EXPERIMENTAL CHARACTERISATION OF PROTEIN-PROTEIN AND PROTEIN-EXCIPIENT INTERACTIONS AND THEIR IMPACT ON PROTEIN CONFORMATION AND AGGREGATION

A Thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Science and Engineering

2020

MATJA ZALAR

SCHOOL OF NATURAL SCIENCES

Department of Chemistry

Table of Contents

WORD COUNT: 62386

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

Abbreviations

1D	one dimensional
2D	two dimensional
3D	three dimensional
^{1}H	proton
^{2}H	deuterium
$^{2}H_2O$	deuterated water
^{13}C	carbon -13
^{15}N	nitrogen -15
A	absorbance
AF^{STD}	STD amplification factor
AIR	ambiguous intermolecular restraints
Arg	arginine
ArgGlu	arginine glutamate
ArgHCl	arginine hydrochloride
AUC	analytical ultracentrifugation
B_{22}	second virial coefficient
BCM	barycentric mean
BCS	biopharmaceutics classification system
c	concentration
CD	circular dichroism
CDR	complementarity-determining region
C_H	constant region of the heavy chain
C_L	constant region of the light chain
CLEANEX-PM	phase-modulated CLEAN chemical exchange
CPMG	Carr-Purcell-Meiboom-Gill
CSP	chemical shift perturbation
D	apparent coefficient of self-diffusion
DLS	dynamic light scattering
DS	domain-swapped
DSC	differential scanning calorimetry
DSF	differential scanning fluorimetry
ELISA	enzyme linked immunosorbent assay
FBDD	fragment based drug design
FT	Fourier transform
F-W	Finke-Watzky
G	gradient strength
GB1	B1 domain of the immunoglobulin G binding protein
Glu	glutamate
Gly	glycine

H-D	proton-deuterium exchange
HPLC	high pressure liquid chromatography
HSA	human serum albumin
HSQC	heteronuclear single quantum coherence
I_f/I_0	residue-specific ratios of signal intensities at the end and at the beginning of the experiment
Ig	immunoglobulin
IgG	immunoglobulin G
IR	infrared
ITC	isothermal calorimetry
k_{agg}	rate of aggregation
K_d	dissociation constant
k_D	mutual diffusion coefficient
k_{ex}	intrinsic proton-deuterium exchange rates
k_{HD}	rate of H-D exchange
k_{nuc}	rate of nucleation
L1	loop 1
L2	loop 2
LB	lysogeny broth
mAb	monoclonal antibody
MD	molecular dynamics
MM-GBSA	molecular mechanics generalized Born solvent area
NDS	non-domain-swapped
NEP	neprilisyn
NH	amide group
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NUS	non-uniform sampling
P	protection factor
PDI	polydispersity index
Pro	proline
PTM	post-translational modifications
RF	radiofrequency
R_h	the apparent hydrodynamic radius
RI	refractive index
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
SASA	solvent accessible surface area
scFv	single-chain fragments of variability
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SEC-MALS	SEC coupled with multi-angle light scattering
SLS	static light scattering

SPR	surface plasmon resonance
SQT	Stefin A quadruple mutant Tracy
SQT-1	SQT variant with AU1 and c-Myc insertion peptides
SQT-1C	SQT-1 with cleavable histidine tag at the C-terminus
SQT-1N	SQT-1 with cleavable histidine tag at the N-terminus
STD	saturation transfer difference
SteA	stefin A
SteB	stefin B
T	temperature
T_1	longitudinal relaxation time
T_2	transverse relaxation time
T_{agg}	onset of aggregation temperature
T_m	melting temperature
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimised spectroscopy
UV	ultraviolet
UV-Vis	ultraviolet-visible
V_H	variable domain of the heavy chain
V_L	variable domain of the light chain
WaterLOGSY	water-ligand observed via gradient spectroscopy
WT	wildtype
Γ	gyromagnetic ratio
Δ	diffusion time
Δ	gradient length and chemical shift
τ_C	rotational correlation time

Abstract

In the last decades, biopharmaceutical proteins, especially monoclonal antibodies (mAbs), have become one of the fastest growing classes of pharmaceuticals, due to their high affinity, specificity, non-toxicity, and low immunogenicity. Despite their versatility and wide applicability, their production is costly and poses unique challenges due to their high molecular weight and structural complexity. One of the major challenges in formulation development is thus prevention of protein self-association that may lead into formation of potentially immunogenic aggregates. To achieve this, excipients are often added in relatively high concentrations to the final formulations. However, there is a lack of information available on the impact of co-solutes, including excipients on protein stability. In light of these inherent problems of mAbs, alternative systems, such as protein aptamers, have been developed for high affinity binders for specific targets. Some advantages of peptide aptamers are their smaller size, versatility, specificity for chosen targets, and ease of production. However, the insertion of binding loops may cause destabilisation of protein, leading to formation of inactive oligomeric species.

In the first part of the Thesis we have used a stefin A derived scaffold protein, SQT, as a model to study solution behaviour and self-association of such systems. We have determined its crystal structure, and investigated possible pathways of oligomer formation. We studied the oligomerisation kinetics and determined the limiting steps of dimer and tetramer formation. Using mutagenesis, we created a more stable variant of the SQT scaffold. In the second part of this Thesis we have applied solution NMR spectroscopy to comprehensively evaluate protein-excipient interactions between seven biotherapeutically relevant proteins and a set of eleven commonly used excipients. Additionally, we evaluated the effect of excipients on thermal and colloidal protein stability, the aggregation kinetics at elevated temperatures and on the protein storage stability at accelerated conditions. In the last part of the Thesis we have combined both protein and ligand-observed NMR approaches to evaluate interactions between the stabilised SQT variant and various excipients. We used isotopically labelled protein to evaluate the potential interaction sites on the protein surface and estimated the binding constants for individual excipients. Furthermore, we compared the sensitivity of both protein- and ligand-observed methods and discussed how they could be applied to biopharmaceutical formulations.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Acknowledgements

Undertaking this PhD has been an exciting journey that has allowed me to grow professionally and personally. None of it would have been possible without the help and support of the people I have encountered on this journey.

First and foremost, I would like to thank my supervisor Alexander P. Golovanov for the continuous feedback, support and encouragement throughout my PhD. Thank you for your patient guidance, useful critiques and motivational words needed to complete the project. I would also like to thank Jim Warwicker, Robin Curtis and all other members of the protein aggregation group for letting me use their research facilities, and providing helpful feedback on my work. Furthermore, I would like to express my great appreciation to all members of the PIPPI consortium, especially to Hristo Svilenov, Gerhard Winter, Sowmya Indrakumar, Pernille Harris, and Gunther H.J. Peters for hosting me during my secondments and allowing me to gain experience in their labs. I am also very grateful for the opportunity to host PIPPI fellows in our lab which reminded me why I love research and helped me decide what the next step in my career should be.

Big thanks go to all research staff at the MIB and UoM Bioanalysis Facility for providing initial training and technical support with running various instruments. I would especially like to thank Colin Levy for going beyond traditional approaches to solve the crystal structure and providing constant guidance during the process. My special thanks go to our NMR facility manager, Matt Cliff, for teaching me various aspects of NMR theory and spectrometer maintenance, skills I never thought I would obtain.

I am also grateful to other members of the Golovanov group I had the pleasure to work with, Richard Tunnicliffe, John Edwards and Jack Bramham. Richard, thank you for introducing me to protein biochemistry, for sharing your little tips and tricks in the lab, and providing advice whenever needed. Jack, thank you very much for your help with automating my analysis and patiently debugging scripts I have butchered beyond recognition.

I would also like to thank the MIB lunch crew, my dear friends and colleagues, Aisling Roche, Nikita Vekaria, Jas Kalayan and Nicole Sibanda who were of great support in deliberating over our problems and findings as well as providing happy distractions during lunch breaks and outside of work.

Completing this PhD would not have been possible without the support of my friends and family. I am grateful for every one of them and it means a world to me that we maintained our friendship even though we are scattered all over Europe. Andreja Bratovš, Eva Lučka Kozak, Tamara Marić, and Urška Slapšak, thank you for always being just a call away, your endless encouraging words and the adventures we shared on our weekend getaways.

To my family, Maja, Borut and Vita; Being away from you was the hardest part of this PhD journey but your unconditional support and unprecedented belief in my capabilities kept me going at my darkest moments. Thank you for reminding me of what matters most and sending little pieces of home to Manchester. You are, and always will be, my safe haven.

Trudie, thank you for welcoming me into your family with an open heart and giving me a second home. I will always be grateful for all the kind words and the care packages you sent when I needed them most.

Last but certainly not least, Jack. Thank you for always being there for me, especially during the rough patches. Thank you for planning all those bike rides that kept me sane and for always putting a smile on my face. Your unconditional support and encouragement throughout my PhD, as well as personal life, helped me to get to this point. Thank you.

Alternative format for submission and authorship details

Authors in alphabetical order:

A.P.G. – Alexander P. Golovanov, S.I. – Sowmya Indrakumar, C.W.L. – Colin W. Levy, G.H.J.P. – Günther H.J. Peters, H.L.S. – Hristo L. Svilenov, R.B.T – Richard B. Tunnicliffe, M.Z. – Matja Zalar

Chapter 2: Studies of the oligomerisation mechanism of a cystatin-based engineered protein scaffold

M.Z. prepared the protein samples, performed the experiments, evaluated the data and wrote the first draft of the manuscript. C.W.L assisted with setting up the crystallisation trials, acquired and processed the x-ray data, finalised the crystal structure refinement and deposited it to the PBD database. R.B.T. made valuable contributions to increase the extent of the NMR assignments. S.I. assisted with setting up molecular dynamics simulations and provided guidance at data analysis. M.Z., A.P.G and G.H.J.P. conceived the presented work and planned the experiments. A.P.G and G.H.J.P. provided conceptual guidance. All authors contributed to paper writing.

Chapter 3: New disulphide bond in cystatin-based protein scaffold prevents domain-swapped-mediated oligomerisation and stabilised the functionally active form

M.Z. prepared all protein samples, performed the experiments, analysed the data and wrote the first draft of the manuscript. M.Z. and A.P.G. contributed to planning the experiments. A.P.G. provided conceptual guidance and made significant contributions to the final draft of the manuscript.

Chapter 4: Binding of excipients is a poor predictor for aggregation kinetics of biopharmaceutical proteins

M.Z. prepared all samples, performed the experiments, analysed the data and wrote the first draft of the manuscript. H.L.S. assisted at setting up protein stability screen. H.L.S. and A.P.G provided guidance at data analysis. M.Z., H.L.S. and A.P.G conceived the presented work and planned the experiments. A.P.G provided conceptual guidance and resources.

Chapter 5: Use of protein- and ligand-observed NMR to comprehensively study protein-excipient interactions

M.Z. prepared all protein samples, performed the experiments, analysed the data and wrote the first draft of the manuscript. A.P.G. provided conceptual guidance and contributed to writing of the manuscript.

1 General Introduction

The general introduction will begin with an overview of the structural biology of immunoglobulin G (IgG) proteins and their use as biotherapeutics, and continue with an introduction of engineered scaffold proteins as an alternative to monoclonal antibodies (mAbs). The introduction will then consider protein physical instabilities leading to the formation of oligomers, aggregates, and fragments, and introduce potential strategies to mitigate aggregation. Next, we will demonstrate examples of the use of NMR in drug development, and protein formulation, followed by an explanation of specific NMR methods relevant to the work in this Thesis. Finally, the aims and objectives of the Thesis will be introduced.

1.1 Biopharmaceutical proteins

1.1.1 Overview

Since approval of recombinant insulin in 1982[1], biopharmaceuticals have been on the rise, with 316 biopharmaceutical products on the market in the EU and US in 2018[2]. Biopharmaceuticals are 'a protein- or nucleic acid-based pharmaceutical substance used for therapeutic or *in vivo* diagnostic purposes, which is produced by means other than direct extraction from a native (non-engineered) biological source[3]. According to the Biopharmaceutics Classification System (BCS), they are divided into four classes based on their solubility and permeability[4]. Alternatively, they can be grouped into three major categories: nucleic acids, proteins, and cell therapies[5], with proteins representing the majority of approved therapeutics[2].

Biopharmaceuticals differ from traditional small-molecule drugs in various aspects, including product sources, composition, structure, and identity of active substances, manufacturing methods and equipment needed, as well as in intellectual property and regulations[6]. Traditionally drugs are mostly small molecules or other chemical substances, synthesised from chemical precursors using standardised chemical processes. Their production is highly consistent, and the purity of the final product can be easily analysed and demonstrated at the atomic level. Due to their structural simplicity, substances with high purity are assumed to be similar or even identical for all practical purposes, even if they are produced via different synthesis protocols[6]. In contrast, biopharmaceuticals are produced in

biological systems and are much larger molecules which exhibit various levels of structural complexity. The main advantages of biopharmaceuticals over small molecule drugs are their high specificity and activity at relatively low concentrations[7-8]. Also, biopharmaceuticals are in principle less toxic than small molecule drugs as they degrade to amino acids and sugars, but form aggregates that may be immunogenic[9-10].

1.1.2 Monoclonal antibodies

Monoclonal antibodies (mAbs) are large blood plasma proteins involved in immune response against infection by pathogenic organisms. They belong to the immunoglobulin (Ig) superfamily, and consist of two heavy and two light chains. Based on the amino acid sequence of the heavy chain, antibodies can be divided into five isotypes, namely IgG, IgM, IgA, IgE and IgD (Figure 1.1)[11]. While the function of IgD antibodies is mostly unknown, IgM antibodies are involved in a primary immune response, IgA molecules are integral at protecting mucosal surfaces against pathogens, and IgE molecules are crucial in allergic reactions[12]. IgG is the most abundant of the antibodies classes, representing 75% of all antibodies in the serum. IgG molecules are involved in secondary immune response as well as in neutralisation of pathogens [12].

