **2011**, 7 (12), 5654–5659. <https://doi.org/10.1039/c1sm05413a>.
- (9) Mcmanus, J. J.; Charbonneau, P.; Zaccarelli, E.; Asherie, N. The Physics of Protein Self-Assembly. *Curr. Opin. Colloid Interface Sci.* **2016**, 22, 73–79. <https://doi.org/10.1016/j.cocis.2016.02.011>.
- (10) Klostermeier, D.; Rudolph, M. G. *Biophysical Chemistry*; 2017. <https://doi.org/10.1201/9781315156910>.
- (11) Boyle, A. L. Applications of de Novo Designed Peptides. In *Peptide Applications in Biomedicine, Biotechnology and Bioengineering*; Elsevier Ltd, 2018; pp 51–86. <https://doi.org/10.1016/B978-0-08-100736-5.00003-X>.
- (12) Su, M.; Ling, Y.; Yu, J.; Wu, J.; Xiao, J. Small Proteins: Untapped Area of Potential Biological Importance. *Front. Genet.* **2013**, 4 (DEC), 1–9. <https://doi.org/doi:10.3389/fgene.2013.00286>.
- (13) Erickson, H. P. Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online* **2009**, 11 (1), 32–51. <https://doi.org/10.1007/s12575-009-9008-x>.
- (14) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; Tompa, P.; Fuxreiter, M.

- Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **2018**, *28* (6), 420–435. <https://doi.org/10.1016/j.tcb.2018.02.004>.
- (15) Gomes, E.; Shorter, J. The Molecular Language of Membraneless Organelles. *J. Biol. Chem.* **2019**, *294* (18), 7115–7127. <https://doi.org/10.1074/jbc.TM118.001192>.
- (16) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176* (3), 419–434. <https://doi.org/10.1016/j.cell.2018.12.035>.
- (17) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q. X.; Nixon, B. T.; Rosen, M. K. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483* (7389), 336–340. <https://doi.org/10.1038/nature10879>.
- (18) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* (80-.). **2009**, *324* (5935), 1729–1732. <https://doi.org/10.1126/science.1172046>.
- (19) Raoult, D.; Forterre, P. Redefining Viruses: Lessons from Mimivirus. *Nat. Rev. Microbiol.* **2008**, *6* (4), 315–319. <https://doi.org/10.1038/nrmicro1858>.
- (20) Bruinsma, R. F.; Wuite, G. J. L.; Roos, W. H. Physics of viral dynamics <http://dx.doi.org/10.1038/s42254-020-00267-1>. <https://doi.org/10.1038/s42254-020-00267-1>.
- (21) Perlmutter, J. D.; Hagan, M. F. Mechanisms of Virus Assembly. *Annu. Rev. Phys. Chem.* **2015**, *66* (1), 217–239. <https://doi.org/10.1146/annurev-physchem-040214-121637>.
- (22) Buzón, P.; Maity, S.; Roos, W. H. Physical Virology: From Virus Self-Assembly to Particle Mechanics. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **2020**, *12* (4), 1–22. <https://doi.org/10.1002/wnan.1613>.
- (23) Ghebremedhin, B. Human Adenovirus: Viral Pathogen with Increasing Importance. *Eur. J. Microbiol. Immunol.* **2014**, *4* (1), 26–33. <https://doi.org/10.1556/eujmi.4.2014.1.2>.
- (24) Nooraei, S.; Bahrulolum, H.; Hoseini, Z. S.; Katalani, C.; Hajizade, A.; Easton, A. J.; Ahmadian, G. Virus-like Particles: Preparation, Immunogenicity and Their Roles as Nanovaccines and Drug Nanocarriers. *J. Nanobiotechnology* **2021**, *19* (1), 1–27. <https://doi.org/10.1186/s12951-021-00806-7>.
- (25) Garmann, R. F.; Comas-Garcia, M.; Knobler, C. M.; Gelbart, W. M. Physical Principles in the Self-Assembly of a Simple Spherical Virus. *Acc. Chem. Res.* **2016**, *49* (1), 48–55. <https://doi.org/10.1021/acs.accounts.5b00350>.
- (26) Mason, T. O.; Shimanovich, U. Fibrous Protein Self-Assembly in Biomimetic Materials. *Adv. Mater.* **2018**, *30* (41), 1–11. <https://doi.org/10.1002/adma.201706462>.
- (27) Scheibel, T. Protein Fibers as Performance Proteins : New Technologies and Applications.

- Curr. Opin. Biotechnol.* **2005**, *16*, 427–433.
<https://doi.org/10.1016/j.copbio.2005.05.005>.
- (28) Bulleid, N. J.; Dalley, J. A.; Lees, J. F. The C-Propeptide Domain of Procollagen Can Be Replaced with a Transmembrane Domain without Affecting Trimer Formation or Collagen Triple Helix Folding during Biosynthesis. *EMBO J.* **1997**, *16* (22), 6694–6701.
<https://doi.org/10.1093/emboj/16.22.6694>.
- (29) Canty, E. G.; Kadler, K. E. Procollagen Trafficking, Processing and Fibrillogenesis. *J. Cell Sci.* **2005**, *118* (7), 1341–1353. <https://doi.org/10.1242/jcs.01731>.
- (30) DeFrates, K. G.; Moore, R.; Borgesi, J.; Lin, G.; Mulderig, T.; Beachley, V.; Hu, X. Protein-Based Fiber Materials in Medicine: A Review. *Nanomaterials* **2018**, *8* (7), 1–26.
<https://doi.org/10.3390/nano8070457>.
- (31) Forns, P.; Lauer-Fields, J. L.; Gao, S.; Fields, G. B. Induction of Protein-like Molecular Architecture by Monoalkyl Hydrocarbon Chains. *Biopolymers* **2000**, *54* (7), 531–546.
[https://doi.org/10.1002/1097-0282\(200012\)54:7<531::AID-BIP60>3.0.CO;2-X](https://doi.org/10.1002/1097-0282(200012)54:7<531::AID-BIP60>3.0.CO;2-X).
- (32) Novelli, F.; Strofaldi, A.; De Santis, S.; Del Giudice, A.; Casciardi, S.; Galantini, L.; Morosetti, S.; Pavel, N. V.; Masci, G.; Scipioni, A. Polymorphic Self-Organization of Lauroyl Peptide in Response to PH and Concentration. *Langmuir* **2020**, *36* (14), 3941–3951. <https://doi.org/10.1021/acs.langmuir.9b02924>.
- (33) Benedek, G. B. Cataract as a Protein Condensation Disease. *Invest. Ophthalmol. Vis. Sci.* **1997**, *38* (10), 1911–1921.
- (34) Hodson, R. Alzheimer’s Disease. *Nat. Outlook* **2018**, *559* (7715), S1.
<https://doi.org/10.1056/NEJM198604103141506>.
- (35) Hippus, H.; Neundörfer, G. The Discovery of Alzheimer’s Disease. *Dialogues Clin. Neurosci.* **2003**, *5* (1), 101–108. <https://doi.org/10.31887/DCNS.2003.5.1/hhippus>.
- (36) Chen, G. F.; Xu, T. H.; Yan, Y.; Zhou, Y. R.; Jiang, Y.; Melcher, K.; Xu, H. E. Amyloid Beta: Structure, Biology and Structure-Based Therapeutic Development. *Acta Pharmacol. Sin.* **2017**, *38* (9), 1205–1235. <https://doi.org/10.1038/aps.2017.28>.
- (37) Hamley, I. W. The Amyloid Beta Peptide: A Chemist’s Perspective. Role in Alzheimer’s and Fibrillization. *Chem. Rev.* **2012**, *112* (10), 5147–5192.
<https://doi.org/10.1021/cr3000994>.
- (38) Inouye, H.; Fraser, P. E.; Kirschner, D. A. Structure of Beta-Crystallite Assemblies Formed by Alzheimer Beta-Amyloid Protein Analogues: Analysis by x-Ray Diffraction. *Biophys. J.* **1993**, *64* (2), 502–519. [https://doi.org/10.1016/S0006-3495\(93\)81393-6](https://doi.org/10.1016/S0006-3495(93)81393-6).
- (39) Haass, C.; Selkoe, D. J. Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer’s Amyloid β -Peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (2), 101–112.
<https://doi.org/10.1038/nrm2101>.
- (40) Ladiwala, A. R. A.; Dordick, J. S.; Tessier, P. M. Aromatic Small Molecules Remodel