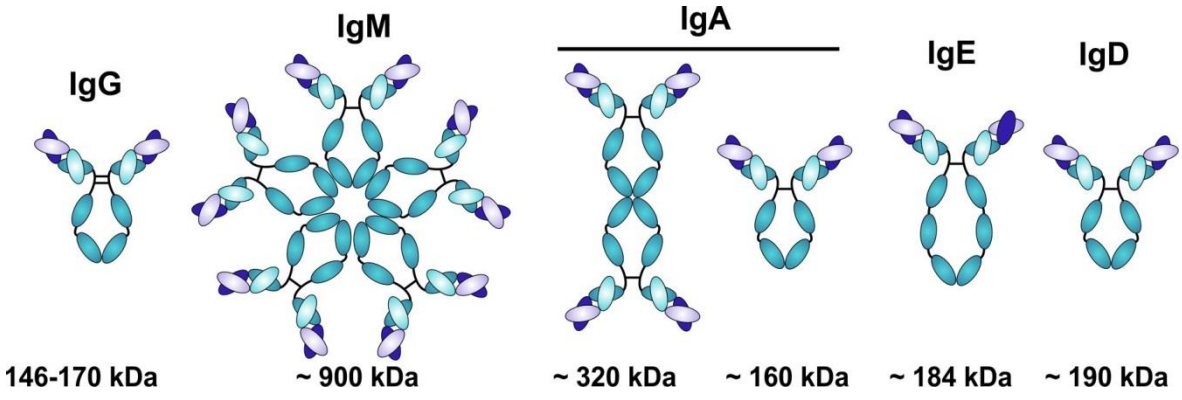

Figure 1.1: Different classes of antibodies

Monomeric IgGs (~150 kDa) consists of four polypeptide chains: two identical heavy (50 kDa) and two identical light (25 kDa) chains linked together by disulfide bonds forming the typical 'Y'-shaped structure (Figure 1.2). Binding of antigens occurs at Fab regions while the Fc fragment has an effector function, triggering a highly specific immune response[12].

The effector function depends on glycosylation patterns[13-14] which may also affect antibody structure, thus requiring careful consideration during drug development[15]. The heavy chain comprises four Ig fold domains: three constant regions (C_H1, C_H2 and C_H3) and a variable domain (V_H). The light chain has two Ig fold domains: a constant (C_L) and a variable (V_L) domain. The variable region of the Fab fragment contains three complementarity-determining regions (CDRs) responsible for the binding of the antigen. High variability of these regions ensures the specificity and diversity of antibodies[16].

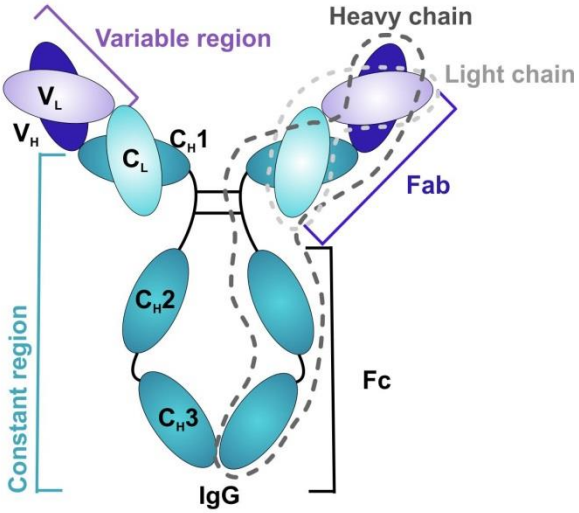

Figure 1.2: Schematic structure of an IgG molecule

The first monoclonal antibodies were produced *in vitro* in 1975, with the first mAb therapeutic approved in the US in 1986 – a murine mAb used to prevent kidney transplant rejection[17]. It was, however, removed from the market due to issues related to short half-life in humans, reduced efficacy and severe immunogenicity[18]. Since then more than 80 mAb-based biopharmaceuticals have been approved in the EU and US. They are used to treat several types of cancer, rheumatoid arthritis, Crohn's disease, lupus and other diseases[2]. Monoclonal antibodies now represent the fastest growing class of biopharmaceutics with market value projected to reach USD 131.766 billion by 2023[19]. Moreover, antibodies are also the most commonly used molecules for affinity binding studies, immunoprecipitation and ELISA assays, as well as uses as bio-imaging agents and diagnostics tools[20]. Despite their versatility and broad applicability, production of antibodies is costly and poses many challenges due to their high molecular weight and structural complexity, as well as chemical and physical instabilities that often lead to protein degradation or formation of immunogenic aggregates[21-23]. Production of monoclonal antibodies is also associated with high

development and production costs, resulting in expensive treatments and reagents. The origin of these instabilities and strategies to prevent them will be discussed further in section 1.2.2. In light of these inherent problems of mAbs, alternative approaches utilising recombinant DNA technology and offering design of molecules with improved specificity and efficacy with better reproducibility and lower production costs have been developed. The first class of antibody alternatives are recombinant antibody fragments, such as Fab fragments[24], single-chain fragments of variability (scFv)[25-27], nanobodies[28], diabodies[29], bi-specific and tri-specific antibodies[29-31]. The second big class of monoclonal antibody alternatives are non-immunoglobulin protein scaffolds[32-33] which will be discussed further in the next section. To date, more than 50 structurally diverse non-immunoglobulin scaffolds have been developed[34-35].

1.1.3 Engineered scaffolds as an alternative to monoclonal antibodies

Scaffold proteins, also known as peptide aptamers, consist of short specific amino acid sequences placed within a small and stable backbone, the 'scaffold'. Scaffold proteins have to be non-toxic, rigid, stable, soluble, and preferably monomeric proteins of known structure. Furthermore, their folding should not be affected by the insertions of various peptides, and, most importantly, they should not interact with any other cellular proteins or influence any biological pathways on their own[36]. As peptide aptamers can specifically bind to the target protein with high affinity and modulate its activity, they are useful tools in studies of protein function[37-38] and molecular interactions [39] within complex networks, as well as bio-imaging agents[40], diagnostic tools[41], biosensors[42-44] and as aids in the development of small molecule drugs[45] and biotherapeutics[46].

Activity of the scaffold-based binding proteins very much depends on maintaining the structure and solution properties of the base scaffold. Therefore, the conformation and solubility of scaffolds should not be affected by peptide insertions that are typically up to 15 residues long. This issue is well recognised; with most scaffolds engineered to improve their structural robustness and thermodynamic stability. Point mutations are often implemented to improve structural stability and resistance towards chemical denaturation[47], while terminal peptide extensions are used to improve the solubility of scaffolds[48]. Despite improvements in the design of scaffold proteins, insertion of specific peptides, or peptides outside a defined length range can significantly perturb scaffold structure[36, 49-50]. Structural

analysis of scaffold-based binding proteins revealed that while most of them interact with their binding partner only through inserted loops, alternative binding sites may be present, especially where the topology of inserted loops is not complementary to the binding interface of the target protein[50-52]. While the design of scaffold-based binding proteins appears relatively straightforward, instabilities introduced by engineering for target binding are less well understood, and therefore a better understanding of the problems associated with their design is required.

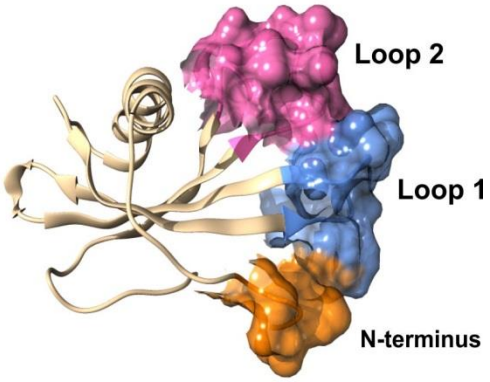

Figure 1.3: Model of SQT with insertion sites

The first scaffold protein was designed by inactivation of thioredoxin from E. *coli* [37], followed by GFP and staphylococcal nuclease based scaffolds[53], affibodies[54], DARPins[55] adhirons[50] and affimers, a group of human stefin A (SteA) based scaffolds; STM[36], SQM[51] and SQT[56]. SteA is a monomeric, single domain, 11 kDa protein that has been studied extensively as an inhibitor of cathepsins. Its structure has been determined both by NMR[57] and x-ray crystallography[58]. Its binding interface includes three independent peptides: loop one, loop two and N-terminus[58-59]. The first in the SteA based series of scaffold proteins was STM. Mutations of three amino acids crucial for its inhibitor activity neutralised its biological function. Furthermore, an insertion site for short peptides into loop 2 was introduced[36]. SQM scaffold was an improved version of STM that enabled the replacement of SteA loop 2 with desired peptide as well as insertions into loop 1 and the N-terminus, which provided a scaffold for production of a large variety of peptide aptamers. However, peptide insertions into the N-terminus severely impacted the folding of SQM, while its stability was adversely affected upon longer insertions into the loop regions[51]. In the SQT variant (Figure 1.3) of SteA based scaffolds, reduction of the number of SteA amino acids

replaced upon peptide insertion into the loop 2 significantly improved the overall stability of the protein, and rendered insertion at all three insertion sites possible individually or simultaneously[56]. Affimers and adhirons have been patented by Avacta and are commercially available. Recently, they have been developed for various applications, including binding to BCL-2 family proteins[60], toxin neutralization[61], visualisation of ubiquitination[62], diagnosis of plant viruses[63] and as affinity reagents[64-65].

1.2 Stability of proteins

1.2.1 Defining protein stability

Protein stability is an essential consideration in the development of both antibody biopharmaceuticals and scaffold proteins. Protein stability is a broad term often used to describe different phenomena. It has various aspects that can be, in general, divided into chemical and physical stability. Chemical stability reports on the propensity of proteins to undergo chemical modifications, while physical stability reports on the ability of the proteins to remain in their stable form under defined conditions. In this thesis, we have primarily focused on physical stability, with stabilisation of proteins against chemical changes outside the scope of this work and as such discussed only briefly. We have therefore divided the term of protein physical stability into thermal, colloidal and conformational stability. Conformational stability is defined as the ability of the protein to remain in its folded active conformation, without converting into unfolded inactive states[66]. Thermal stability refers to protein stability in terms of ability to resist denaturation in response to temperature changes[67]. It is measured primarily in terms of the apparent melting temperatures (T_m). Colloidal stability refers to the propensity of a protein to interact with the neighbouring protein molecules. It reflects in the apparent aggregation temperatures (T_{agg}), the second virial coefficient (B_{22}), and the mutual diffusion coefficient (k_D)[68].

1.2.2 Challenges of biopharmaceutical protein formulation

The major challenges during bioprocessing and formulation of biopharmaceuticals occur due to their large molecular weight, complex structure, structural instabilities and susceptibility to degradation and aggregation. In proteins, each amino acid residue represents a possible site of structural variation that affects not only its primary, but also higher orders

of structure. Structural changes and post-translational modifications (PTMs) that vary between batches may alter function, efficiency, and safety-related characteristics of biopharmaceuticals. One of the biggest concerns in antibody production and formulation is the formation of immunogenic aggregates. Though the immune response to biopharmaceuticals may be without any clinical impact, it may result in decreased efficacy, neutralisation of the endogenous counterpart or even in life-threatening severe immune system effects[9].

Protein degradation may occur due to chemical or physical instabilities. Chemical instabilities lead to new chemical entities. The most common chemical instabilities reported for therapeutic proteins involve deamination of asparagine residues; oxidation of histidine, methionine, cysteine, tryptophan or tyrosine residues; incomplete disulphide bond formation or their rearrangement; and glycation of proteins[69-70]. As mechanisms of chemical degradation have been studied extensively and are well understood, established guidelines on how to prevent chemical degradation of proteins during bioprocessing and formulation stages exist[21, 71-72].

Physical degradation pathways include denaturation, self-association, aggregation, precipitation, liquid-liquid phase separation and surface adsorption[69]. They depend not only on protein properties but also on solvent and co-solutes that are present in protein formulation. Antibody formulations are typically administered subcutaneously in small volumes (1-2 mL) of highly concentrated protein solutions (>150 mg/mL)[73] as mAbs have limited permeability through cell membrane and degrade rapidly in the gastrointestinal tract[21, 74]. Under these high concentration conditions, protein self-association and aggregation are more likely to occur. These are detrimental for industry, not only due to possible immunogenicity, but also as monomeric protein species are the desired protein form. Protein self-association and aggregation are thus considered the major issues during biopharmaceutical production.

1.2.3 Pathways of protein self-association and aggregation

Proteins in solution are in dynamic equilibrium between native state and partially unfolded intermediates[75]. Formation of protein particles may occur through the association of native protein monomers, which is referred to as self-association[76]. Such oligomers are stabilised by weak electrostatic and hydrophobic interactions. They are typically reversible,

but may act as precursors to irreversible aggregates[77]. Alternatively, proteins may form well defined soluble oligomers through various mechanisms that do not result in the formation of larger aggregates. Processes that result in molecular assemblies larger than the monomeric native protein and contain unfolded, partially unfolded or misfolded protein chains will be referred to as aggregation, which is considered irreversible[76, 78]. This process is similar but should not be confused with amyloid aggregation during which partially unfolded proteins form well-defined fibrils with β-strand secondary structure[79].