- Toxic Soluble Oligomers of Amyloid β through Three Independent Pathways. *J. Biol. Chem.* **2011**, *286* (5), 3209–3218. <https://doi.org/10.1074/jbc.M110.173856>.
- (41) Bloom, G. S. Amyloid- β and Tau. *JAMA Neurol.* **2014**, *71* (4), 505. <https://doi.org/10.1001/jamaneurol.2013.5847>.
- (42) d'Errico, P.; Meyer-Luehmann, M. Mechanisms of Pathogenic Tau and A β Protein Spreading in Alzheimer's Disease. *Front. Aging Neurosci.* **2020**, *12* (265). <https://doi.org/10.3389/fnagi.2020.00265>.
- (43) Sachse, C.; Grigorieff, N.; Fändrich, M. Nanoscale Flexibility Parameters of Alzheimer Amyloid Fibrils Determined by Electron Cryo-Microscopy. *Angew. Chemie* **2010**, *49* (7), 1321–1323. <https://doi.org/10.1002/anie.200904781>.
- (44) Kato, G. J.; Piel, F. B.; Reid, C. D.; Gaston, M. H.; Ohene-Frempong, K.; Krishnamurti, L.; Smith, W. R.; Panepinto, J. A.; Weatherall, D. J.; Costa, F. F.; Vichinsky, E. P. Sick Cell Disease. *Nat. Rev. Dis. Prim.* **2018**, *4*, 1–22. <https://doi.org/10.1038/nrdp.2018.10>.
- (45) Wishner, B. C.; Ward, K. B.; Lattman, E. E.; Love, W. E. Crystal Structure of Sick Cell Deoxyhemoglobin at 5 Å Resolution. *J. Mol. Biol.* **1975**, *98* (1), 179–194. [https://doi.org/10.1016/S0022-2836\(75\)80108-2](https://doi.org/10.1016/S0022-2836(75)80108-2).
- (46) Vekilov, P. G. Sick Cell Haemoglobin Polymerization: Is It the Primary Pathogenic Event of Sick Cell Anaemia? *Br. J. Haematol.* **2007**, *139* (2), 173–184. <https://doi.org/10.1111/j.1365-2141.2007.06794.x>.
- (47) Ashley-Koch, A.; Yang, Q.; Olney, R. S. Sick Cell Hemoglobin (Hb S) Allele and Sick Cell Disease: A HuGE Review. *Am. J. Epidemiol.* **2000**, *151* (9), 839–845. <https://doi.org/10.1093/oxfordjournals.aje.a010288>.
- (48) Benedek, G. B. Cataract as a Protein Condensation Disease: The Proctor Lecture. *Invest. Ophthalmol. Vis. Sci.* **1997**, *38* (10), 1911–1921. [https://doi.org/10.1016/S0002-9394\(99\)80119-6](https://doi.org/10.1016/S0002-9394(99)80119-6).
- (49) Eaton, W. A.; Hofrichter, J. The Biophysics of Sick Cell Hydroxyurea Therapy. *Science (80-.)*. **1995**, *268* (5214), 1142–1143. <https://doi.org/10.1126/science.7539154>.
- (50) Faes, C.; Ilich, A.; Sotiaux, A.; Sparkenbaugh, E. M.; Henderson, M. W.; Buczek, L.; Beckman, J. D.; Ellsworth, P.; Noubouossie, D. F.; Bhoopat, L.; Piegore, M.; Renoux, C.; Bergmeier, W.; Park, Y.; Ataga, K. I.; Cooley, B.; Wolberg, A. S.; Key, N. S.; Pawlinski, R. Red Blood Cells Modulate Structure and Dynamics of Venous Clot Formation in Sick Cell Disease. *Blood* **2019**, *133* (23), 2529–2541. <https://doi.org/10.1182/blood.2019000424>.
- (51) Bloemendal, H. The Vertebrate Eye Lens. *Science (80-.)*. **1977**, *197* (4299), 127–138. <https://doi.org/10.1126/science.877544>.
- (52) V  r  tout, F.; Tardieu, A. The Protein Concentration Gradient within Eye Lens Might Originate from Constant Osmotic Pressure Coupled to Differential Interactive Properties

- of Crystallins. *Eur. Biophys. J.* **1989**, *17* (2), 61–68. <https://doi.org/10.1007/BF00257103>.
- (53) Tardieu, A.; Véré tout, F.; Krop, B.; Slingsby, C. Protein Interactions in the Calf Eye Lens: Interactions between β -Crystallins Are Repulsive Whereas in γ -Crystallins They Are Attractive. *Eur. Biophys. J.* **1992**, *21* (1), 1–12. <https://doi.org/10.1007/BF00195438>.
- (54) Siezen, R. J.; Fisch, M. R.; Slingsby, C.; Benedek, G. B. Opacification of γ -Crystallin Solutions from Calf Lens in Relation to Cold Cataract Formation. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (6), 1701–1705. <https://doi.org/10.1073/pnas.82.6.1701>.
- (55) Graw, J. Genetics of Crystallins: Cataract and Beyond. *Exp. Eye Res.* **2009**, *88* (2), 173–189. <https://doi.org/10.1016/j.exer.2008.10.011>.
- (56) Pande, A.; Annunziata, O.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. Decrease in Protein Solubility and Cataract Formation Caused by the Pro23 to Thr Mutation in Human γ D-Crystallin. *Biochemistry* **2005**, 2491–2500. <https://doi.org/10.1021/bi0479611>.
- (57) Pande, A.; Pande, J.; Asherie, N.; Lomakin, A.; Ogun, O.; King, J.; Benedek, G. B. Crystal Cataracts: Human Genetic Cataract Caused by Protein Crystallization. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (11), 6116–6120. <https://doi.org/10.1073/pnas.101124798>.
- (58) Kmoch, S.; Brynda, J.; Asfaw, B.; Bezouška, K.; Novák, P.; Rezácová, P.; Ondrová, L.; Filipец, M.; Sedláček, J.; Elleder, M. Link between a Novel Human γ D-Crystallin Allele and a Unique Cataract Phenotype Explained by Protein Crystallography. *Hum. Mol. Genet.* **2000**, *9* (12), 1779–1786. <https://doi.org/10.1093/hmg/9.12.1779>.
- (59) Doye, J. P. K.; Louis, A. A.; Vendruscolo, M. Inhibition of Protein Crystallization by Evolutionary Negative Design. *Phys. Biol.* **2004**, *1* (1). <https://doi.org/10.1088/1478-3967/1/1/P02>.
- (60) Doye, J. P. K.; Poon, W. C. K. Protein Crystallization in Vivo. *Curr. Opin. Colloid Interface Sci.* **2006**, *11* (1), 40–46. <https://doi.org/10.1016/j.cocis.2005.10.002>.
- (61) Schönherr, R.; Rudolph, J. M.; Redecke, L. Protein Crystallization in Living Cells. *Biol. Chem.* **2018**, *399* (7), 751–772. <https://doi.org/10.1515/hsz-2018-0158>.
- (62) Dodson, G.; Steiner, D. The Role of Assembly in Insulin's Biosynthesis. *Curr. Opin. Struct. Biol.* **1998**, *8* (2), 189–194. [https://doi.org/10.1016/S0959-440X\(98\)80037-7](https://doi.org/10.1016/S0959-440X(98)80037-7).
- (63) McPherson, A.; Gavira, J. A. Introduction to Protein Crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **2014**, *70* (1), 2–20. <https://doi.org/10.1107/S2053230X13033141>.
- (64) Krauss, I. R.; Merlino, A.; Vergara, A.; Sica, F. An Overview of Biological Macromolecule Crystallization. *Int. J. Mol. Sci.* **2013**, *14* (6), 11643–11691. <https://doi.org/10.3390/ijms140611643>.
- (65) Wen, F.; Rubin-pitel, S. B.; Zhao, H. Engineering of Therapeutic Proteins. In *Protein Engineering and Design*; 2010; pp 153–176.
- (66) Lagassé, H. A. D.; Alexaki, A.; Simhadri, V. L.; Katagiri, N. H.; Jankowski, W.; Sauna,

- Z. E.; Kimchi-Sarfaty, C. Recent Advances in (Therapeutic Protein) Drug Development. *F1000Research* **2017**, *6*, 1–7. <https://doi.org/10.12688/f1000research.9970.1>.
- (67) Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F. Physical Stability of Proteins in Aqueous Solution : Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharm. Res.* **2003**, *20* (9), 1325–1336. <https://doi.org/10.1023/A:1025771421906>.
- (68) Shire, S. J.; Shahrokh, Z.; Liu, J. Challenges in the Development of High Protein Concentration Formulations. *J. Pharm. Sci.* **2004**, *93* (6), 1390–1402. <https://doi.org/10.1002/jps.20079>.
- (69) Wang, W. Protein Aggregation and Its Inhibition in Biopharmaceutics. *Int. J. Pharm.* **2005**, *289* (1–2), 1–30. <https://doi.org/10.1016/j.ijpharm.2004.11.014>.
- (70) Roldão, A.; Mellado, M. C. M.; Castilho, L. R.; Carrondo, M. J. T.; Alves, P. M. Virus-like Particles in Vaccine Development. *Expert Rev. Vaccines* **2010**, *9* (10), 1149–1176. <https://doi.org/10.1586/erv.10.115>.
- (71) Marsian, J.; Fox, H.; Bahar, M. W.; Kotecha, A.; Fry, E. E.; Stuart, D. I.; Macadam, A. J.; Rowlands, D. J.; Lomonosoff, G. P. Plant-Made Polio Type 3 Stabilized VLPs-A Candidate Synthetic Polio Vaccine. *Nat. Commun.* **2017**, *8*, 1–9. <https://doi.org/10.1038/s41467-017-00090-w>.
- (72) Fernández, E.; Toledo, J. R.; Méndez, L.; González, N.; Parra, F.; Martín-Alonso, J. M.; Limonta, M.; Sánchez, K.; Cabrales, A.; Estrada, M. P.; Rodríguez-Mallón, A.; Farnós, O. Conformational and Thermal Stability Improvements for the Large-Scale Production of Yeast-Derived Rabbit Hemorrhagic Disease Virus-Like Particles as Multipurpose Vaccine. *PLoS One* **2013**, *8* (2). <https://doi.org/10.1371/journal.pone.0056417>.
- (73) Kern, N.; Frenkel, D. Fluid-Fluid Coexistence in Colloidal Systems with Short-Ranged Strongly Directional Attraction. *J. Chem. Phys.* **2003**, *118* (21), 9882–9889. <https://doi.org/10.1063/1.1569473>.
- (74) Sear, R. P. Low-Temperature Interface between the Gas and Solid Phases of Hard Spheres with a Short-Ranged Attraction. *Phys. Rev. E* **1999**, *59* (6), 6838–6841. <https://doi.org/10.1103/PhysRevE.59.6838>.
- (75) Lomakin, A.; Asherie, N.; Benedek, G. B. Aeolotopic Interactions of Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (17), 9465–9468. <https://doi.org/10.1073/pnas.96.17.9465>.
- (76) McElroy, H. E.; Sisson, G. W.; Schoettlin, W. E.; Aust, R. M.; Villafranca, J. E. Studies on Engineering Crystallizability by Mutation of Surface Residues of Human Thymidylate Synthase. *J. Cryst. Growth* **1992**, *122*, 265–272. [https://doi.org/10.1016/0022-0248\(92\)90255-H](https://doi.org/10.1016/0022-0248(92)90255-H).
- (77) Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. *Structure* **2004**, *12*, 529–535. <https://doi.org/10.1016/j.str.2004.03.008>.

- (78) Figueira-Mansur, J.; Aguilera, E. A.; Stoque, R. M.; Ventura, G. T.; Mohana-Borges, R. Mutations in the Dimer Interfaces of the Dengue Virus Capsid Protein Affect Structural Stability and Impair RNA-Capsid Interaction. *Sci. Rep.* **2019**, *9* (2829), 1–14. <https://doi.org/10.1038/s41598-019-39185-3>.
- (79) Ceres, P.; Stray, S. J.; Zlotnick, A. Hepatitis B Virus Capsid Assembly Is Enhanced by Naturally Occurring Mutation F97L. *J. Virol.* **2004**, *78* (17), 9538–9543. <https://doi.org/10.1128/jvi.78.17.9538-9543.2004>.
- (80) Hill, T. L. *An Introduction to Statistical Thermodynamics*; 1961. <https://doi.org/10.1063/1.3057470>.
- (81) Gunton, J. D.; Shiryayev, A.; Pagan, D. L. *Protein Condensation: Kinetic Pathways to Crystallization and Disease*; 2007.
- (82) Pate, K.; Safier, P. Chemical Metrology Methods for CMP Quality. In *Advances in Chemical Mechanical Planarization (CMP)*; 2016; pp 299–324. <https://doi.org/10.1016/B978-0-08-100165-3.00012-7>.
- (83) Israelachvili, J. N. *Intermolecular and Surface Forces: Third Edition*; 1997. <https://doi.org/10.1016/C2011-0-05119-0>.
- (84) Curtis, R. A.; Blanch, H. W.; Prausnitz, J. M. Calculation of Phase Diagrams for Aqueous Protein Solutions. *J. Phys. Chem. B* **2001**, *105*, 2445–2452. <https://doi.org/10.1021/jp003087j>.
- (85) Atkins, P.; De Paula, J. *Physical Chemistry for the Life Sciences*; 2006.
- (86) Ten Wolde, P. R.; Frenkel, D. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science* (80-.). **1997**, *277* (5334), 1975–1978. <https://doi.org/10.1126/science.277.5334.1975>.
- (87) Asherie, N.; Lomakin, A.; Benedek, G. B. Phase Diagram of Colloidal Solutions. *Phys. Rev. Lett.* **1996**, *77* (23), 4832–4835. <https://doi.org/10.1103/PhysRevLett.77.4832>.
- (88) Fusco, D.; Charbonneau, P. Soft Matter Perspective on Protein Crystal Assembly. *Colloids Surfaces B Biointerfaces* **2016**, *137*, 22–31. <https://doi.org/10.1016/j.colsurfb.2015.07.023>.
- (89) Muschol, M.; Rosenberger, F. Interactions in Undersaturated and Supersaturated Lysozyme Solutions: Static and Dynamic Light Scattering Results. *J. Chem. Phys.* **1995**, *103* (24), 10424–10432. <https://doi.org/10.1063/1.469891>.
- (90) Curtis, R. A.; Ulrich, J.; Montaser, A.; Prausnitz, J. M.; Blanch, H. W. Protein-Protein Interactions in Concentrated Electrolyte Solutions: Hofmeister-Series Effects. *Biotechnol. Bioeng.* **2002**, *79* (4), 367–380. <https://doi.org/10.1002/bit.10342>.
- (91) Eberstein, W.; Georgalis, Y.; Saenger, W. Molecular Interactions in Crystallizing Lysozyme Solutions Studied by Photon Correlation Spectroscopy. *J. Cryst. Growth* **1994**, *143* (1–2), 71–78. [https://doi.org/10.1016/0022-0248\(94\)90369-7](https://doi.org/10.1016/0022-0248(94)90369-7).

- (92) Curtis, R. A.; Lue, L. A Molecular Approach to Bioseparations: Protein-Protein and Protein-Salt Interactions. *Chem. Eng. Sci.* **2006**, *61* (3), 907–923. <https://doi.org/10.1016/j.ces.2005.04.007>.
- (93) McMillan, W. G.; Mayer, J. E. The Statistical Thermodynamics of Multicomponent Systems. *J. Chem. Phys.* **1945**, *13* (7), 276–305. <https://doi.org/10.1063/1.1724036>.
- (94) Muschol, M.; Rosenberger, F. Liquid–Liquid Phase Separation in Supersaturated Lysozyme Solutions and Associated Precipitate Formation/Crystallization. *J. Chem. Phys.* **1997**, *107* (6), 1953–1962. <https://doi.org/10.1063/1.474547>.
- (95) Stradner, A.; Schurtenberger, P. Potential and Limits of a Colloid Approach to Protein Solutions. *Soft Matter* **2020**, *16* (2), 307–323. <https://doi.org/10.1039/c9sm01953g>.
- (96) McManus, J. J.; Lomakin, A.; Basan, M.; Pande, A.; Ogun, O.; Pande, J.; Benedek, G. Altered Phase Diagram Due to a Single Point Mutation in Human γ D-Crystallin. *Proc. Natl. Acad. Sci.* **2007**, *104* (43), 16856–16861. <https://doi.org/10.1073/pnas.0707412104>.
- (97) James, S.; Quinn, M. K.; Mcmanus, J. J. The Self Assembly of Proteins; Probing Patchy Protein Interactions. *Phys. Chem. Chem. Phys.* **2015**, *17*, 5413–5420. <https://doi.org/10.1039/C4CP05892E>.
- (98) Serrano, M. D.; Galkin, O.; Yau, S. T.; Thomas, B. R.; Nagel, R. L.; Hirsch, R. E.; Vekilov, P. G. Are Protein Crystallization Mechanisms Relevant to Understanding and Control of Polymerization of Deoxyhemoglobin S? *J. Cryst. Growth* **2001**, *232* (1–4), 368–375. [https://doi.org/10.1016/S0022-0248\(01\)01068-5](https://doi.org/10.1016/S0022-0248(01)01068-5).
- (99) Lafont, S.; Veessler, S.; Astier, J. P.; Boistelle, R. Solubility and Prenucleation of Aprotinin (BPTI) Molecules in Sodium Chloride Solutions. *J. Cryst. Growth* **1994**, *143* (3–4), 249–255. [https://doi.org/10.1016/0022-0248\(94\)90064-7](https://doi.org/10.1016/0022-0248(94)90064-7).
- (100) Petsev, D. N.; Thomas, B. R.; Yau, S. T.; Tsekova, D.; Nanev, C.; William Wilson, W.; Vekilov, P. G. Temperature-Independent Solubility and Interactions between Apoferritin Monomers and Dimers in Solution. *J. Cryst. Growth* **2001**, *232* (1–4), 21–29. [https://doi.org/10.1016/S0022-0248\(01\)01095-8](https://doi.org/10.1016/S0022-0248(01)01095-8).
- (101) Ishimoto, C.; Tanaka, T. Critical Behaviour of a Binary Mixture of Protein and Salt Water. *Phys. Rev. Lett.* **1977**, *39* (8), 474–477. <https://doi.org/10.1103/PhysRevLett.39.474>.
- (102) Taratuta, V. G.; Holschbach, A.; Thurston, G. M.; Blankschtein, D.; Benedek, G. B. Liquid-Liquid Phase Separation of Aqueous Lysozyme Solutions: Effects of PH and Salt Identity. *J. Phys. Chem.* **1990**, *94* (5), 2140–2144. <https://doi.org/10.1021/j100368a074>.
- (103) Thomson, J. a.; Schurtenberger, P.; Thurston, G. M.; Benedek, G. B. Binary Liquid Phase Separation and Critical Phenomena in a Protein/Water Solution. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (20), 7079–7083. <https://doi.org/10.1073/pnas.84.20.7079>.
- (104) San Biagio, P. L.; Palma, M. U. Spinodal Lines and Flory-Huggins Free-Energies for Solutions of Human Hemoglobins HbS and HbA. *Biophys. J.* **1991**, *60* (2), 508–512.