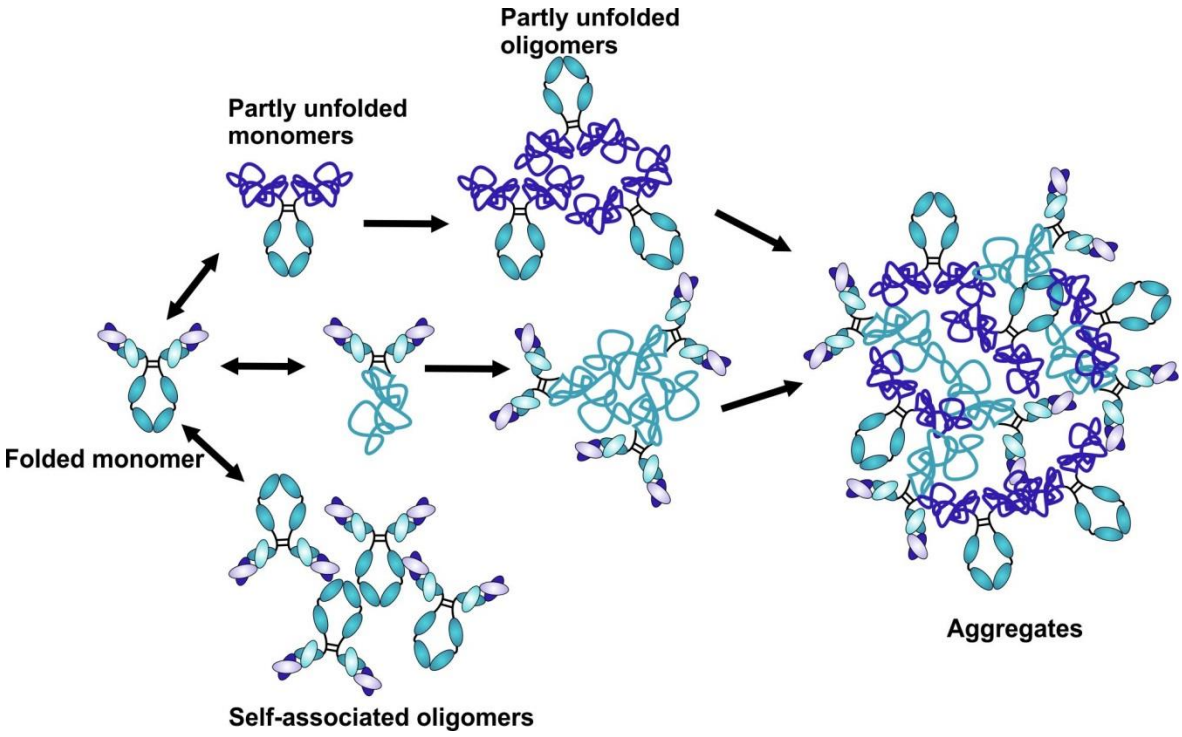

Figure 1.4: Scheme of protein aggregation pathways

Aggregation can occur through different pathways (Figure 1.4); however, a degree of conformational distortion is typically needed to expose the highly aggregation-prone regions and allow the formation of strong non-covalent interactions between polypeptide chains that stabilise the aggregates[78]. The potential pathways of protein aggregation and limiting steps in each of them are described in detail in several reviews[76, 78, 80-82]. Here we will consider only three mechanisms on which this thesis focuses. The first one is domain-swap oligomerisation, the second is reversible protein self-association, and the third one is aggregation through the addition of partially unfolded protein monomers.

1.2.4 Oligomerisation through domain-swap

Three dimensional (3D) domain swapping is a process in which one protein molecule exchanges domains in space with an identical partner, forming oligomeric species[83]. The exchanging domain connects to the rest of the protein through a so-called hinge region that typically adopts an extended conformation in the domain-swapped dimer and folds back on itself while in monomeric form. Except for the hinge region, the resulting domain-swapped oligomer has an identical structure to the monomer species (Figure 1.5).

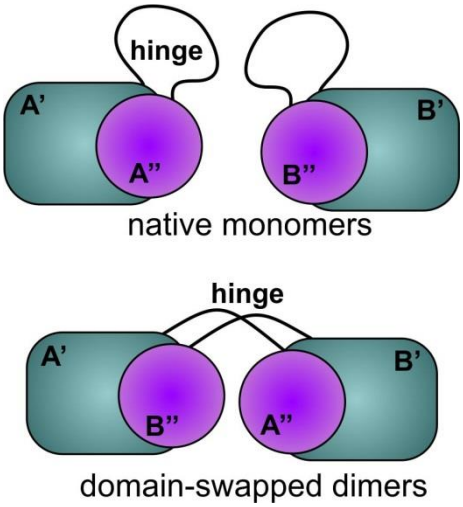

Figure 1.5: Schematic representation of domain-swap dimer formation

Domain-swap oligomerisation may be one of the mechanisms by which proteins have evolved the ability to oligomerise. Such oligomerisation may be involved in regulating the activity of proteins[84], however more commonly it results in the formation of inactive oligomers or fibrils that are often associated with the onset of diseases[85]. For example, mutation in human cystatin C, which causes spontaneous domain swap is related to the hereditary cystatin C amyloid angiopathy[86]. It has been shown, however, that domain-swap can be controlled by introducing mutations in the hinge region[87-89]. More recently it was established that domain swap could be induced in otherwise non-domain swapping proteins by introducing hinge regions from domain-swapping proteins[90-91]. This opened the possibility of building protein switches where domain swapping regulates protein activity[91].

1.2.5 Reversible self-association

Reversible self-association occurs when native protein molecules interact with each other through weak non-covalent interactions that are reversible, for example, upon dilution in buffer[75, 92]. The tendency to self-associate is intrinsic to the protein itself and is governed primarily by its primary sequence. The forces involved in reversible self-association are dipole-dipole, electrostatic, Van der Waals and hydrophobic interactions on protein surfaces, with electrostatic interactions being the predominant force in self-association process[93-94]. Additionally, non-specific long-range range repulsion and short-range attraction forces are crucial for cluster formation[74, 95-96]. Although these interactions are generally weak and reversible, they represent a challenge at high protein concentrations, where short-range attraction forces prevail, making the protein more likely to interact with neighbouring molecules[74]. Additionally, self-association can increase the viscosity of protein formulations, presenting processing and administration challenges[92, 97-98]. Protein self-association is also sensitive to solution conditions and can be altered by changing pH, temperature, ionic strength, and the presence of cosolutes. Strategies for preventing reversible self-association will be discussed further in section 1.3.2.

1.2.6 Aggregation through addition of partially unfolded protein monomers

Generally speaking, protein aggregation proceeds as a series of sequential events. Initially, partial unfolding or other structural perturbation of the native species occurs, exposing ‘sticky’ patches through which proteins interact with each other. This first results in the formation of smaller oligomers, followed by aggregate growth and ultimately the formation of subvisible and visible particles[81]. In solution, proteins undergo various structural perturbations and hence exist in equilibrium with various partially unfolded states that typically fold back into the native state. However, structural perturbations may expose patches that are aggregation-prone, and through which the initial non-native oligomers form. These aggregation-prone regions can be located throughout the protein structure. The initial nucleation rate of aggregate formation correlates primarily with the protein conformational stability and flexibility[99-101]. While electrostatic interactions are the main driving force of initial oligomerisation, other protein-protein interactions play a role as well, yet their contributions are not fully understood and are discussed at length elsewhere[81].

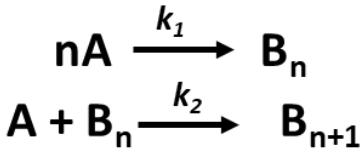

Figure 1.6: The F-W mechanism for protein aggregation

Initially formed oligomers can grow further into larger aggregates. Aggregate growth can proceed through various mechanisms that are often protein-specific and have been summarised in detail elsewhere[76, 78, 80]. Here, we will mention only the most straightforward model where aggregate growth occurs through a polymerisation reaction via addition of partially unfolded monomer species to the initially formed oligomers[102]. Exact protein aggregation pathways usually include multiple steps, which may co-occur or proceed too fast to be experimentally observed. Therefore, aggregation mechanisms are often simplified. The simplest option is the general two-step, minimal, 'Ockham's razor' model, also known as the Finke-Watzky (F-W) mechanism which was initially used to describe autocatalytic surface growth of metal nanoclusters[103]. It has been later applied to biological processes such as protein aggregation and formation of amyloids[104-106]. The F-W mechanism (Figure 1.6) describes protein aggregation as a two-step mechanism in which slow, continuous nucleation is followed by fast aggregate growth[106].

Conformational and colloidal stability of the protein control aggregation growth as well as nucleation. However, different molecular features of proteins do not necessarily correlate well with overall conformational and colloidal stability, let alone with aggregate growth. It is, therefore, challenging to predict aggregation rates from molecular descriptors such as exposed hydrophobic patches or charged amino acids alone[107].

1.2.7 Factors contributing to aggregation

Factors like protein amino acid sequence and structure, and environmental factors such as ionic strength, pH, temperature, surface interactions and exposure to mechanical stress[70, 77] influence protein aggregation. As aggregation may occur at any stage of bioprocessing, all factors that potentially influence aggregation must be considered and closely monitored to prevent a reduction in product quality[77, 108].

Protein sequence and structure

Amino acid sequence and structure of a protein are one of the defining factors for protein aggregation and are inherent to the protein itself. Content of hydrophobic amino acids is especially relevant as they may form aggregation-prone regions, and the introduction of new hydrophobic amino acids may substantially increase the rate of aggregation[109-110]. Amino acid composition can be altered through rational design and mutagenesis. However, introducing such mutations may lead to conformational instabilities or loss of activity, both of which are not acceptable[111]. Additionally, glycosylation patterns can alter the conformational and thermal stability of proteins and in turn, affect their aggregation propensity as well[112-113]. The secondary and tertiary structures of protein can also influence their aggregation propensity as they determine the spatial orientation of aggregation-prone regions. Additionally, increases in protein concentration can result in enhanced aggregation, which presents a significant challenge in the formulation of high-concentration biopharmaceuticals for subcutaneous administration[114].

External factors

Temperature affects protein aggregation by influencing protein-protein interactions, protein solubility, its conformational stability, and increasing the fraction of partially unfolded species. Increases in temperature typically result in accelerated aggregation, while decreases in temperature may result in increased reversible self-association that can potentially lead to opalescence, gelation, increased viscosity, and liquid-liquid phase separation[81]. Other solution conditions may also significantly affect protein physical stability. pH has the most significant influence on protein aggregation as it alters the surface charge of the proteins, thus influencing the nature and extent of protein-protein interactions, as well as the reactivity of partially unfolded species. Another critical attribute is the ionic strength of the solution that may alter the nature of protein-protein interactions due to screening or enhancement of electrostatic interactions. Additionally, ions can interact specifically or non-specifically with the protein, altering its aggregation propensity[115-117].

1.3 Strategies for mitigating protein self-association and aggregation

As protein self-association and aggregation are influenced by a wide range of factors, altering any of these factors can enhance or reduce the prevalence of these instabilities. In general, proteins can be stabilised by changing their structural characteristics through mutagenesis, by changing solvent composition, or by addition of excipients. In the next sections, we will discuss the three different approaches.

1.3.1 Protein engineering

It has been postulated that certain structural features, called aggregation-prone regions, drive protein aggregation. Multiple computational tools have been developed to identify them[118], and rational design and engineering strategies developed to modulate protein stability through protein engineering of these regions. These include site-directed mutagenesis, incorporation of post-translational modifications and use of fusion partners[119]. Thermal and colloidal stability of proteins can be maximised using rational design methods. One of the most straightforward stabilisation strategies is to replace free cysteines with other small amino acids, such as serine, in order to prevent the formation of inter-protein disulphide bonds and reduce the formation of irreversible aggregates[120]. Conversely, introduction of additional disulphide bonds can also significantly improve the thermal stability of proteins and prevent the opening of the structure and subsequent oligomer formation[121]. Substituting exposed nonpolar amino acids with polar residues or altering the protein net charge and its isoelectric point (pI) improves the solubility of proteins[122]. Rational design and engineering of biopharmaceutical proteins that are less prone to aggregation has had limited success as it is still unclear how specific mutations influence aggregation and it is still challenging to predict the impact of introduced mutations on proteins overall stability[76].

1.3.2 Solvent conditions

Physiochemical properties and chemical degradation of proteins depend on pH. Therefore, optimisation of buffer conditions is critical, especially in biopharmaceutical proteins stored for extended periods of time. pH of protein solutions is maintained using various buffering agents. The chosen buffering system should favour solubility and stability of the proteins and other molecules in the formulation, and should have an appropriate overall ionic strength. Proteins themselves have a buffering capacity, that increases with

increasing protein concentration, and which has to be taken into consideration when choosing the formulation buffer[123]. The most common buffers used in protein formulations are citrate, phosphate and acetate which are present in the majority of formulations approved by FDA[124]. In case of mAbs, histidine and phosphate are the most commonly used formulation buffers[125]. Some buffering agents change pH significantly upon temperature alterations; therefore, caution is needed when choosing storage conditions, or during temperature ramp experiments to study protein physical stability[126].

1.3.3 Excipients

The addition of various excipients can further improve the protein stability. By definition, excipients are inactive substances added to the protein formulation to enhance or maintain solubility and stability of the active ingredient[127]. In general, excipients should be soluble, non-toxic and non-immunogenic. Their stabilising effects depend on excipient concentration, type of the protein, and other solution factors such as pH and ionic strength. Therefore, systematic screening is needed to determine optimal formulations[128]. Typically, excipients are most effective at moderate (0.1M) to high concentrations (1M). The main classes of excipients include salts, sugars, polyols, amino acids, polymers and surfactants.