[https://doi.org/10.1016/S0006-3495\(91\)82078-1](https://doi.org/10.1016/S0006-3495(91)82078-1).

- (105) Grouazel, S.; Bonneté, F.; Astier, J. P.; Ferté, N.; Perez, J.; Veessler, S. Exploring Bovine Pancreatic Trypsin Inhibitor Phase Transitions. *J. Phys. Chem. B* **2006**, *110* (39), 19664–19670. <https://doi.org/10.1021/jp0627123>.
- (106) Wang, Y.; Lomakin, A.; Latypov, R. F.; Benedek, G. B. Phase Separation in Solutions of Monoclonal Antibodies and the Effect of Human Serum Albumin. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (40), 16606–16611. <https://doi.org/10.1073/pnas.1112241108>.
- (107) Mason, B. D.; Enk, J. Z.; Zhang, L.; Jr, R. L. R.; Zhang, J. Liquid-Liquid Phase Separation of a Monoclonal Antibody and Nonmonotonic Influence of Hofmeister Anions. *Biophys. journal* **2010**, *99* (11), 3792–3800. <https://doi.org/10.1016/j.bpj.2010.10.040>.
- (108) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; Pozniakovski, A.; Poser, I.; Maghelli, N.; Royer, L. A.; Weigert, M.; Myers, E. W.; Grill, S.; Drechsel, D.; Hyman, A. A.; Alberti, S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162* (5), 1066–1077. <https://doi.org/10.1016/j.cell.2015.07.047>.
- (109) Li, H.; Chiang, W.; Chou, P.; Wang, W.; Huang, J. TAR DNA-Binding Protein 43 (TDP-43) Liquid – Liquid Phase Separation Is Mediated by Just a Few Aromatic Residues. *J. biolo* **2018**, *43*, 6090–6098. <https://doi.org/10.1074/jbc.AC117.001037>.
- (110) Majumdar, A.; Dogra, P.; Maity, S.; Mukhopadhyay, S. Liquid-Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* **2019**, *10* (14), 3929–3936. <https://doi.org/10.1021/acs.jpcllett.9b01731>.
- (111) Guo, B.; Kao, S.; Mcdonald, H.; Asanov, A.; Combs, L. L.; Wilson, W. W. Correlation of Second Virial Coefficients and Solubilities Useful in Protein Crystal Growth. *J. Cryst. Growth* **1999**, *196*, 424–433. [https://doi.org/10.1016/S0022-0248\(98\)00842-2](https://doi.org/10.1016/S0022-0248(98)00842-2).
- (112) Petsev, D. N.; Wu, X.; Galkin, O.; Vekilov, P. G. Thermodynamic Functions of Concentrated Protein Solutions from Phase Equilibria. *J. Phys. Chem. B* **2003**, *107* (16), 3921–3926. <https://doi.org/10.1021/jp0278317>.
- (113) Lomakin, A.; Asherie, N.; Benedek, G. B. Monte Carlo Study of Phase Separation in Aqueous Protein Solutions. *J. Phys. Chem.* **1996**, *104* (October 1995), 1646–1656. <https://doi.org/10.1063/1.470751>.
- (114) Platten, F.; Valadez-Pérez, N. E.; Castañeda-Priego, R.; Egelhaaf, S. U. Extended Law of Corresponding States for Protein Solutions. *J. Chem. Phys.* **2015**, *142* (17), 1–15. <https://doi.org/10.1063/1.4919127>.
- (115) Noro, M. G.; Frenkel, D. Extended Corresponding-States Behavior for Particles with Variable Range Attractions. *J. Chem. Phys.* **2000**, *113* (8), 2941–2944. <https://doi.org/10.1063/1.1288684>.

- (116) Gögelein, C.; Nägele, G.; Tuinier, R.; Gibaud, T.; Stradner, A.; Schurtenberger, P. A Simple Patchy Colloid Model for the Phase Behavior of Lysozyme Dispersions. *J. Chem. Phys.* **2008**, *129*, 1–12. <https://doi.org/10.1063/1.2951987>.
- (117) Sun, G.; Wang, Y.; Lomakin, A.; Benedek, G. B.; Stanley, H. E.; Xu, L.; Buldyrev, S. V. The Phase Behavior Study of Human Antibody Solution Using Multi-Scale Modeling. *Journal of Chemical Physics*. 2016. <https://doi.org/10.1063/1.4966972>.
- (118) Fusco, D.; Headd, J. J.; De Simone, A.; Wang, J.; Charbonneau, P. Characterizing Protein Crystal Contacts and Their Role in Crystallization : Rubredoxin as a Case Study. *Soft Matter* **2014**, *10*, 290–302. <https://doi.org/10.1039/c3sm52175c>.
- (119) Khan, A. R.; James, S.; Quinn, M. K.; Altan, I.; Charbonneau, P.; McManus, J. J. Temperature-Dependent Interactions Explain Normal and Inverted Solubility in a γ D-Crystallin Mutant. *Biophys. J.* **2019**, *117*, 930–937. <https://doi.org/10.1016/j.bpj.2019.07.019>.
- (120) Quinn, M. K.; Gnan, N.; James, S.; Ninarello, A.; Sciortino, F.; Zaccarelli, E. How Fluorescent Labelling Alters the Solution Behaviour of Proteins. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31177–31187. <https://doi.org/10.1039/C5CP04463D>.
- (121) Quinn, M. K.; James, S.; McManus, J. J. Chemical Modification Alters Protein-Protein Interactions and Can Lead to Lower Protein Solubility. *J. Phys. Chem. B* **2019**, *123* (20), 4373–4379. <https://doi.org/10.1021/acs.jpcc.9b02368>.
- (122) Mcguffee, S. R.; Elcock, A. H. Atomically Detailed Simulations of Concentrated Protein Solutions : The Effects of Salt , PH , Point Mutations , and Protein Concentration in Simulations of 1000-Molecule Systems. *J. Am. Chem. Soc. Amer* **2006**, *128*, 12098–12110. <https://doi.org/10.1021/ja0614058>.
- (123) Qin, S.; Zhou, H. Calculation of Second Virial Coefficients of Atomistic Proteins Using Fast Fourier Transform. *J. Phys. Chem. B* **2019**, *123*, 8203–8215. <https://doi.org/10.1021/acs.jpcc.9b06808>.
- (124) Elcock, A. H.; Mccammon, J. A. Calculation of Weak Protein-Protein Interactions : The PH Dependence of the Second Virial Coefficient. *Biophys. J.* **2001**, *80* (2), 613–625. [https://doi.org/10.1016/S0006-3495\(01\)76042-0](https://doi.org/10.1016/S0006-3495(01)76042-0).
- (125) Neal, B. L.; Asthagiri, D.; Velev, O. D.; Lenhoff, A. M.; Kaler, E. W. Why Is the Osmotic Second Virial Coefficient Related to Protein Crystallization? *J. Cryst. Growth* **1999**, *196*, 377–387. [https://doi.org/10.1016/S0022-0248\(98\)00855-0](https://doi.org/10.1016/S0022-0248(98)00855-0).
- (126) Curtis, R. A.; Steinbrecher, C.; Heinemann, M.; Blanch, H. W.; Prausnitz, J. M. Hydrophobic Forces between Protein Molecules in Aqueous Solutions of Concentrated Electrolyte. *Biophys. Chem.* **2002**, *98*, 249–265. [https://doi.org/10.1016/s0301-4622\(02\)00071-6](https://doi.org/10.1016/s0301-4622(02)00071-6).
- (127) Chang, R. C.; Asthagiri, D.; Lenhoff, A. M. Measured and Calculated Effects of Mutations

- in Bacteriophage T4 Lysozyme on Interactions in Solution. *Proteins Struct. Funct. Genet.* **2000**, *41*, 123–132..
- (128) Asherie, N.; Pande, J.; Pande, A.; Zarutskie, J. A.; Lomakin, J.; Lomakin, A.; Ogun, O.; Stern, L. J.; King, J.; Benedek, G. B. Enhanced Crystallization of the Cys18 to Ser Mutant of Bovine γ B Crystallin. *J. Mol. Biol. Mol. Biol.* **2001**, *314*, 663–669. <https://doi.org/10.1006/jmbi.2001.5155>.
- (129) Rupp, B. Origin and Use of Crystallization Phase Diagrams. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **2015**, *71* (iii), 247–260. <https://doi.org/10.1107/S2053230X1500374X>.
- (130) Chayen, N. E. Turning Protein Crystallisation from an Art into a Science. *Curr. Opin. Struct. Biol.* **2004**, *14*, 577–583. <https://doi.org/10.1016/j.sbi.2004.08.002>.
- (131) George, A.; Wilson, W. Predicting Protein Crystallization from a Dilute Solution Property. *Acta Crystallogr.* **1994**, *D50* (0907–4449), 361–365. <https://doi.org/10.1107/S0907444994001216>.
- (132) Asherie, N. Protein Crystallization and Phase Diagrams. *Methods* **2004**, *34* (3), 266–272. <https://doi.org/10.1016/j.ymeth.2004.03.028>.
- (133) Asherie, N. A Dialogue about Protein Crystallization and Phase Diagrams. *Protein Pept. Lett.* **2012**, *19* (7), 708–713. <https://doi.org/10.2174/092986612800793217>.
- (134) Saridakis, E.; Chayen, N. E. Systematic Improvement of Protein Crystals by Determining the Supersolubility Curves of Phase Diagrams. *Biophys. J.* **2003**, *84* (2 I), 1218–1222. [https://doi.org/10.1016/S0006-3495\(03\)74936-4](https://doi.org/10.1016/S0006-3495(03)74936-4).
- (135) Li, L.; Ismagilov, R. F. Protein Crystallization Using Microfluidic Technologies Based on Valves, Droplets, and SlipChip. *Annu. Rev. Biophys.* **2010**, *39* (1), 139–158. <https://doi.org/10.1146/annurev.biophys.050708.133630>.
- (136) Vekilov, P. G. Solvent Entropy Effects in the Formation of Protein Solid Phases. In *Methods in Enzymology*; 2003; pp 84–105. [https://doi.org/10.1016/S0076-6879\(03\)68006-7](https://doi.org/10.1016/S0076-6879(03)68006-7).
- (137) Derewenda, Z. S.; G., V. P. Entropy and Surface Engineering in Protein Crystallization. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62*, 116–124. <https://doi.org/10.1107/S0907444905035237>.
- (138) Schall, C. A.; Arnold, E.; Wienczek, J. M. Enthalpy of Crystallization of Hen Egg-White Lysozyme. *J. Cryst. Growth* **1996**, *165* (3), 293–298. [https://doi.org/10.1016/0022-0248\(96\)00180-7](https://doi.org/10.1016/0022-0248(96)00180-7).
- (139) Yau, S.; Petsev, D. N.; Thomas, B. R.; Vekilov, P. G. Molecular-Level Thermodynamic and Kinetic Parameters for the Self-Assembly of Apoferritin Molecules into Crystals. *Journal Mol. Biol.* **2000**, *303*, 667–678. <https://doi.org/10.1006/jmbi.2000.4171>.
- (140) Gliko, O.; Neumaier, N.; Fischer, M.; Haase, I.; Bacher, A.; Weinkauff, S.; Vekilov, P. G.

- Dense Liquid Droplets as a Step Source for the Crystallization of Lumazine Synthase. *J. Cryst. Growth* **2005**, *275*, e1409–e1416. <https://doi.org/10.1016/j.jcrysgro.2004.11.291>.
- (141) Durbin, S. D.; Feher, G. Protein Crystallization. *J. Cryst. Growth* **1996**, *166* (1–4), 40–54. <https://doi.org/0.1146/annurev.physchem.47.1.171>.
- (142) Nanev, C. N. Peculiarities of Protein Crystal Nucleation and Growth. *Crystals* **2018**, *8* (11). <https://doi.org/10.3390/cryst8110422>.
- (143) Finkelstein, A. V.; Janin, J. The Price of Lost Freedom: Entropy of Bimolecular Complex Formation. *Protein Eng.* **1989**, *3*, 1–3. <https://doi.org/10.1093/protein/3.1.1>.
- (144) Nandi, N.; Bagchi, B. Dielectric Relaxation of Biological Water. *J. Phys. Chem. B* **1997**, *101* (50), 10954–10961. <https://doi.org/10.1021/jp971879g>.
- (145) Pal, S. K.; Peon, J.; Zewail, A. H. Biological Water at the Protein Surface: Dynamical Solvation Probed Directly with Femtosecond Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (4), 1763–1768. <https://doi.org/10.1073/pnas.042697899>.
- (146) Sear, R. P. Quantitative Studies of Crystal Nucleation at Constant Supersaturation: Experimental Data and Models. *Cryst. engineering Commun.* **2014**, *16* (6506), 6506–6522. <https://doi.org/10.1039/c4ce00344f>.
- (147) Vekilov, P. G. Nucleation of Protein Crystals. *Prog. Cryst. Growth Charact. Mater.* **2016**, *62* (2), 136–154. <https://doi.org/10.1016/j.pcrysgrow.2016.04.007>.
- (148) Boistelle, R.; Astier, J. P. Crystallization Mechanisms in Solution. *J. Cryst. Growth* **1988**, *90*, 14–30. [https://doi.org/doi.org/10.1016/0022-0248\(88\)90294-1](https://doi.org/doi.org/10.1016/0022-0248(88)90294-1).
- (149) Myerson, A. S. Concluding Remarks. *Faraday Discuss.* **2015**, *179*, 543–547. <https://doi.org/10.1039/c5fd00042d>.
- (150) García-Ruiz, J. M. Nucleation of Protein Crystals. *J. Struct. Biol.* **2003**, *142*, 22–31. [https://doi.org/10.1016/S1047-8477\(03\)00035-2](https://doi.org/10.1016/S1047-8477(03)00035-2).
- (151) Galkin, O.; Vekilov, P. G. Control of Protein Crystal Nucleation around the Metastable Liquid – Liquid Phase Boundary. *Proc. Natl. Acad. Sci.* **2000**, *97* (12), 6277–6281. <https://doi.org/10.1073/pnas.110000497>.
- (152) Vivarès, D.; Kaler, E. W.; Lenhoff, A. M. Quantitative Imaging by Confocal Scanning Fluorescence Microscopy of Protein Crystallization via Liquid-Liquid Phase Separation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61* (0907–4449), 819–825. <https://doi.org/10.1107/S090744490402949X>.
- (153) Luo, J.; Liu, Z.; Guo, Y.; Li, M. A Structural Dissection of Large Protein-Protein Crystal Packing Contacts. *Sci. Rep.* **2015**, *5*, 1–13. <https://doi.org/10.1038/srep14214>.
- (154) Rayment, I.; Rypniewski, W. R.; Schmidt-base, K.; Smith, R.; Tomchick, D. R.; Benning, M. M.; Winkelmann, D. A.; Wesenberg, G.; Holden, H. M. Three-Dimensional Structure of Myosin Subfragment-1: A Molecular Motor. *Science (80-.).* **1993**, *261*, 50–58. <https://doi.org/10.1126/science.8316857>.

- (155) Derewenda, Z. S.; Godzik, A. The “Sticky Patch” Model of Crystallization and Modification of Proteins for Enhanced Crystallizability. In *Protein Crystallography: Methods and Protocols*; 2017; pp 77–115. <https://doi.org/10.1007/978-1-4939-7000-1>.
- (156) Longenecker, K. L.; Garrard, S. M.; Sheffield, P. J.; Derewenda, Z. S. Protein Crystallization by Rational Mutagenesis of Surface Residues: Lys to Ala Mutations Promote Crystallization of RhoGDI. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2001**, *57* (5), 679–688. <https://doi.org/10.1107/S0907444901003122>.
- (157) Derewenda, Z. S. It ’ s All in the Crystals . . . *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *D67* (0907–4449), 243–248. <https://doi.org/10.1107/S0907444911007797>.
- (158) Prevelige, P. E.; Thomas, D. J.; King, J. Nucleation and Growth Phases in the Polymerization of Coat and Scaffolding Subunits into Icosahedral Procapsid Shells. *Biophys. J.* **1993**, *64* (March), 824–835. [https://doi.org/10.1016/S0006-3495\(93\)81443-7](https://doi.org/10.1016/S0006-3495(93)81443-7).
- (159) McPherson, A. Micelle Formation and Crystallization as Paradigms for Virus Assembly. *BioEssays* **2005**, *27* (4), 447–458. <https://doi.org/10.1002/bies.20196>.
- (160) Ramalho, R.; Rankovic, S.; Zhou, J.; Aiken, C.; Rousso, I. Analysis of the Mechanical Properties of Wild Type and Hyperstable Mutants of the HIV-1 Capsid. *Retrovirology* **2016**, *13* (1), 1–7. <https://doi.org/10.1186/s12977-016-0250-4>.
- (161) Brault, V.; Bergdoll, M.; Mutterer, J.; Prasad, V.; Pfeffer, S.; Erdinger, M.; Richards, K. E.; Ziegler-Graff, V. Effects of Point Mutations in the Major Capsid Protein of Beet Western Yellows Virus on Capsid Formation, Virus Accumulation, and Aphid Transmission. *J. Virol.* **2003**, *77* (5), 3247–3256. <https://doi.org/10.1006/viro.1997.8476>.
- (162) Díaz-Valle, A.; Falcón-González, J. M.; Carrillo-Tripp, M. Hot Spots and Their Contribution to the Self-Assembly of the Viral Capsid: In Silico Prediction and Analysis. *Int. J. Mol. Sci.* **2019**, *20* (23). <https://doi.org/10.3390/ijms20235966>.
- (163) Valdez-Lara, A. G.; Andrade-Medina, M.; Aleman-Vilis, J. A.; Perez-Montoya, A. A.; Pineda-Anguilar, N.; Martinez-Guerra, E.; Roldan-Salgado, A.; Gaytan, P.; Carrillo-Tripp, M. Hot-Spots and Their Contribution to the Self-Assembly of the Viral Capsid: In-Vitro Validation. *bioRxiv* **2019**. <https://doi.org/10.1101/723023>.
- (164) Hong, P.; Koza, S.; Bouvier, E. S. P. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and Their Aggregates. *J. Liq. Chromatogr. Relat. Technol.* **2012**, *35* (20), 2923–2950. <https://doi.org/10.1080/10826076.2012.743724>.
- (165) Hammes, G. G. *Spectroscopy for the Biological Sciences*; 2005. <https://doi.org/10.1002/0471733555>.
- (166) Kueltzo, L. A.; Ersoy, B.; Ralston, J. P.; Middaugh, C. R. Derivative Absorbance Spectroscopy and Protein Phase Diagrams as Tools for Comprehensive Protein Characterization: A BGCSF Case Study. *J. Pharm. Sci.* **2003**, *92* (9), 1805–1820. <https://doi.org/10.1002/jps.10439>.