Salts

Salts are usually added to protein solutions in the form of buffering agents, as well as viscosity and tonicity modifiers. They can either have a positive, no or a negative effect on protein stability. Based on the effect salts have on protein stability and solubility ions and cations could be ordered into the so-called Hofmeister series, which was named after the first researcher that tried ordering the salts based on their effect on protein stability[129-130]. It is also known that cations effect the protein stability to a smaller extant then anions[131-132]. The magnitude of the effect depends on the protein itself and on the ionic strength of the salt added. Typically, Hofmeister effects are observed at medium to high salt concentrations (100-1000 mM)[117, 132]. Additionally, anions and cations can have opposite effects on protein stability; therefore, careful selection and optimisation is needed[128]. Apart from buffering agents, the most commonly used salts are potassium chloride (KCl) and sodium chloride (NaCl) that are highly effective in stabilising proteins and reducing viscosity through screening of the repulsive repulsion between solutes[133-134].

Sugars and sugar alcohols

Sugars and sugar alcohols (polyols) are one of the most widely used excipients in protein formulation. They are often added as tonicity agents, stabilisers, or to protect during freezing (cryoprotectants) or lyophilisation (lyoprotectants). The most commonly used sugars and sugar polyols are sucrose, trehalose, maltose, lactose, mannitol, sorbitol and glycerol. They are added to protein formulations to improve protein thermal and colloidal stability of the proteins. The stabilising effect of sugars and sugar polyols depends on their concentrations, with relatively high concentrations (>0.3M) often required to achieve the desired effect[128]. Their mechanism of action, extensively studied in the 1980s, is most likely through a preferential exclusion mechanism, where the sugar/polyol molecules are excluded from the protein surface due to their size. This exclusion is thermodynamically unfavourable, which in turn shifts the protein towards the smallest solvent-exposed surface and stabilises the native structure, preventing aggregation[135-138]. While sugars and polyols are effective against thermal stress, they can promote aggregate formation upon mechanical stress[139].

Surfactants

Surfactants are added to protein formulations to inhibit protein aggregation upon interfacial stress, prevent protein surface adsorption and protect against mechanical stress-induced aggregation. They are present in the majority of protein formulations, with polysorbate 20 and 80 (also known as Tween 20 and Tween 80), and poloxamer 407 being the most common ones. However, they are susceptible to degradation, forming reactive oxygen species that chemically modify proteins, causing their degradation and aggregation[140]. The common consensus is that in some cases surfactants bind to the protein surface, especially to hydrophobic patches, increasing protein colloidal stability, but that major stabilising effect arises from the competitive adsorption of surfactants at the interfaces, keeping the protein in solution[140].

Amino acids

Amino acids like histidine (His), methionine (Met), glycine (Gly), proline (Pro) and arginine hydrochloride (ArgHCl) are commonly used in protein formulation. His is primarily used as a buffering agent in mAbs formulation[141], Met as an antioxidant[142], Gly as a buffering and bulking agent[143], Pro as a stabiliser[144]. ArgHCl is used throughout biopharmaceutical production as a solubilising agent, additive in the mobile phase in analytical size exclusion and as an excipient reducing the viscosity of protein formulation and preventing protein aggregation[143]. Among amino acids, ArgHCl is the most studied due to its unique properties. Despite numerous studies, the mechanisms by which ArgHCl stabilises proteins is not fully

understood. Generally speaking, three possible mechanisms have been proposed: stabilisation by surface tension effects, direct interaction with the protein, or stabilisation by preferential exclusion. However, none of these has been conclusively confirmed[145-148]. Recently, it has been shown that arginine in combination with alternative anions, such as glutamate (ArgGlu) has beneficial effects on protein stability[149-151].

It is important to note that while effects of excipients on protein stability are extensively studied, these studies are often limited to a handful of excipients in combination with a single protein or consider only one aspect of protein stability. The effect of excipients is often protein-specific, which means that the results of these studies are often conflicting and do not take into account the complexity of these systems and inter-relations between different parameters. This ambiguity ultimately means that we still have limited understanding of how the stability of proteins depends on the solution conditions. Therefore, for practical applications, creating a stable protein formulation still requires time- and material-consuming combinatorial screening[152]. While high-throughput screening methods enable fast characterisation of formulations, there is a need to gain a more fundamental understanding of aggregation mechanisms and the impact of co-solutes, including excipients, on different aspects of protein stability.

1.4 Tools for characterising protein solutions

1.4.1 Traditional tools for assessing protein stability

Biopharmaceutics require thorough characterisation and quantification of chemical and physical instabilities as well as its degradation profile throughout bioprocessing, formulation, manufacturing and storage process to ensure structural and biological integrity of the product[153]. There is a large pool of methods available to characterise the thermal, colloidal and conformational stability of the proteins as well as study formation of oligomers and aggregates. Due to the limitations of each method, the most reliable results are obtained by comparing data from orthogonal techniques[77].

Apparent melting temperatures (T_m) are a measure of protein thermal stability, where an increase of T_m values reflects the stabilising effect of selected conditions on protein thermal stability. There are numerous methods available to obtain the T_m values, including differential scanning calorimetry (DSC)[154], differential scanning fluorimetry (DSF)[155] using

external fluorescent probes and its modified version that measures intrinsic fluorescence (nanoDSF)[156], measuring the maximum of intrinsic fluorescence[157], circular dichroism (CD)[158] and others.

The most straightforward measure of the protein colloidal stability is the onset of aggregation temperature (T_{agg}), which be determined by measuring static light scattering (SLS) or sample turbidity[159]. More accurate measures of the protein colloidal stability are diffusion interaction parameter (k_D)[160] and second osmotic virial coefficient (B_{22})[68], both determined using dynamic light scattering (DLS), and effective surface charge potential measured by electrophoretic light scattering[161].

Various biophysical methods are available to evaluate the amount and morphology of protein aggregates present in solution. As protein aggregates vary in size, complete physicochemical characterisation of protein and its aggregates cannot be achieved by a single analysis method[162]. The most commonly used technique is size exclusion chromatography (SEC) that separates particles based on their size. When coupled with a multi angle laser light scattering (MALS) detector, it also determines the size of soluble protein aggregates or their fragments[163]. Other commonly used techniques to study protein aggregation are dynamic light scattering (DLS) and static light scattering (SLS), which can determine the size of protein aggregates even in small amounts and in a high throughput manner[164]. Size of protein aggregates can also be assessed by analytical ultracentrifugation (AUC); however, the method is relatively low-throughput[165]. Infrared (IR), ultraviolet-visible (UV-Vis), CD and Raman spectroscopy indirectly detect changes in secondary, tertiary and quaternary structures that may occur due to aggregation[143]. Size and morphology of subvisible particles can be studied by small-angle x-ray scattering or by electron and atomic force microscopy. Methods mentioned so far offer an indication of size and topology of aggregates but fail to provide in-depth structural information needed for a mechanistic understanding of protein-excipient and protein-protein interactions.

While changes in protein's primary sequence are relatively easy to analyse, differences in 3D structure or conformation are much more difficult and time-consuming to characterise or are, in some cases, even indistinguishable by state-of-the-art analytical tools. Recently, various computational tools were developed to predict the stability of proteins and guide protein formulation. They are discussed in section 1.4.2. One of the experimental methods that can be applied to investigate structural changes upon protein-protein or protein-excipient

interactions in solution is nuclear magnetic resonance (NMR) spectroscopy and will be discussed in more detail in sections 1.4.3 – 1.4.6.

1.4.2 Computational tools for assessing protein stability

Computational approaches are applicable in every step of the design and development of a biotherapeutic product. *In-silico* methods available allow for mechanistic studies, including protein (mis)folding and protein-protein interactions, as well as predict aggregation propensity and identify possible aggregation-prone regions (APRs).

Some of the conventional computational approaches are homology modelling methods to guide protein design, docking, molecular dynamics (MD) or coarse-grain modelling to identify protein-protein interfaces and mutational studies to evaluate the effect of specific mutations on protein stability[166-168]. MD simulations are typically applied to smaller systems to explore forces and residues that drive the formation of oligomers[169], while coarse-grain models can be used to study bigger aggregates[170].

In protein formulation, computational approaches focus on using the molecular information (sequence or structure) to predict aggregation propensity and protein stability, which aid product development from design through to formulation stages[118, 171]. Standard view of protein aggregation is that partial unfolding results in the association of APRs, therefore a significant effort is going into accurately predicting these regions. Computational algorithms that predict protein solubility and aggregation propensity generally divide into those based on sequence and those based on structure. Sequence-based schemes include many factors but are usually based on the β-forming propensity of short linear segments in the amino acid sequence[172-175]. Software like Aggrescan3D[176], SAP[177-178]or CamSol[179] additionally take into account protein structure, eliminating the contribution of hydrophobic residues buried within the protein fold. They have been successfully applied in the screening of antibody libraries[180]. In contrast, Protein-sol predicts protein solubility by taking into account charge-based features as well as non-polar features of seven-amino acids long stretches in the sequence[181]. Additionally, tools for predicting hydrophobic and charged patches on the protein surface have been developed [182] and shown to correlate relatively well with experimental data[182-183]. Recently, several machine learning protocols were developed in an attempt to aid biopharmaceutical development[184-185]. These approaches, however, depend mainly on the quality of the dataset used to train the neural networks and results obtained can be difficult to interpret[186]. While computational tools cannot replace the

experimental methods, they can provide time and cost-efficient ways of guiding them which could significantly speed up the drug discovery and development[187].

1.4.3 Overview of biomolecular NMR to study protein-protein and protein-ligand interactions

NMR detects nuclei of isotopes with spin angular momentum in a magnetic field providing structural/physical information about the molecule under investigation. In the case of protein biopharmaceuticals, these nuclei are typically ^{1}H, ^{15}N, ^{13}C and ^{19}F. The most significant advantage of NMR over other analytical techniques is the detail and variety of obtained molecular information. In principle, each NMR active nucleus acts as a probe that reports on its chemical environment, local flexibility and apparent molecular size through several observable and measurable parameters, such as chemical shifts (δ), line width, signal intensity, and relaxation and diffusion rates. NMR can thus be used to probe slight structural changes occurring in proteins upon self-association or interactions with other molecules at the atomic level. NMR methods for studying interactions can be grouped into two main categories based on the observed molecule. Excipient-observed methods monitor signals of the binding molecule, while protein-based experiments detect resonances of the target protein.

However, certain limitations apply especially for protein-observed methods, as the increasing size of the observed system results in signal overlap. Therefore, proteins have to be sufficiently small or isotopically labelled using ^{15}N, ^{13}C or both in order to observe changes on per-residue bases in proteins. Additionally, line-widths in NMR spectrum depend on rotational correlation time, which is proportional to molecular mass, and thus large proteins result in broad NMR signals[188]. In practice this limited use of protein-observed solution NMR for proteins smaller than 40 kDa for a long time. However, the development of novel specific isotopic labelling methods[189-190] together with advances in transverse relaxation-optimized spectroscopy (TROSY) has enabled NMR based studies for proteins and protein complexes with molecular weights greater than 120 kDa[191-193]. With these techniques, some structural information can be obtained even for systems up to 900 kDa[194]. Conversely, ligand observed methods, utilised to study protein-ligand interactions, are not limited by the size of the proteins, and are thus valuable tool in fragment-based drug design (FBDD)[195], and more generally in studies of protein-ligand interactions[196].

The following sections provide an overview of the application of NMR to protein formulation and highlight the NMR methods used for characterisation of protein-excipient and protein-protein interactions in this Thesis.

1.4.4 Application of NMR to protein formulation

NMR has been successfully applied in detecting transient oligomer formation[197] of various small proteins and peptides, and has proved to be especially useful in determining the sparsely populated transient states on the amyloid fibril aggregation pathways[198]. More recently, an approach to measure protein aggregation through observation of transverse relaxation time (T_2) of water signal at low magnetic fields has been developed[199-201]. In therapeutically relevant proteins, NMR has been used to compare the high-order structure of biosimilar drugs with the reference product using 2D NMR. Initially, this approach was used on a smaller recombinant human granulocyte macrophage-colony stimulation factor[202] and filgrastim[203]. In parallel, ^{1}H profiling of antibodies in different formulations was developed as well[204]. Recently, a similar approach has been applied to map monoclonal antibody structure by 2D ^{13}C, and ^{15}N NMR spectra at natural abundance[205-206], which is particularly useful for biotherapeutic proteins where isotopic labelling is not feasible but achieving high protein concentration is usually not an issue. However, this typically requires enzymatic cleavage of mAbs to Fab and Fc fragments, and mAbs with high colloidal stability to ensure acceptable line shapes. Additionally, chemometrics analysis approach was developed to compare the higher-order structures of monoclonal antibody therapeutics[207-208].

NMR has also been applied to formulation design of pharmaceutically relevant monoclonal antibodies, using 1D ^{1}H methods, where they explored the effect of excipient on protein self-association and aggregation, revealing the favourable formulation conditions[149, 209-210]. Similarly, rotational and translational diffusion methods were used to assess the solution behaviour of biopharmaceuticals[211-212]. Furthermore, 1D T_2 filter experiments were applied to identify bioprocessing contaminants[213-214]. Additionally, ligand-observed NMR methods have been applied to study interactions of macrocycles[215] and surfactants with mAbs[216] while protein-salt interactions were monitored using protein-observed methods[117]. More recently, an approach using ^{19}F labelled biopharmaceutical proteins to monitor reversible protein self-association has been developed[217-219].