- (167) Schmid, F.-X. Biological Macromolecules: UV-Visible Spectrophotometry. *Encycl. Life Sci.* **2001**, 1–4. <https://doi.org/10.1038/npg.els.0003142>.
- (168) Broide, M. L.; Berland, C. R.; Pande, J.; Ogun, O. O.; Benedek, G. B. Binary-Liquid Phase Separation of Lens Protein Solutions. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (13), 5660–5664. <https://doi.org/10.1073/pnas.88.13.5660>.
- (169) Biancalana, M.; Koide, S. Molecular Mechanism of Thioflavin-T Binding to Amyloid Fibrils. *Biochim. Biophys. Acta* **2010**, *1804* (7), 1405–1412. <https://doi.org/10.1016/j.bbapap.2010.04.001>.
- (170) Makin, O. S.; Serpell, L. C. Structures for Amyloid Fibrils. *FEBS J.* **2005**, *272* (23), 5950–5961. <https://doi.org/10.1111/j.1742-4658.2005.05025.x>.
- (171) Voropai, E. S.; Samtsov, M. P.; Kaplevskii, K. N.; Maskevich, A. A.; Stepuro, V. I.; Povarova, O. I.; Kuznetsova, I. M.; Turoverov, K. K.; Fink, A. L.; Uverskii, V. N. Spectral Properties of Thioflavin T and Its Complexes with Amyloid Fibrils. *J. Appl. Spectrosc.* **2003**, *70* (6), 868–874. <https://doi.org/10.1023/B:JAPS.0000016303.37573.7e>.
- (172) Harel, M.; Sonoda, L. K.; Silman, I.; Sussman, J. L.; Rosenberry, T. L. Crystal Structure of Thioflavin T Bound to the Peripheral Site of Torpedo Californica Acetylcholinesterase Reveals How Thioflavin T Acts as a Sensitive Fluorescent Reporter of Ligand Binding to the Acylation Site. *J. Am. Chem. Soc.* **2008**, *130* (25), 7856–7861. <https://doi.org/10.1021/ja7109822>.
- (173) Groenning, M.; Olsen, L.; van de Weert, M.; Flink, J. M.; Frokjaer, S.; Jørgensen, F. S. Study on the Binding of Thioflavin T to β -Sheet-Rich and Non- β -Sheet Cavities. *J. Struct. Biol.* **2007**, *158* (3), 358–369. <https://doi.org/10.1016/j.jsb.2006.12.010>.
- (174) Meehan, S.; Knowles, T. P. J.; Baldwin, A. J.; Smith, J. F.; Squires, A. M.; Clements, P.; Treweek, T. M.; Ecroyd, H.; Tartaglia, G. G.; Vendruscolo, M.; MacPhee, C. E.; Dobson, C. M.; Carver, J. A. Characterisation of Amyloid Fibril Formation by Small Heat-Shock Chaperone Proteins Human AA-, AB- and R120G AB-Crystallins. *J. Mol. Biol.* **2007**, *372* (2), 470–484. <https://doi.org/10.1016/j.jmb.2007.06.060>.
- (175) E. Van Holde, W. Curtis Johnson, P. Shing Ho, K. *Principles of Physical Biochemistry.* <https://doi.org/10.1021/ed076p474.1>.
- (176) Mertens, H. D. T.; Svergun, D. I. Structural Characterization of Proteins and Complexes Using Small-Angle X-Ray Solution Scattering. *J. Struct. Biol.* **2010**, *172* (1), 128–141. <https://doi.org/10.1016/j.jsb.2010.06.012>.
- (177) Makowski, L. Characterization of Proteins with Wide-Angle X-Ray Solution Scattering (WAXS). *J. Struct. Funct. Genomics* **2010**, *11* (1), 9–19. <https://doi.org/10.1007/s10969-009-9075-x>.
- (178) Tencé-Girault, S.; Lebreton, S.; Bunau, O.; Dang, P.; Bargain, F. Simultaneous SAXS-WAXS Experiments on Semi-Crystalline Polymers: Example of PA11 and Its Brill

- Transition. *Crystals* **2019**, *9* (5), 1–17. <https://doi.org/10.3390/cryst9050271>.
- (179) Jacques, D. A.; Trehwella, J. Small-Angle Scattering for Structural Biology - Expanding the Frontier While Avoiding the Pitfalls. *Protein Sci.* **2010**, *19* (4), 642–657. <https://doi.org/10.1002/pro.351>.
- (180) Sasview. SASview Spherical Form Factor Documentation https://www.sasview.org/docs/old_docs/3.1.2/user/models/model_functions.html
- (181) Sasview. Sticky hard-sphere model documentation <https://www.sasview.org/docs/user/models/stickyhardsphere.html>.
- (182) SASview. Ornstein-Zernike model documentation <https://www.sasview.org/docs/user/models/lorentz.html>.
- (183) Murphy, D. B.; Spring, K. R.; Fellers, T. J.; Davison, M. W. Principles of Birefringence Introduction to Optical Birefringence <https://www.microscopyu.com/techniques/polarized-light/principles-of-birefringence>.
- (184) Hammond, C. *The Basics of Crystallography and Diffraction*; 2009.
- (185) Drenth, J. *Principles of Protein X-Ray Crystallography*; 2007.
- (186) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oefner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwillinger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *D66* (0907–4449), 213–221. <https://doi.org/10.1107/S0907444909052925>.
- (187) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *D66*, 486–501. <https://doi.org/10.1107/S0907444910007493>.
- (188) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal Mol. Biol.* **2007**, *372*, 774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>.
- (189) Vendra, V. P. R.; Khan, I.; Chandani, S.; Muniyandi, A.; Balasubramanian, D. Gamma Crystallins of the Human Eye Lens. *Biochim. Biophys. Acta - Gen. Subj.* **2016**, *1860* (1), 333–343. <https://doi.org/10.1016/j.bbagen.2015.06.007>.
- (190) Ramkumar, S.; Fan, X.; Wang, B.; Yang, S.; Monnier, V. M. Reactive Cysteine Residues in the Oxidative Dimerization and Cu²⁺ Induced Aggregation of Human γ D-Crystallin: Implications for Age-Related Cataract. *Biochim. Biophys. Acta - Mol. Basis Dis.* **2018**, *1864* (11), 3595–3604. <https://doi.org/10.1016/j.bbadis.2018.08.021>.
- (191) Srikanthan, D.; Bateman, O. A.; Purkiss, A. G.; Slingsby, C. Sulfur in Human Crystallins. *Exp. Eye Res.* **2004**, *79*, 823–831. <https://doi.org/10.1016/j.exer.2004.05.009>.
- (192) Pande, A.; Pande, J.; Asherie, N.; Lomakin, A.; Ogun, O.; King, J. a; Lubsen, N. H.; Walton, D.; Benedek, G. B. Molecular Basis of a Progressive Juvenile-Onset Hereditary

- Cataract. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (5), 1993–1998. <https://doi.org/10.1073/pnas.040554397>.
- (193) Giblin, F. J. Glutathione a Vital Lens Antioxidant. *J. Ocul. Pharmacol. Ther.* **2000**, *16*, 121–135.
- (194) Rowley, D. A.; Halliwell, B. Superoxide-Dependent Formation of Hydroxyl Radicals in the Presence of Thiol Compounds. *FEBS Lett.* **1982**, *138*, 33–37. [https://doi.org/10.1016/0014-5793\(82\)80388-8](https://doi.org/10.1016/0014-5793(82)80388-8).
- (195) Schafheimer, N.; King, J. Tryptophan Cluster Protects Human γ D-Crystallin from Ultraviolet Radiation-Induced Photoaggregation in Vitro. *Photochem. Photobiol.* **2013**, *89* (5), 1106–1115. <https://doi.org/10.1111/php.12096>.
- (196) Serebryany, E.; Takata, T.; Erickson, E.; Schafheimer, N.; Wang, Y.; King, J. A. Aggregation of Trp>Glu Point Mutants of Human Gamma-D Crystallin Provides a Model for Hereditary or UV-Induced Cataract. *Protein Sci.* **2016**, *25* (6), 1115–1128. <https://doi.org/10.1002/pro.2924>.
- (197) Schafheimer, N.; Wang, Z.; Schey, K.; King, J. Tyrosine/Cysteine Cluster Sensitizing Human γ D-Crystallin to Ultraviolet Radiation-Induced Photoaggregation in Vitro. *Biochemistry* **2014**, *53* (6), 979–990. <https://doi.org/10.1021/bi401397g>.
- (198) Zhao, H.; Brown, P. H.; Magone, M. T.; Schuck, P. The Molecular Refractive Function of Lens γ -Crystallins. *J. Mol. Biol.* **2011**, *411* (3), 680–699. <https://doi.org/10.1016/j.jmb.2011.06.007>.
- (199) Mahendiran, K.; Elie, C.; Nebel, J. C.; Ryan, A.; Pierscionek, B. K. Primary Sequence Contribution to the Optical Function of the Eye Lens. *Sci. Rep.* **2014**, *4*, 20–23. <https://doi.org/10.1038/srep05195>.
- (200) Hanson, S. R. A.; Smith, D. L.; Smith, J. B. Deamidation and Disulfide Bonding in Human Lens γ -Crystallins. *Exp. Eye Res.* **1998**, *67* (3), 301–312. <https://doi.org/10.1006/exer.1998.0530>.
- (201) Wang, Y.; Lomakin, A.; McManus, J. J.; Ogun, O.; Benedek, G. B. Phase Behavior of Mixtures of Human Lens Proteins Gamma D and Beta B1. *Proc. Natl. Acad. Sci.* **2010**, *107* (30), 13282–13287. <https://doi.org/10.1073/pnas.1008353107>.
- (202) Basak, A.; Bateman, O.; Slingsby, C.; Pande, A.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. High-Resolution X-Ray Crystal Structures of Human γ D Crystallin (1.25 Å) and the R58H Mutant (1.15 Å) Associated with Aculeiform Cataract. *J. Mol. Biol.* **2003**, *328* (5), 1137–1147. [https://doi.org/10.1016/S0022-2836\(03\)00375-9](https://doi.org/10.1016/S0022-2836(03)00375-9).
- (203) James, S. *The Self-Assembly of Proteins: Probing Anisotropic Protein-Protein Interactions Using Phase Diagrams*, Maynooth University, 2015.
- (204) Strofaldi, A.; Khan, A. R.; Mcmanus, J. J. Surface Exposed Free Cysteine Suppresses Crystallization of Human γ D-Crystallin. *J. Mol. Biol.* **2021**, *433* (22), 167252.

<https://doi.org/10.1016/j.jmb.2021.167252>.

- (205) Vorontsova, M. A.; Chan, H. Y.; Lubchenko, V.; Vekilov, P. G. Lack of Dependence of the Sizes of the Mesoscopic Protein Clusters on Electrostatics. *Biophys. Journal* **2015**, *109* (9), 1959–1968. <https://doi.org/10.1016/j.bpj.2015.09.025>.
- (206) Maes, D.; Vorontsova, M. A.; Potenza, M. A. C.; Tiziano, S.; Sleutel, M.; Giglio, M.; Vekilov, P. G. Do Protein Crystals Nucleate within Dense Liquid Clusters? *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2015**, *F71*, 815–822. <https://doi.org/10.1107/S2053230X15008997>.
- (207) Vekilov, P. G. Phase Diagrams and Kinetics of Phase Transitions in Protein Solutions. *J. Phys. Condens. matter* **2012**, *24* (193101), 1–16. <https://doi.org/10.1088/0953-8984/24/19/193101>.
- (208) Islam, M. M.; Kobayashi, K.; Kidokoro, S. I.; Kuroda, Y. Hydrophobic Surface Residues Can Stabilize a Protein through Improved Water–Protein Interactions. *FEBS J.* **2019**, *286* (20), 4122–4134. <https://doi.org/10.1111/febs.14941>.
- (209) Grdadolnik, J.; Merzel, F.; Avbelj, F. Origin of Hydrophobicity and Enhanced Water Hydrogen Bond Strength near Purely Hydrophobic Solutes. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (2), 322–327. <https://doi.org/10.1073/pnas.1612480114>.
- (210) Head-Gordon, T.; Sorenson, J. M.; Pertsemlidis, A.; Glaeser, R. M. Differences in Hydration Structure near Hydrophobic and Hydrophilic Amino Acids. *Biophys. J.* **1997**, *73* (4), 2106–2115. [https://doi.org/10.1016/S0006-3495\(97\)78241-9](https://doi.org/10.1016/S0006-3495(97)78241-9).
- (211) Jamadagni, S. N.; Godawat, R.; Garde, S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2* (1), 147–171. <https://doi.org/10.1146/annurev-chembioeng-061010-114156>.
- (212) Fan, Y.; Joachimiak, A. Enhanced Crystal Packing Due to Solvent Reorganization through Reductive Methylation of Lysine Residues in Oxidoreductase from *Streptococcus Pneumoniae*. *J. Struct. Funct. Genomics* **2010**, *11* (2), 101–111. <https://doi.org/10.1007/s10969-010-9079-6>.
- (213) Parks, C.; Koswara, A.; Devilbiss, F.; Tung, H.; Nere, K. N.; Bordawekar, S.; Nagy, Z. K.; Ramkrishna, D. Solubility Curves and Nucleation Rates from Molecular Dynamics for Polymorph Prediction – Moving beyond Lattice Energy Minimization. *Phys. Chem. Chem. Phys.* **2017**, *19*, 5285–5295. <https://doi.org/10.1039/c6cp07181c>.
- (214) Shentu, X.; Yao, K.; Xu, W.; Zheng, S.; Hu, S.; Gong, X. Special Fasciculiform Cataract Caused by a Mutation in the GammaD-Crystallin Gene. *Mol. Vis.* **2004**, *10* (March 2004), 233–239.
- (215) Pande, A.; Ghosh, K. S.; Banerjee, P. R.; Pande, J. Increase in Surface Hydrophobicity of the Cataract-Associated P23T Mutant of Human γ D-Crystallin Is Responsible for Its Dramatically Lower, Retrograde Solubility. *Biochemistry* **2010**, *49* (29), 6122–6129.

<https://doi.org/10.1021/bi100664s>.

- (216) Vekilov, P. G.; Feeling-Taylor, A. R.; Yau, S. T.; Petsev, D. Solvent Entropy Contribution to the Free Energy of Protein Crystallization. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2002**, *58*, 1611–1616. <https://doi.org/10.1107/S0907444902014312>.
- (217) Pande, A.; Zhang, J.; Banerjee, P. R.; Puttamadappa, S. S.; Shekhtman, A.; Pande, J. NMR Study of the Cataract-Linked P23T Mutant of Human γ D-Crystallin Shows Minor Changes in Hydrophobic Patches That Reflect Its Retrograde Solubility. *Biochem. Biophys. Res. Commun.* **2009**, *382* (1), 196–199. <https://doi.org/10.1016/j.bbrc.2009.03.007>.
- (218) Vetri, V.; Foderà, V. The Route to Protein Aggregate Superstructures : Particulates and Amyloid-like Spherulites. *FEBS Lett.* **2015**, *589*, 2448–2463. <https://doi.org/10.1016/j.febslet.2015.07.006>.
- (219) Crist, B.; Schultz, J. M. Polymer Spherulites : A Critical Review. *Prog. Polym. Sci.* **2016**, *56*, 1–63. <https://doi.org/10.1016/j.progpolymsci.2015.11.006>.
- (220) Heijna, M. C. R.; Theelen, M. J.; Van Enckevort, W. J. P.; Vlieg, E. Spherulitic Growth of Hen Egg-White Lysozyme Crystals. *J. Phys. Chem. B* **2007**, *111* (7), 1567–1573. <https://doi.org/10.1021/jp0643294>.
- (221) Krebs, M. R. H.; Devlin, G. L.; Donald, A. M. Protein Particulates: Another Generic Form of Protein Aggregation? *Biophys. J.* **2007**, *92* (4), 1336–1342. <https://doi.org/10.1529/biophysj.106.094342>.
- (222) Bromley, E. H. C.; Krebs, M. R. H.; Donald, A. M. Mechanisms of Structure Formation in Particulate Gels of β -Lactoglobulin Formed near the Isoelectric Point. *Eur. Phys. J. E* **2006**, *21* (2), 145–152. <https://doi.org/10.1140/epje/i2006-10055-7>.
- (223) Krebs, M. R. H.; Devlin, G. L.; Donald, A. M. Amyloid Fibril-like Structure Underlies the Aggregate Structure across the PH Range for β -Lactoglobulin. *Biophys. J.* **2009**, *96* (12), 5013–5019. <https://doi.org/10.1016/j.bpj.2009.03.028>.
- (224) Krebs, M. R. H.; Bromley, E. H. C.; Rogers, S. S.; Donald, A. M. The Mechanism of Amyloid Spherulite Formation by Bovine Insulin. *Biophys. J.* **2005**, *88* (3), 2013–2021. <https://doi.org/10.1529/biophysj.104.051896>.
- (225) Foderà, V.; Vetri, V.; Wind, T. S.; Noppe, W.; Cornett, C.; Donald, A. M.; Morozova-Roche, L. A.; Vestergaard, B. Observation of the Early Structural Changes Leading to the Formation of Protein Superstructures. *J. Phys. Chem. Lett.* **2014**, *5* (18), 3254–3258. <https://doi.org/10.1021/jz501614e>.
- (226) Juárez, J.; Taboada, P.; Goy-López, S.; Cambón, A.; Madec, M. B.; Yeates, S. G.; Mosquera, V. Additional Supra-Self-Assembly of Human Serum Albumin under Amyloid-like-Forming Solution Conditions. *J. Phys. Chem. B* **2009**, *113* (36), 12391–12399. <https://doi.org/10.1021/jp904167e>.