1.4.5 Protein-observed NMR methods

Protein-observed methods provide straightforward, reliable and unambiguous information about structural changes occurring in the receptor protein upon formation of the complex. In theory, protein-observed analysis is not limited by the size or by the affinity of the ligand. It can reveal different binding sites, enables direct distinction between specific and unspecific binding and provides a range of structural information. However, several limitations apply. Information at the residue-specific level can only be obtained if the spectral assignments for the target protein are known. Assignment of amino acids can be quite challenging, especially for larger proteins, and often requires uniform ^{13}C and ^{15}N isotopic labelling of the target molecule. Even though recent advances in multidimensional and relaxation-optimised techniques, such as TROSY[193, 220], combined with specific isotopic labelling of methyl groups enabled studies of proteins with molecular weights up to 120 kDa[190], the size of the target protein and formed complex still imposes the biggest limitation and a significant factor in determining what NMR experiments can be used. Additionally, specific isotopic labelling strategies are generally impractical for biopharmaceutical mAbs produced on an industrial scale in mammalian cell lines.

Chemical-shift perturbation (CSP)

Chemical shift perturbation (CSP) analysis is arguably the most frequently used NMR method to map protein interaction interfaces[221]. It follows changes in the chemical shifts of a protein upon interaction with a small-molecule ligand or another protein. This information can be utilised to determine the location of the binding site, calculate the affinity of ligand binding[222] or be used as constraints in protein-protein docking[223]. Alternatively, the affinity of the binding molecule can be determined even without assignment of the target protein, which is a useful approach in protein-based screening assays in drug design[224].

Interaction of the binding molecule with the target protein typically induces environmental changes on the protein surface which reflects in chemical shift perturbations of the nuclei on the binding interface as well as those around it. Perturbations in the chemical shifts can be observed in a variety of NMR spectra, however ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) and TROSY experiments are most commonly used as they probe every amino acid (via its amide signal) in the protein (except prolines) and are highly sensitive experiments with well-dispersed signals. It can be foreseen that the protein of

interest has to be uniformly ^{15}N labelled for these experiments to work on an appropriate timescale. More recently, a ^{1}H-^{13}C- HSQC based CSP method has been proposed for proteins containing ^{13}C-labelled methyl groups, which extended the use of CSP to proteins with molecular weight over 100 kDa[225].

CSP is usually performed by acquiring a series of NMR spectra of choice with increasing amounts of unlabelled binding partner added to the target protein. The interpretation of results is, in principle reasonably simple: peaks that move the most are more likely to map to the interaction surface[222]. However, caution is needed when interpreting data as additional changes in the spectrum may be observed due to long-range or allosteric effects of ligand binding[221]. Chemical shifts perturbation induced by 'real' interaction vs long-range effects of binding can be distinguished by comparing the chemical shift changes of a protein upon addition of another, closely related ligand[226].

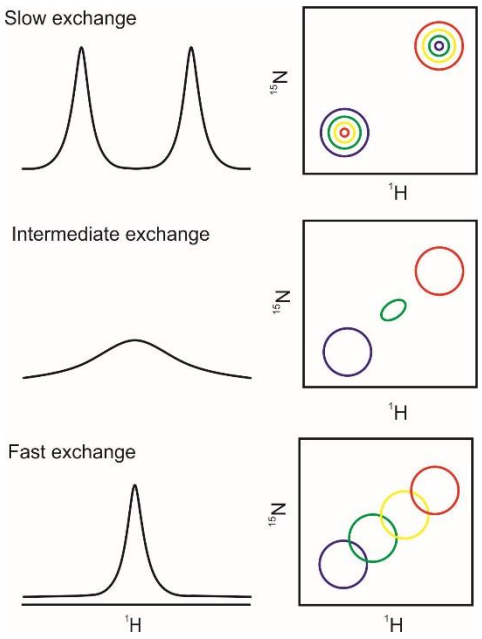

Figure 1.7: Dependence of chemical shift perturbation on the exchange rate. Left: Shape and number of ^{1}H signals for molecules in slow, intermediate and fast exchange regime Right: Number and shape of signals in 2D ^{1}H-^{15}N HSQC spectrum during gradual titration of the ligand as observed for slow, intermediate and fast exchange regime of bound and free state of the protein. Blue colour denotes free, green and yellow partially bound and red fully bound states of the protein.

For CSP analysis, it is essential to discuss the different chemical exchange regimes that are a result of the modulation in the isotropic chemical shift due to microsecond and

millisecond motions. At any given temperature, the observed molecule is macroscopically at equilibrium but each nucleus exchanges among different states which influence its magnetic properties and hence its chemical shift[227]. Chemical exchange phenomena can be grouped into three distinct regimes in terms of the 'NMR timescale': Firstly, if two distinct signals are observed for an individual nucleus, the exchange regime is referred to as slow. Secondly, if line broadening is observed upon addition of the ligand, the nucleus is in intermediate exchange. Thirdly, if a single average resonance is observed for a signal, it is in fast exchange[228]. Typical chemical shift perturbations patterns based on the exchange rate are depicted in Figure 1.7.

In the case of slow off-rate, typical for tight binding molecules, two sets of signals are observed for the target protein, one corresponding to the bound and second one for a free state. Intensities of the signals change during the titration; signals corresponding to the bound state increase at each addition of binding partner, while those associated with the free state slowly diminish (Figure 1.7 top). Even though position of signals, corresponding to bound state cannot be predicted unless separate assignment of bound complex is performed, the binding constant can still be estimated. However, other methods, such as SPR and ITC, may be more suitable for the determination of affinities for tightly bound complexes[229]. In the case of an intermediate exchange, affected resonances exhibit severe signal broadening, often causing signals to 'disappear' from the NMR spectrum (Figure 1.7 middle). In principle, resonances that are directly affected by interaction will disappear much faster than those that are further away from the binding interface. Using this approach, crude estimation of the interaction surface is still possible even though the estimation of binding parameters for complexes in intermediate exchange is challenging[221, 230]. However, a method enabling accurate evaluation of data in intermediate exchange based on the direct fitting of the 2D experiments has recently been introduced[231-232].

In the case of fast complex dissociation (fast exchange), resonances of bound and free protein are averaged, which means that a single resonance that moves linearly during titration is observed (Figure 1.7 bottom). Fast dissociation constants are observed for weaker interactions (μM - mM range). Shape of trajectories for interactions in fast exchange regime indicates the stoichiometry of the complex. Linear trajectories, occurring at the same rate for all peaks denote single binding site, while different rates and curved trajectories point to multiple binding events[221].

Relaxation experiments

Rather than being rigid proteins are flexible molecules. Furthermore, there is a direct link between protein dynamics and its molecular function. Specific recognition of ligand molecules and their binding are often accompanied by structural fluctuations that in turn affect protein dynamics. Additionally, dynamics of protein may influence rates and pathway of its misfolding or aggregation[233]. NMR can be used to study these dynamic processes at the residue-specific level in terms of heteronuclear (^{15}N-H, ^{13}C-H) relaxation rates and relaxation dispersion. Similar to chemical shifts, relaxation rates of each nucleus are affected by the chemical environment as well as by dynamic processes such as chemical exchange between two states.

Binding of a small molecule or another macromolecule to observed protein affects its structure and local motions, with the majority of changes occurring around the interaction interface. It can be foreseen that these changes are reflected in changes to the relaxation rates of nuclei affected by the interaction. Even though experiments that measure longitudinal (R_1) and transverse (R_2) relaxation rates of ^{13}C and ^{15}N were introduced in the 1950s, they have only relatively recently been applied to study protein dynamics[234]. Relaxation rates of ^{13}C and ^{15}N nuclei indirectly probe dynamic processes occurring on pico- to nano-second time scale[235]. Processes that happen in this timeframe include bond vibration, side-chain rotamer inversion, loop motions and backbone torsion angle rotation. As these structural changes may occur during binding of a ligand molecule to the observed protein, changes in relaxation rates can be used to probe conformational changes during binding or aggregation events, their effect on structural flexibility of the target protein as well as to map the binding interface.

Proton-deuterium (H-D) exchange

Hydrogen exchange rates of amide groups (NH) measured by NMR can be used to characterise protein structure, stability and dynamics at amino acid resolution[236]. In order for proton-deuterium exchange to occur, the NH group has to be directly exposed to the solvent. This process is significantly decreased if the NH forms a hydrogen bond with neighbouring groups. Therefore, HD exchange rates can be used in the mapping of secondary structural elements and protein folding intermediates[236] and evaluating protein dynamics[237]. However, it should be noted that the exchange rates are temperature and pH-dependent, making them difficult to measure in certain conditions[238-239].

In the most basic proton-deuterium exchange (H-D) NMR experiment, decay of signal intensity in HSQC spectra is monitored over time upon exposure of protein to fully deuterated solvent. In this experiment, the exchange that occurs within minutes to hours can be detected[240]. Faster exchange rates can be measured using Phase-Modulated CLEAN Chemical Exchange (CLEANEX-PM) experiment[241].

1.4.6 Ligand –observed methods

As mentioned in previous sections, NMR parameters of molecules depend on their size, shape and chemical environment (Table 1.1). Changes in line widths, NOE values and relaxation rates can, therefore, characterise the binding of a small molecule to a target protein. Ligand-observed methods are particularly useful if the target protein is too big to be directly studied by NMR, labelled samples are not available, or assignment is not possible[242], as is the typical case for mAbs.

Table 1.1: NMR properties of excipients in free and bound states

Free	Ligand properties	Bound
solvent	Chemical environment	protein
δ_{free}	Chemical shift	δ_{bound}
fast	Rotational tumbling	slow
slow	Transverse relaxation	fast
fast	Translational diffusion	slow
positive	NOE	negative

Most NMR experiments used to characterise protein-ligand interactions exploit the transfer of NMR properties from the protein to the small molecule and rely on the difference between these to distinguish between binding and non-binding molecules. For example, small molecules with short correlation times exhibit weak positive NOEs, while macromolecules that have long correlation times exhibit negative NOEs of much stronger intensity. Upon binding to a protein target, ligand adopts tumbling rate of the receptor and its correlation time significantly increases, which ultimately results in strong negative NOEs[243-244]. Direct observation of proton resonances of ligand bound to protein target is often difficult or even impossible due to broadening of signals[243]. In the case of slow

exchange between free and bound states of excipient, only signals of free excipient are observed while proton resonances observed in the case of fast exchange regime are an average of signals from both, free and bound states. In case of fast exchange regime, chemical exchange occurs many times before the NMR signal relaxes, which leads to transfer of negative NOE build-up regime of bound excipient to easily detectable proton resonances of free excipient[245]. In practice, this means that signals corresponding to ligands that bind to protein target exhibit negative signals, while unbound ligands have positive signals. The same principle applies to all of the observables measured by experiments described below.

T_2-filter

In NMR, relaxation describes the process in which the magnetisation restores to equilibrium after the radiofrequency pulse (RF) has been applied. The two main relaxation processes are the spin-lattice (longitudinal, T_1) and spin-spin (transverse, T_2) relaxation. While longitudinal relaxation describes the restoration of the magnetisation in the direction of the external magnetic field, the transverse relaxation depends on the fluctuations of the local magnetic field which leads to the dispersion of the phase coherence in the x-y plane[246]. The T_2 relaxation of the small molecule significantly increases when it interacts with the protein, which can be measured by a standard Carr-Purcell-Meiboom-Gill (CPMG) pulse-echo sequence[247]. CPMG uses a train of successive 180° pulses with variable delay between them during which the magnetisation relaxes. The experiment is typically acquired as a pseudo 2D experiment, where the number of refocusing elements is varied. This allows for accurate determination of transverse relaxation rates that are sensitive to protein-excipient interactions. However, such an experiment has fairly long acquisition times. Therefore, various alternative approaches have been developed in which CPMG spectra are acquired at only two different lengths of the echo train, and the ratio between signal intensities is used to evaluate binding to the protein[248-249].

Differential line broadening

Linewidth of signals in NMR spectra depend primarily on T_2 relaxation. Longer relaxation times result in narrow line widths while faster relaxation (short T_2) result in broader signals. Upon binding of a small molecule to a large protein receptor, its transverse relaxation rate is significantly accelerated, which reflects in broadening of the ligand signals in ^{1}H spectra[224]. Broadening of ligand resonances upon addition of potential protein target is

more significant for resonances that are in direct contact or in closer proximity to target protein therefore a clear indication of binding. By comparing the relative line broadening upon binding to the receptor protein, the receptor-binding site of the ligand can be mapped[194]. The same principle can be applied to identify the binding site of the receptor protein, provided we have uniformly ^{15}N labelled sample with known assignment. In case of large protein complexes, mapping of interaction site can be achieved by methyl TROSY NMR spectroscopy of ^{13}C methyl labelled protein interacting with non-labelled binding partner[250].

Saturation transfer difference (STD)

Saturation transfer difference (STD) NMR spectroscopy allows qualitative and quantitative analysis of small molecules binding to macromolecular molecules[251]. Initially, the STD experiment was designed to screen libraries of small molecules for compounds with binding activity to the target protein[251]. It has since been successfully implemented in the characterisation of ligand-receptor interactions and analysing binding kinetics of ligand-receptor complexes[242].

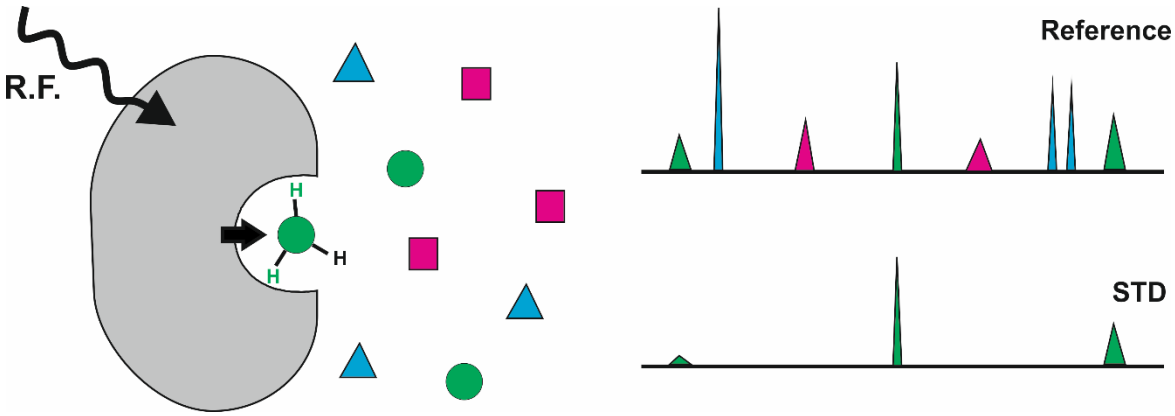

Figure 1.8: Schematic representation of the STD experiment

STD experiments rely on intermolecular magnetisation transfer from protein target to its binding partner. While in contact with the receptor, the ligand is subject to the same NMR properties as the protein. Therefore, saturation applied spreads not only to all protein protons, but also to the protein-bound ligand via dipolar interactions. Saturation of ligand results in decreased intensity of its signals, compared to the spectrum of free ligand, which

is exploited in STD NMR spectroscopy. In STD pulse sequence, protein is first saturated with selective irradiation in a spectral range that does not contain ligand resonances. During selective saturation time, magnetisation spreads from protein to bound ligand. Spectrum acquired in this way is often referred to as '*on-resonance'*. A reference spectrum, also referred to as *'off-resonance'*, is obtained using selective saturation applied far off-resonance of the receptor (usually 30 ppm). Finally, the STD spectrum is obtained by subtracting the on-resonance spectrum from the off-resonance one[251]. In this difference spectrum only signals from bound ligands are present (Figure 1.8). Ligand protons in closer proximity to the protein target receive higher degrees of saturation, which reflects in more substantial STD effect. This phenomenon is used to map the binding epitope of the ligand at atomic resolution[252].

The STD effect increases with longer saturation times and signal to noise ratio improves with higher ligand to protein ratio[252]. The effect also depends on the kinetic parameters of complex formation and can be thus used to determine the K_D value of binding ligands. As absolute STD effects cannot be directly compared, STD amplification factor (AF^{STD}) is calculated using the equation 1.1 where AF^{STD} is the STD amplification factor, I_{off} the intensity of excipient signal in the off-resonance spectrum, I_{on} the intensity of signal in the on resonance spectrum, and *[L]* and *[P]* represent total concentrations of the ligand and the protein, respectively.

$$AF^{STD} = \frac{I_{off} - I_{on}}{I_{off}} * \frac{[L]}{[P]} \qquad \text{(Eq 1.1)}$$

The build-up curve of AF^{STD} versus ligand concentration can be fitted using a simple one-site binding model, assuming specific binding of the ligand. However, this approach does not result in accurate binding affinity constants[253-254]. More accurate K_d values can be obtained by performing competition binding studies[255] or by determining STD initial growth rates[256]. The main advantages of the STD experiment are the small amounts of unlabelled protein needed for experiments, its excellent sensitivity and the relatively short experimental time compared to other NMR experiments. It can be applied to any protein with rapid spin-diffusion (typically proteins >10 kDa), and does not impose any upper limit to the molecular weight of the protein. Furthermore, STD experiments are not limited to soluble proteins, but can also be applied to immobilized proteins[257] or transmembrane proteins embedded in

liposomes[258]. It is, however, limited to systems where free and bound states of the ligand are in fast exchange with respect to the T_1 relaxation time of the protein target. In practice this translates into K_D values limited between µM and mM range[242]. For complexes with tighter binding, STD effect cannot be observed as the exchange between bound and free states of the ligand is too slow to reflect in attenuated signal intensities in on resonance spectrum, while magnetisation transfer is not sufficient enough to detect binding for protein-ligand systems with weaker binding affinity[259].

Water-Ligand Observed via Gradient Spectroscopy (WaterLOGSY)

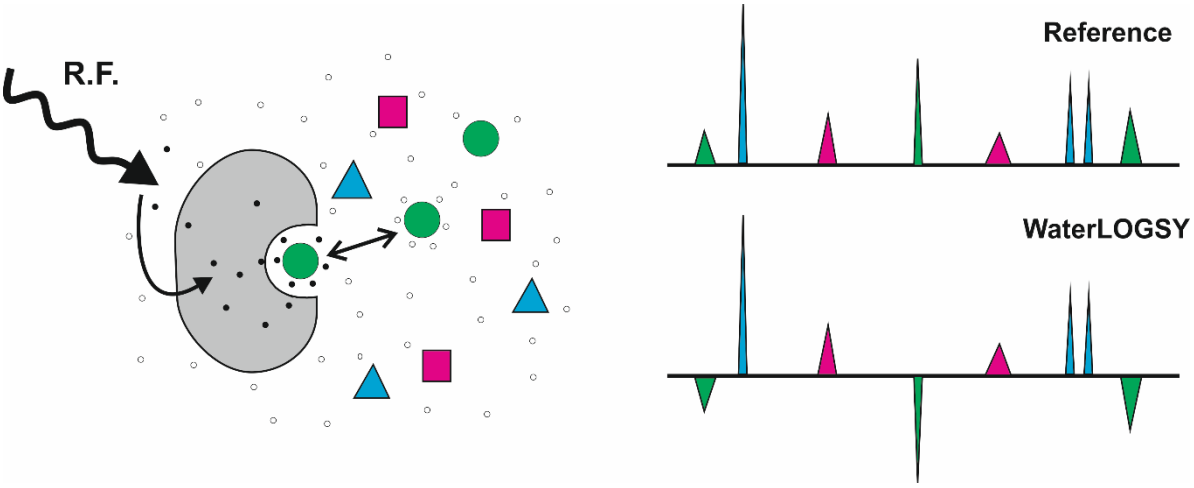

Figure 1.9: Schematic representation of the WaterLOGSY experiment.

WaterLOGSY is an experiment derived from STD NMR. Both methods detect changes in the spectrum of free ligand upon transfer of magnetisation to the bound ligand. In WaterLOGSY, magnetisation is indirectly transferred by excitation of the bulk water magnetisation[260]. The excited water interacts with protein-excipient complex and transfers magnetisation through cross-relaxation and proton exchange, which is retained by the ligand after it dissociates. Non-binding and binding compounds in the mixture are easily distinguished from each other as they either have decreased intensity or the opposite sign of signals in WaterLOGSY spectrum (Figure 1.9)[261]. The reference spectrum of excipients mixture in the absence of protein is needed to eliminate possible false-positive signals

corresponding to exchangeable protons of excipients. Due to excess of solvent and large number of exchangeable protons in protein-excipient complex, WaterLOGSY is much more sensitive than other ligand-based screening methods. Therefore, WaterLOGSY is often used for screening mixtures of compounds against the target of interest in drug design[261-262].

1.5 Introduction to the Thesis

1.5.1 PIPPI consortium

This PhD Thesis was part of the Innovative Training Network (ITN) named Protein-excipient Interactions and Protein-Protein Interactions in formulation (PIPPI) funded by the European Horizon 2020 programme. The consortium consisted of both academic and industrial partners across Europe, namely Danish Technical University, Copenhagen University, Ludwig-Maximilians University Munich, Lund University, University of Manchester, Novozymes, AstraZeneca, Wyatt Technologies Europe, Roche, NanoTemper Technologies and MAX IV. More information about the consortium and contributing partners is available on the PIPPI website http://www.pippi.kemi.dtu.dk/contact.

The overall aim of the PIPPI consortium was to systematically investigate the biophysical behaviour of biotherapeutic proteins to create a comprehensive database (PIPPI-data) and a platform for successful protein formulation. Fifteen Early Stage Researchers (PhD students) were employed within the consortium, all using the same proteins supplied by the industrial partners but each of us focusing on a different aspect of protein formulation and different methods. My work focused on applying NMR to biopharmaceutical formulations using both unlabelled proteins provided by consortium members, and an isotopically labelled protein scaffold SQT expressed in-house.

Within the consortium, special attention was given to the collaborative aspect of research that was organised in two levels. Firstly, each student had to invest part of their time in performing screening experiments for the database (PIPPI-data). In my case, the task was to apply NMR approaches to evaluate protein stability and protein-excipient interactions in biopharmaceutical formulations. Secondly, each student had to visit other partners within the consortium for the cumulative time of at least sixteen weeks and host other PIPPI fellows at their institution. I had the opportunity to spend ten weeks in the lab of Dr Gunther H.J. Peters at DTU, where we have applied various molecular dynamics approaches to complement our experimental results, and twelve weeks in the lab of Prof. Dr Gerhard Winter at LMU where I enhanced my knowledge of protein analytics. During my PhD, we have also hosted four PIPPI fellows from other universities in our lab, whom I assisted with planning, acquiring, processing and interpretation of NMR data that complemented their research.

These collaborations resulted in several publications, listed below, which are not included in this Thesis as results chapters. While some of these have already been published, others are still under review or in preparation at the moment. My contributions to each of these papers are outlined in section 7.1.

- M. Martinez Morales, M. Zalar, S Sonzini, A. P. Golovanov, C. F. van der Walle, and J.P. Derrick, **Interaction of a Macrocycle with an Aggregation-Prone Region of a Monoclonal Antibody,** Molecular Pharmaceutics 2019 16 (7), 3100-3108.

- S. Indrakumar, M. Zalar, C. Pohl, A. Nørgaard, W. Streicher, P. Harris, A. P. Golovanov, and G.H.J. Peters, **Conformational Stability Study of a Therapeutic Peptide Plectasin Using Molecular Dynamics Simulations in Combination with NMR,** The Journal of Physical Chemistry B, 2019, 123(23), 4867-4877.

- H.L. Svilenov, A.V. Kulakova, M. Zalar, A. P. Golovanov, P. Harris and G. Winter, **Orthogonal Techniques to Study the Effect of pH, Sucrose and Arginine Salts on Monoclonal Antibody Physical Stability and Aggregation During Long-term Storage**, Journal of Pharmaceutical Sciences, 2020 109(1), 584-94.

- S. Indrakumar, M. Zalar., N. Tschammer, A. Nørgaard, W. Streicher, P. Harris, A.P. Golovanov, G.H.J. Peters, **Synergistic applications of molecular dynamics, microscale thermophoresis and NMR to probe excipient interactions with therapeutic peptide Plectasin**, 2020, *in review*.

- C. Pohl, M. Zalar, I. El Bialy, S. Indrakumar, G.H.J. Peters, A. P. Golovanov, W. Streicher, A. Nørgaard and P. Harris, **The effect of point mutations on the protein properties in solution: development of a screening protocol of an anti-microbial peptide,** 2020, *Molecular Pharmaceutics, accepted.*

- A.V. Kulakova, L. Gentiluomo, H.L. Svilenov, D. Augustijn, I. El Bialy, ML. Greco, S. Indrakumar, S Mahapatra, M. Martinez Morales, C. Pohl, A. Roche, M. Zalar A Tosstorff, R. Curtis, J. P. Derrick, A.Nørgaard, Tarik A. Khan, A.P. Golovanov, G.H.J. Peters, A. Pluen, Å. Rinnan, W. Streicher, C. F. van der Walle, S. Uddin, G Winter, D Roessner, Wolfgang Frieß and P Harris, **Advancing excipient selection for protein formulation through comprehensive biophysical characterisation**, 2020, *in preparation.*

1.5.2 Aims and structure of the Thesis

The main aim of this thesis was to characterise protein-protein and protein-excipient interactions in biopharmaceutical formulations using NMR spectroscopy and evaluate how these affect various aspects of protein stability using a range of complementary techniques.

In Chapter 2, we used an SQT variant with AU1 and c-Myc insertion peptides (SQT-1C) to study the effect of peptide insertions on scaffold protein structure and oligomer formation. The possible oligomer topologies were determined using NMR restrained protein-protein docking, and two possible mechanisms of oligomerisation were compared using molecular dynamics simulation. In Chapter 3, we studied in detail the oligomerisation kinetics of SQT-1C, determined the oligomerisation pathway and explored the factors contributing to tetramer formation. Furthermore, a new, more stable variant of SQT was developed using rational design. In Chapter 4, we employed NMR screening to characterise protein-excipient interactions, using PIPPI supplied proteins and commonly used excipient. We have also evaluated the effect of excipients on different aspects of protein stability and aggregation kinetics and explored the correlations between them. In Chapter 5, we employed ligand- and protein-observed screening of interactions between stabilised SQT-1C and commonly used excipients to evaluate how they can be applied to evaluation of protein-excipient interactions and compare the two approaches.

Each results chapter is in the form of a manuscript, including separate Introduction, Material and Methods, Results and Discussion sections, with additional Supplementary Information. Chapter 6 is a general discussion and conclusion where the overall results of this Thesis and future outlook are discussed with regard to the initial aims of the project.

BLANK PAGE

2 Studies of the oligomerisation mechanism of a cystatin-based engineered protein scaffold

This chapter was written, peer-reviewed and published as:

Matja Zalar[1], Sowmya Indrakumar[2], Colin W. Levy[1], Richard B. Tunnicliffe[1], Günther H.J. Peters[2], Alexander P. Golovanov[1,*], Studies of the oligomerisation mechanism of a cystatin-based engineered protein scaffold, Scientific Reports, doi: https://doi.org/10.1038/s41598-019-45565-6

[1] Manchester Institute of Biotechnology and Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester, M1 7DN, UK

[2] Department of Chemistry, Technical University of Denmark, Lyngby, Denmark

* Correspondence to: A.Golovanov@manchester.ac.uk

Author contributions:

M.Z. prepared the protein samples, performed the experiments, evaluated the data and wrote the first draft of the manuscript. C.W.L. assisted at setting up the crystallisation trials, acquired and processed the x-ray data, finalised the crystal structure refinement and deposited it to the PBD database. R.B.T. made valuable contributions to increase the extent of the NMR assignments. S.I. assisted at setting up molecular dynamics simulations and provided guidance at data analysis. M.Z., A.P.G and G.H.J.P. conceived the presented work and planned the experiments. A.P.G and G.H.J.P. provided conceptual guidance. All authors contributed to paper writing.