- (227) Pande, A.; Annunziata, O.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. Decrease in Protein Solubility and Cataract Formation Caused by the Pro23 to Thr Mutation in Human γ D-Crystallin. *Biochemistry* **2005**, *44* (7), 2491–2500. <https://doi.org/10.1021/bi0479611>.
- (228) Matthews, B. W. Determination of Molecular Weight from Protein Crystals. *J. Mol. Biol.* **1974**, *82* (4), 513–526. [https://doi.org/10.1016/0022-2836\(74\)90245-9](https://doi.org/10.1016/0022-2836(74)90245-9).
- (229) Altan, I.; Khan, A. R.; James, S.; Quinn, M. K.; McManus, J. J.; Charbonneau, P. Using Schematic Models to Understand the Microscopic Basis for Inverted Solubility in γ D-Crystallin. *J. Phys. Chem. B* **2019**, *123* (47), 10061–10072. <https://doi.org/10.1021/acs.jpcc.9b07774>.
- (230) Quinn, M. K. Alterations in the Phase Behaviour of Human Γ D-Crystallin Due to Mutagenesis and Chemical Modification of the Protein Surface, Maynooth University, 2017.
- (231) Krebs, M. R. H.; Domike, K. R.; Donald, A. M. Protein Aggregation: More than Just Fibrils. *Biochem. Soc. Trans.* **2009**, *37* (4), 682–686. <https://doi.org/10.1042/BST0370682>.
- (232) VanOudenhove, J.; Anderson, E.; Krueger, S.; Cole, J. L. Analysis of PKR Structure by Small-Angle Scattering. *J. Mol. Biol.* **2009**, *387* (4), 910–920. <https://doi.org/10.1016/j.jmb.2009.02.019>.
- (233) Mo, Z.; Zhang, H. The Degree of Crystallinity in Polymers by Wide-Angle X-Ray Diffraction (WAXD). *J. Macromol. Sci. Part C Polym. Rev.* **1995**, *35*(4), 555–580. <https://doi.org/10.1080/15321799508021751>.
- (234) Domike, K. R.; Donald, A. M. Kinetics of Spherulite Formation and Growth: Salt and Protein Concentration Dependence on Proteins β -Lactoglobulin and Insulin. *Biomacromolecules* **2007**, *8*, 3930–3937. <https://doi.org/10.1016/j.ijbiomac.2008.12.014>.
- (235) Evans, P.; Wyatt, K.; Wistow, G. J.; Bateman, O. A.; Wallace, B. A.; Slingsby, C. The P23T Cataract Mutation Causes Loss of Solubility of Folded γ D-Crystallin. *J. Mol. Biol.* **2004**, *343* (2), 435–444. <https://doi.org/10.1016/j.jmb.2004.08.050>.
- (236) Boatz, J. C.; Whitley, M. J.; Li, M.; Gronenborn, A. M.; Wel, P. C. A. Van Der. Cataract-Associated P23T GD-Crystallin Retains a Native-like Fold in Amorphous-Looking Aggregates Formed at Physiological PH. *Nat. Commun.* **2017**, *8* (15137), 1–10. <https://doi.org/10.1038/ncomms15137>.
- (237) Ng, C. Z.; Lean, Y. L.; Yeoh, S. F.; Lean, Q. Y.; Lee, K. S.; Suleiman, A. K.; Liew, K. Bin; Kassab, Y. W.; Al-Worafi, Y. M.; Ming, L. C. Cold Chain Time-and Temperature-Controlled Transport of Vaccines: A Simulated Experimental Study. *Clin. Exp. Vaccine Res.* **2020**, *9* (1), 8–14. <https://doi.org/10.7774/cevr.2020.9.1.8>.
- (238) Bogale, H. A.; Amhare, A. F.; Bogale, A. A. Assessment of Factors Affecting Vaccine Cold Chain Management Practice in Public Health Institutions in East Gojam Zone of

- Amhara Region. *BMC Public Health* **2019**, *19* (1), 1–6. <https://doi.org/10.1186/s12889-019-7786-x>.
- (239) Ausar, S. F.; Foubert, T. R.; Hudson, M. H.; Vedvick, T. S.; Middaugh, C. R. Conformational Stability and Disassembly of Norwalk Virus-like Particles: Effect of PH and Temperature. *J. Biol. Chem.* **2006**, *281* (28), 19478–19488. <https://doi.org/10.1074/jbc.M603313200>.
- (240) Vragliau, C.; Bufton, J. C.; Garzoni, F.; Stermann, E.; Rabi, F.; Terrat, C.; Guidetti, M.; Josserand, V.; Williams, M.; Woods, C. J.; Viedma, G.; Bates, P.; Verrier, B.; Chaperot, L.; Schaffitzel, C.; Berger, I.; Fender, P. Synthetic Self-Assembling ADDomer Platform for Highly Efficient Vaccination by Genetically Encoded Multiepitope Display. *Sci. Adv.* **2019**, *5* (9), 1–8. <https://doi.org/10.1126/sciadv.aaw2853>.
- (241) Besson, S.; Vragliau, C.; Vassal-Stermann, E.; Dagher, M. C.; Fender, P. The Adenovirus Dodecahedron: Beyond the Platonic Story. *Viruses* **2020**, *12* (7). <https://doi.org/10.3390/v12070718>.
- (242) Sari-Ak, D.; Bufton, J.; Gupta, K.; Garzoni, F.; Fitzgerald, D.; Schaffitzel, C.; Berger, I. VLP-Factory™ and ADDomer©: Self-Assembling Virus-Like Particle (VLP) Technologies for Multiple Protein and Peptide Epitope Display. *Curr. Protoc.* **2021**, *1* (3), 1–25. <https://doi.org/10.1002/cpz1.55>.
- (243) Szolajaska, E.; Burmeister, W. P.; Zochowska, M.; Nerlo, B.; Andreev, I.; Schoehn, G.; Andrieu, J. P.; Fender, P.; Naskalska, A.; Zubieta, C.; Cusack, S.; Chroboczek, J. The Structural Basis for the Integrity of Adenovirus Ad3 Dodecahedron. *PLoS One* **2012**, *7* (9), 1–11. <https://doi.org/10.1371/journal.pone.0046075>.
- (244) Cizmar, P.; Yuana, Y. Detection and Characterization of Extracellular Vesicles by Transmission and Cryo-Transmission Electron Microscopy. In *Methods in molecular biology (Clifton, N.J.)*; 2017; pp 221–232. <https://doi.org/10.1007/978-1-4939-7253-1>.
- (245) Hall, M.; Rubin, J.; Behrens, S. H.; Bommarius, A. S. The Cellulose-Binding Domain of Cellobiohydrolase Cel7A from *Trichoderma Reesei* Is Also a Thermostabilizing Domain. **2011**, *155*, 370–376. <https://doi.org/10.1016/j.jbiotec.2011.07.016>.
- (246) Rexroad, J.; Martin, T. T.; McNeilly, D.; Godwin, S.; Middaugh, C. R. Thermal Stability of Adenovirus Type 2 as a Function of PH. *J. Pharm. Sci.* **2006**, *95* (7), 1469–1479. <https://doi.org/10.1002/jps.20617>.
- (247) Brisco, M.; Haniff, C.; Wilson, T. M. A.; Sattelle, D. B. The Kinetics of Swelling of Southern Bean Mosaic Virus: A Study Using Photon Correlation Spectroscopy. *Virology* **1986**, *148*, 218–220. [https://doi.org/10.1016/0042-6822\(86\)90417-4](https://doi.org/10.1016/0042-6822(86)90417-4).
- (248) Rexroad, J.; Wiethoff, C. M.; Green, A. P.; Kierstead, T. D.; Scott, M. O.; Middaugh, C. R. Structural Stability of Adenovirus Type 5. *J. Pharm. Sci.* **2003**, *92* (3), 665–678. <https://doi.org/10.1002/jps.10340>.

- (249) Duy, C.; Fitter, J. How Aggregation and Conformational Scrambling of Unfolded States Govern Fluorescence Emission Spectra. *Biophys. J.* **2006**, *90* (10), 3704–3711. <https://doi.org/10.1529/biophysj.105.078980>.
- (250) Atri, M. S.; Saboury, A. A.; Moosavi-Movahedi, A. A.; Goliaei, B.; Sefidbakht, Y.; Alijanvand, H. H.; Sharifzadeh, A.; Niasari-Naslaji, A. Structure and Stability Analysis of Cytotoxic Complex of Camel α -Lactalbumin and Unsaturated Fatty Acids Produced at High Temperature. *J. Biomol. Struct. Dyn.* **2011**, *28* (6), 919–928. <https://doi.org/10.1080/07391102.2011.10508618>.
- (251) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2006**, *1* (6), 2876–2890. <https://doi.org/10.1038/nprot.2006.202>.
- (252) Duret, L.; Gasteiger, E.; Perrière, G. LALNVIEW: A Graphical Viewer for Pairwise Sequence Alignments. *CABIOS* **1996**, *12* (6), 507–510.
- (253) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydrophobic Character of a Protein. *J. Mol. Biol.* **1982**, *157* (1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0).