2.1 Abstract

Engineered protein scaffolds are an alternative to monoclonal antibodies in research and drug design due to their small size, ease of production, versatility, and specificity for chosen targets. One key consideration when engineering such proteins is retaining the original scaffold structure and stability upon insertion of target-binding loops. SQT is a stefin A derived scaffold protein that was used as a model to study possible problems associated with solution behaviour of such aptamers. We used an SQT variant with AU1 and c-Myc insertion peptides (SQT-1C) to study the effect of peptide insertions on protein structure and oligomerisation. The X-ray structure of monomeric SQT-1C revealed a cystatin-like fold. Furthermore, we show that SQT-1C readily forms dimers and tetramers in solution. NMR revealed that these oligomers are symmetrical, with inserted loops comprising the interaction interface. Two possible mechanisms of oligomerisation are compared using MD simulations, with domain swap oligomerisation being thermodynamically favoured. We show that retained secondary structure upon peptide insertion is not indicative of unaltered 3D structure and solution behaviour. Therefore, additional methods should be employed to comprehensively assess the consequences of peptide insertions in all aptamers, particularly as uncharacterised oligomerisation may alter binding epitope presentation and affect functional efficiency.

2.2 Introduction

Monoclonal antibodies (mAbs) represent the major class of molecules used for affinity binding studies in research, and are the most commonly used diagnostic and biotherapeutic agents[20, 263]. Despite their versatility and wide applicability, production of mAbs is costly and poses many challenges due to their high molecular weight (>140 kDa) and structural complexity, including post-translational modifications typically requiring mammalian expression systems[264]. In light of these inherent problems of mAbs, alternative approaches to find high-affinity binders for a specific target have been developed, including using smaller antibody fragments[265], or using engineered protein scaffolds (also called protein aptamers) based on non-immunoglobulin proteins[33, 266]. Protein aptamers are designed by insertion of a short (typically up to 10 – 15 residues) peptide containing the desired binding epitope into a loop of a stable protein scaffold[37]. In principle these aptamers mimic the antibody-based molecular recognition mechanism, but have a much smaller frame, simpler design, do not have post-translational modifications, and are often obtained using less

demanding recombinant expression systems[37, 267]. To date, more than 50 structurally diverse non-immunoglobulin protein scaffolds have been developed[268-269]. While they were initially used for construction and screening of combinatorial protein libraries for protein recognition[37, 270], they have since then become widely used in studies of protein function and molecular interaction[271-273], as diagnostic tools[274-275] and biosensors[43], as well as imaging agents[40, 276] and biotherapeutics[277-279].

The correct presentation of peptide epitopes for binding, and hence the activity of the scaffold-based binding proteins, strongly depends on maintaining scaffold structure and solution properties. Often, point mutations are introduced to improve structural robustness and thermodynamic stability of scaffolds[47], while tags are added to improve their solubility[48]. Whereas rigorous protein library evaluation is usually performed to check scaffold's affinity and specificity towards the target protein, their 3D structural characteristics and solution behaviour are not always characterised comprehensively[41, 272, 280-281]. During scaffold development, the effect of loop insertions on thermal stability and retention of secondary structure is typically tested; however, large variations are acceptable[51, 56, 279, 282]. Although it has been shown previously that insertion of specific peptides needed for function can significantly perturb scaffold structure[36, 49], the origins and consequences of such instabilities have not been studied in detail. A better understanding of these problems is required, perhaps by looking at specific case studies first.

SQT is one of many engineered scaffold proteins: it is based on human stefin A (SteA, also known as cystatin A) that has three possible insertion sites for peptides; namely the N-terminus, loop 1 (L1) and loop 2 (L2). Although it has been previously shown to retain secondary structure upon insertion of a variety of peptide combinations into its insertion sites[56], no tertiary or higher-order structural characterisation has been reported.

Here we have used an SQT variant, named SQT-1C, with AU1 and Myc peptides inserted into loops L1 and L2, respectively, as a model to understand the effect of inserted peptides on scaffold structure, stability and oligomerisation properties. X-ray crystallography confirmed that monomeric SQT-1C exhibits typical cystatin fold but only in very specific conditions, when crystallised in the presence of 19% dioxane. However, in solution, in the absence of dioxane, monomeric SQT-1C exists in dynamic equilibrium with domain-swapped dimeric and dominant tetrameric species. NMR was used to determine the amino acid residues involved in oligomerisation of the SQT variant, while molecular dynamics (MD) simulations and molecular mechanics energies combined with the

generalized Born surface area continuum solvation (MM-GBSA) calculations were employed to explore the driving forces of SQT-1C self-association. We show that inserted peptide regions play a crucial role in scaffold protein oligomerisation via domain swapping, leading to significantly different conformations and surface exposure of these binding epitopes in monomeric, dimeric and tetrameric forms. We anticipate that due to general similarity of protein folding driving forces, the instabilities revealed in this case study may also be observed in other small protein aptamers, therefore justifying a thorough characterisation of the "functional" conformations using an appropriate range of techniques, helping to troubleshoot designs early in their development processes.

2.3 Results

2.3.1 Purification of SQT reveals defined oligomeric species

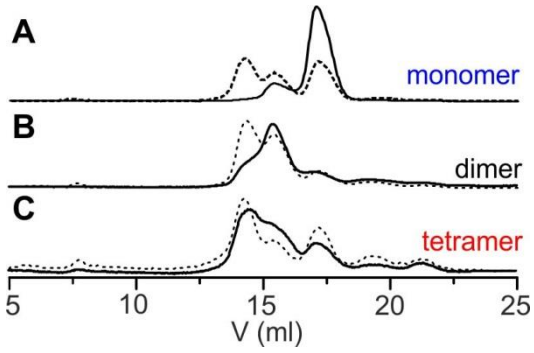

Figure 2.1: Analysis of SQT-1C oligomeric states. SEC traces of A) monomeric, B) dimeric and C) tetrameric SQT-1C reinjected onto the column immediately after purification (solid line) and after 24 h of incubation at 25 °C (dashed line). Protein concentration in monomeric, dimeric and tetrameric fractions were, 0.8, 1 and 0.5 mg/ml, respectively. The peak at 7.5 ml corresponds to a larger aggregate which is not in equilibrium with other species.

The variant of a model scaffold protein SQT[56] with AU1 and Myc peptide insertions and C-terminal hexa-histidine tag, named SQT-1C, was expressed and purified. After separation of refolded SQT-1C on size exclusion column, three distinct elution peaks were identified, corresponding to SQT-1C monomer, dimer and tetramer, as confirmed by size-exclusion chromatography (SEC) coupled with multi-angle light scattering (MALS) (Figure S2.1). In addition, a low intensity polydispersed peak corresponding to higher oligomeric

species was observed; however, its concentration remained marginal even after extended periods of time. The isolated monomeric fraction oligomerised over time, again producing the same three well defined peaks on the chromatogram, with the tetramer species being predominant. Similarly, isolated dimeric and tetrameric species equilibrated into a mixture of all three species within 24 hours (Figure 2.1).

Melting temperature of SQT-1C was 54°C±1°C, as measured by changes in intrinsic fluorescence (Figure S2.1), compared to 79.9°C reported for SQT alone[56], which indicated that the inserted peptides significantly destabilised the scaffold. To understand the underlying mechanism and pathway of SQT-1C oligomerisation we attempted to crystallise fresh monomeric, dimeric and tetrameric SQT-1C fractions.

2.3.2 Crystallised SQT-1C exhibits a cystatin-like monomer fold

SQT-1C crystallised only from the freshly-purified tetrameric fraction, and, in the presence of 19 % dioxane, surprisingly yielded monomeric structure, which was solved to 2.5 Å resolution. Essential data collection and refinement statistics is shown in Table 2.1, and extended statistics in Table S2.1. The structure coordinates were submitted to the Protein Data Bank under accession code 6QB2. Interpretable electron density was observed for polypeptide regions R4-L48 and A93-N113 whilst residues D49-F75 were visible but less well ordered, indicating higher flexibility of this region. The SQT-1C monomer exhibits a cystatin-like architecture with five anti-parallel β-sheets and a perpendicular α-helix (Figure 2.2), showing that peptide insertions into loop regions do not affect the SQT scaffold backbone. For further analysis using MD simulations, the missing polypeptide stretches were added to the X-ray structure with homology modelling (see section 2.5.7 for more details). PDBePISA analysis[283] of protein interfaces did not identify any specific interactions that would result in the formation of stable oligomers indicating that SQT-1C crystallises as a monomer. SQT-1C oligomerisation state in solution under crystallisation conditions was further assessed by SEC, which revealed that in the presence of 19% dioxane SQT-1C is monomeric (Figure S2.2A). NMR spectra of SQT-1C in the presence of dioxane were also consistent with the monomeric form of the protein (Figure S2.2B) additionally confirming that presence of high percentage of dioxane shifts the SQT-1C equilibrium exclusively towards the monomeric state, possibly by strengthening the salt-bridges within the protein[284].

Table 2.1: Crystallography data collection and refinement statistics for SQT-1C

	SQT-1C*
Data Collection	
Space group	P 42_12
Cell dimensions	
a,b,c (Å)	96.12, 96.12, 29.86
α,β,γ (°)	90, 90, 90
Resolution range (Å)	42.99 - 2.5 (2.589 - 2.5)
R_{merge}	0.05822 (0.9239)
$I/\sigma I$	17.39 (2.37)
Completeness (%)	99.79 (99.60)
Redundancy	11.9 (12.6)
Refinement	
Resolution (Å)	2.5
No. reflections	5207 (497)
R_{work}	0.2725 (0.4411)
R_{free}	0.2892 (0.6248)
No. atoms	
Protein	753
Protein residues	95
B-factors ($Å^2$)	
Average	104.03
Protein	104.03
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.63
Number of TLS groups	1

**Statistics for the highest resolution shell are shown in parenthesis*

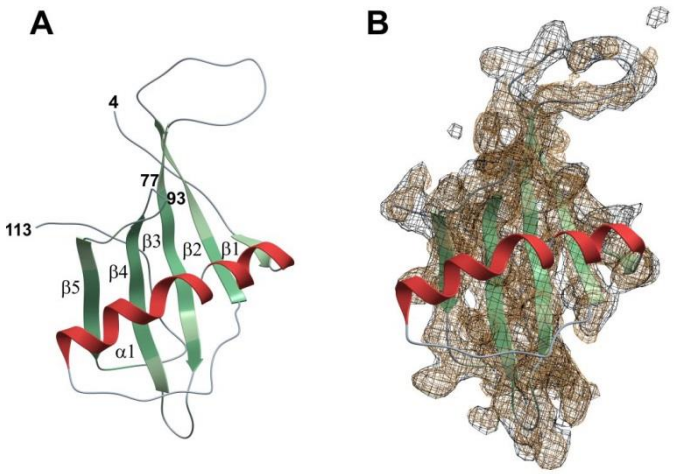

Figure 2.2: Overview of SQT-1C structure. A) Cartoon representation of SQT-1C structure coloured by secondary structure element with termini and chain breaks labelled with residue numbers. B) Electron density and cartoon representation of SQT-1C structure. Black electron density is 2Fo-Fc map contoured at 1 sigma level, and brown electron density composite omit iterative rebuild map contoured at 1 sigma level. This map shows good density for the core of the protein but only weak density for the loops. For clarity, the electron density of the helix region has been omitted.

2.3.3 SQT-1C forms symmetric oligomers in solution

SQT-1C monomer in a normal buffer solution in the absence of dioxane freely oligomerises into dimers and tetramers over time. Far–UV CD spectra of freshly-separated SQT-1C monomers showed little difference when compared to dimers and tetramers, indicating that the overall fold of SQT-1C is retained upon oligomerisation (Figure S2.1C). Structural changes occurring upon SQT-1C oligomerisation were further assessed using NMR spectroscopy. The direct backbone assignment of SQT-1C was not possible due to signal loss and broadening, likely caused by chemical exchange between oligomeric species in the intermediate regime. Therefore, another construct, SQT-1N, was used for the backbone assignment following a standard sequential assignment strategy (submitted to BioMagResBank, ID 27757). The N-terminal tag and residues M20-V23, G55, D68-Y70, Y72, I73, G98, K102 and T128, could not be unambiguously assigned due to lack of cross-peak connectivities in 3D spectra. The sequence-specific assignment of SQT-1C was derived from that of SQT-1N, as described in section 2.5.6. Backbone ^{1}H and ^{15}N chemical shift assignments of both constructs are shown in Figure S2.3. While SQT-1N exhibited better NMR spectral properties, it also formed dimers and tetramers, similarly to SQT-1C (Figure S2.4).

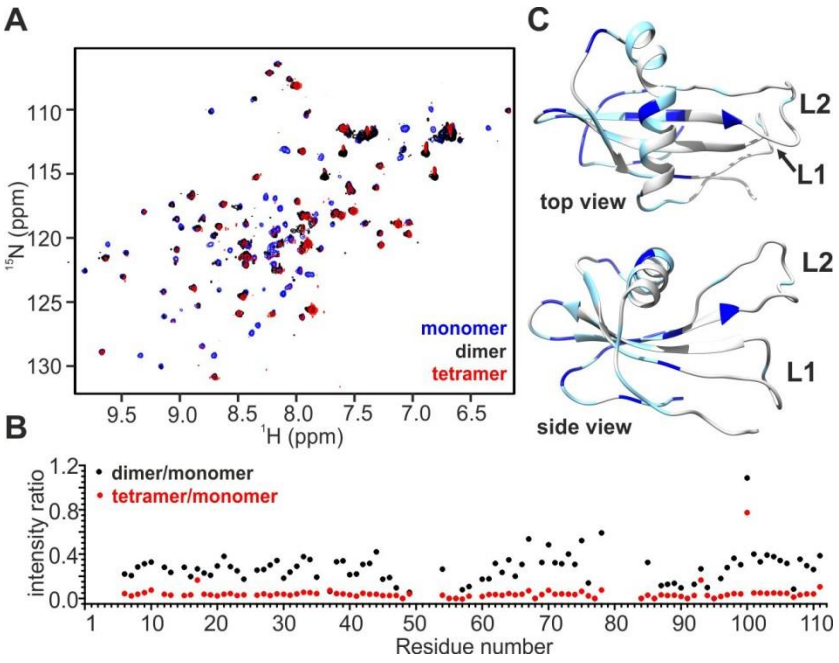

Figure 2.3: NMR characterisation of SQT-1C oligomeric species. A) Comparison of ^{1}H-^{15}N HSQC spectra of freshly isolated monomeric (blue), dimeric (black) and tetrameric (red) SQT-1C indicates symmetrical topology of oligomers and conformational exchange in intermediate regime. B) Dimer-monomer (black) and tetramer-monomer (red) peak intensity ratio plotted against SQT-1C residue number. C) Regions with preserved signal intensities were identified and two arbitrary thresholds were set to 0.48 and 0.25, and mapped onto the SQT-1C structure Residues between threshold values are plotted in light blue while residues with intensity ratio above 0.48 are coloured with dark blue. They indicate regions of SQT that are less affected by oligomerisation.

The 1D ^{1}H NMR spectrum of the freshly-purified SQT-1C dimer exhibits line broadening without significant signal shifts when compared to the spectrum of monomer, which is expected due to increased molecular weight and possibly due to chemical exchange. These effects are even more prominent in the spectrum of the tetrameric fraction (Figure S2.5). Similar to the ^{1}H spectra, line broadening was observed in 2D ^{1}H-^{15}N HSQC spectra of dimers and tetramers, but it was not uniform across all peaks (Figure 2.3). This non-uniformity was indicative of the spectral resolution being influenced not only by an increase of molecular weight and local polypeptide mobility but also by conformational exchange processes in the intermediate regime. The intensity ratio of peaks measured in different oligomeric states of SQT-1C (Figure 2.3B) revealed that residues least affected by line

broadening are mostly situated on the part of the protein distant from the engineered N-terminus, L1 and L2: the signals from latter regions were severely broadened in 2D ^{1}H-^{15}N HSQC spectra of dimers and tetramers, which further suggests their possible involvement in SQT-1C oligomerisation interface (Figure 2.3C). Additionally, no significant chemical shift perturbations or cross peak duplications were observed in 2D ^{1}H-^{15}N HSQC spectra of SQT-1 oligomers indicating that both dimeric and tetrameric species have symmetrical topologies and are formed by self-association of SQT-1C monomeric units.

To probe the local conformational stability of freshly-prepared SQT-1C monomers, residue-specific hydrogen/deuterium (H/D) exchange data on a timescale of minutes were collected using fast acquisition methods and fitted to exponential decays, to obtain exchange rates and protection factors (Figure S2.6). The H/D exchange for the whole protein occurs within less than an hour, suggesting that the whole protein fold is destabilised. However, we also observed some site-specific variations of the exchange times, with the amides located on the surface of the protein and/or in the loop regions exchanging within minutes or seconds. These data supported our hypothesis that SQT-1C has a highly flexible structure undergoing conformational exchange, consistent with spontaneous partial opening.

2.3.4 Monomeric SQT-1C self-associates into oligomers through inserted peptide loops

Self-association of isolated monomeric SQT-1C in solution was monitored by acquiring a series of ^{1}H-^{15}N HSQC spectra over a period of 36 h. During this time, approximately 70% of the signal intensity was lost due to oligomerisation, but no chemical shift perturbations were observed (Figure S2.7). Residue-specific ratios of signal intensities at the end and at the beginning of the experiment (I_f/I_0) report on signal loss due to conformational exchange and increases in rotational correlation time as a consequence of oligomerisation (Figure 2.4). The highest I_f/I_0 values were observed for residues that form β-sheets and the α-helix indicating conformational stability of these regions. Conversely, I_f/I_0 values below average were observed for residues in the N-terminus, L1 and L2 indicating regions most affected by conformational exchange and therefore most likely involved in the process of oligomerisation. Loops connecting α1 and β2 and β3 to β4 exhibited slightly lower I_f/I_0 values than the secondary structure elements. Additionally, the most severe signal broadening was observed for residues Q46-T58 and F77-T96 which

correspond to inserted peptides in L1 (D49-L54) and L2 (E81-R93) and residues in their immediate proximity, as shown in Figure 2.4b. Overall, the NMR experiments performed indicated that inserted loops together with the N-terminal are involved in SQT-1C oligomerisation.

The slow rates of signal loss and oligomerisation imply a relatively slow loop-mediated structural rearrangement leading to symmetrical association of monomers in a non-domain-swapped (NDS) or a domain-swapped (DS) manner. Although oligomerisation via domain swapping has been described for the cystatin protein family[285-286], V48D mutation was engineered in SQT to prevent cystatin-like domain swapping[51, 56]. As the broadening of NMR signals upon oligomerisation and symmetrical nature of SQT-1C oligomers made it impossible to distinguish between NDS and DS mechanisms from NMR data alone, we have explored the relative stability of different oligomer models using MD simulations, to gain insight into the SQT-1C oligomerisation pathway.

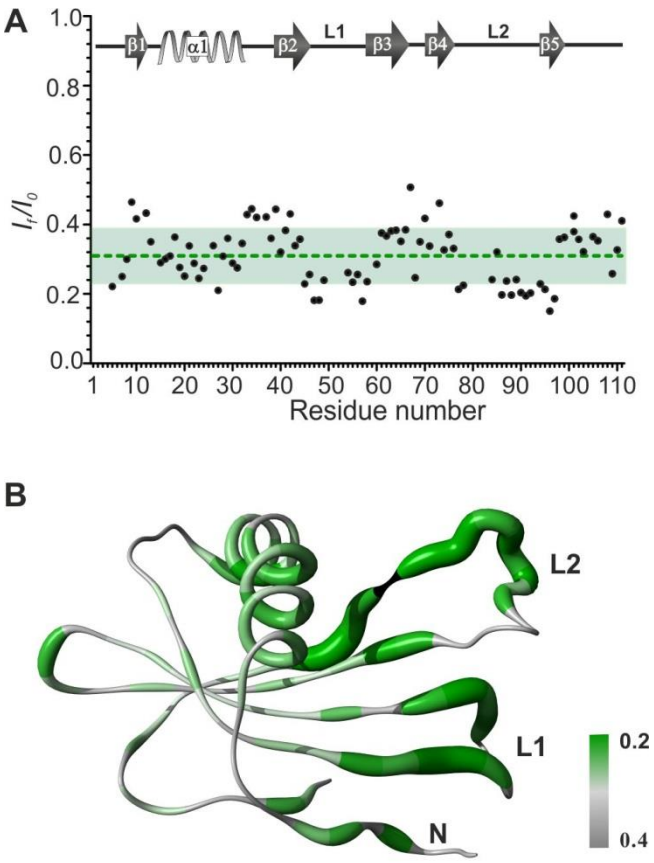

Figure 2.4: NMR signal perturbation mapping of SQT-1C oligomerisation interface. A) The ratio between signal intensity after 36 h incubation at 25 °C (I_f) and intensity immediately after monomer fraction isolation (I_0) plotted against residue number. The mean I_f/I_0 ratio is depicted with a green line and standard deviation of mean depicted in the green region. The position of elements of the secondary structure is annotated for reference. B) Mapping of oligomerisation interface. I_f/I_0 values per residue for SQT-1C are represented on the structure. The thickness of ribbons is inversely correlated to the I_f/I_0 ratio. Residues with lower than average I_f/I_0 ratio are coloured in green while the rest are depicted in grey. Unresolved, unassigned, and proline residues are shown in black.

2.3.5 Domain swap is the preferred pathway of SQT-1C oligomerisation

Initial DS SQT-1C dimer and tetramer structures were obtained by homology modelling based on the domain-swapped stefin B structure (PDB ID 2OCT)[287] using I-TASSER[288-290]. Monomeric unit of stefin B has 43.8% sequence identity to SQT-1C and 2OCT is, to the best of our knowledge, the only tetrameric structure of SQT-1C homolog available for use as a template. NDS dimer and tetramer SQT-1C models were obtained by ab-initio docking using the multi-body interface of HADDOCK 2.2 webserver[223, 291-292]. Structures of all four SQT-1C tetramer models, together with their chain orientations are

shown in Figure S2.8 and Figure S2.9. For full details also see Supplementary Methods in section 2.7.3. All SQT-1C oligomer models were subjected to 120 ns continuous MD simulation. For systems that did not converge within this timeframe, MD simulations were extended to explore the convergence on a longer timescale. MD simulations allowed assessment of relative stability of different models, and for the stable models, to determine the geometry at the binding interface and identify residues crucial for the oligomerisation process.

Based on backbone RMSDs, structures of dimer 2 and tetramer of NDS cluster 2 were quickly disintegrating, indicating that such oligomers would not be stable in solution (Figure S2.10). The other models converged and were stable for the duration of the production runs. For these stable models, we have assessed the effect of oligomerization on loop flexibility, and mapped residues involved. In monomeric SQT-1C, N- and C- terminal residues together with inserted loops, represent regions of increased flexibility, as measured by root mean square fluctuations (RMSF). As expected, these regions exhibit lower RMSF values in the dimer simulations, and these are further decreased in tetramers, as the inserted loops together with terminal residues become constrained within the interaction interface (Figure S2.11). The same effect is also seen as a reduction of solvent accessible areas (SASA) for loop regions when compared to SASA values in monomeric SQT-1C. Whereas solvent exposure of both inserted loops in all of the NDS dimers is similar to that of the monomer, in NDS tetramers, it is decreased for L1 typically to 22-27% of that of a monomer (Table S2.2). In case of DS oligomers, L1 is entirely solvent inaccessible upon tetramer formation, and SASA is significantly decreased in DS dimer, while L2 solvent exposure is decreased to 21% when the DS tetramer is formed, but only marginally affected by DS dimerisation.

To further understand the SQT-1C intermolecular interactions at the atomic level, binding free energies of each amino acid residue in the SQT-1C oligomers were calculated using the molecular mechanics generalised Born solvent area (MM-GBSA) methods implemented in AMBER 16 software[293-294]. We found that while the formation of both NDS and DS SQT-1C oligomers is highly energetically favourable, the total interaction energy of DS oligomers is much higher, indicating that DS oligomers would be more stable than their NDS analogues (Table 2.2). This further suggests that the domain swap mechanism might be the preferred pathway of SQT-1C oligomerisation. Detailed analysis of total free binding energy decomposition revealed that the unfavourable desolvation component (ΔG_{GB}) is compensated for with favourable electrostatic energy (ΔE_{EEL}), van der Waals interaction

contribution (ΔE_{VDW}) and non-polar solvation energies (ΔE_{NPOL} and ΔE_{SURF}) (Table S2.3). This implied that both NDS and DS oligomerisation pathways are driven by both electrostatic and hydrophobic interactions. To assess whether SQT-1C dimer models are consistent with the electron density observed in the crystal, we attempted to superimpose the DS SQT-1C homology model onto the SQT-1C crystal lattice (Figure S2.12). However, such a model was clearly incompatible with the packing observed in the crystalline lattice, encroaching into two of the symmetry-related molecules, in agreement with SQT-1C being monomeric in the crystal state.

Table 2.2: MM-GB binding free energy of SQT-1C oligomers

		ΔG_{TOT} (kcal/mol)
NDS Cluster 1	tetramer	-291±25
	dimer 1	-47±6
	dimer 2	-63±9
NDS Cluster 2	Tetramer*	NC
	dimer 1	-52±10
	dimer 2*	NC
NDS Cluster 3	tetramer	-177±19
	dimer 1	-29±7
	dimer 2	-34±5
DS	tetramer	-662±24
	dimer	-337±15

**denotes conformations that were not stable during MD simulations. No values were calculated for these conformations (NC)*

2.3.6 Inserted loops are crucial for SQT-1C oligomer formation

To further understand SQT-1C oligomerisation at the amino acid level, per residue free energy contributions to tetramerisation binding free energy were calculated using the MM-GBSA method and are shown in Figure 2.5A-D. Residues with an energy contribution larger than 2.5 kcal/mol were considered as interaction hotspots and are highlighted in Figure 5E-G. In the NDS oligomers, even though domain orientation between the three tetramer clusters varies, T50-I55, which are a part of the inserted peptide in L1 region of SQT-1C, were the amino acids with highest free energy contribution to the overall binding free energy. Conversely, the energy contribution of L2 residues, also located within the interaction interface, to the total binding energy is significantly smaller, indicating that in NDS

tetramers, L1 is more crucial for the oligomer formation. In comparison, for the DS dimer, residues L38-R65, comprising the β2, extended L1 and β3 of the SQT-1C monomer and lie on the interface of the two domain-swapped chains have, unsurprisingly, the most significant energy contributions to free energy of dimerisation. Meanwhile, residues located in L2 are crucial for the formation of DS tetramers.

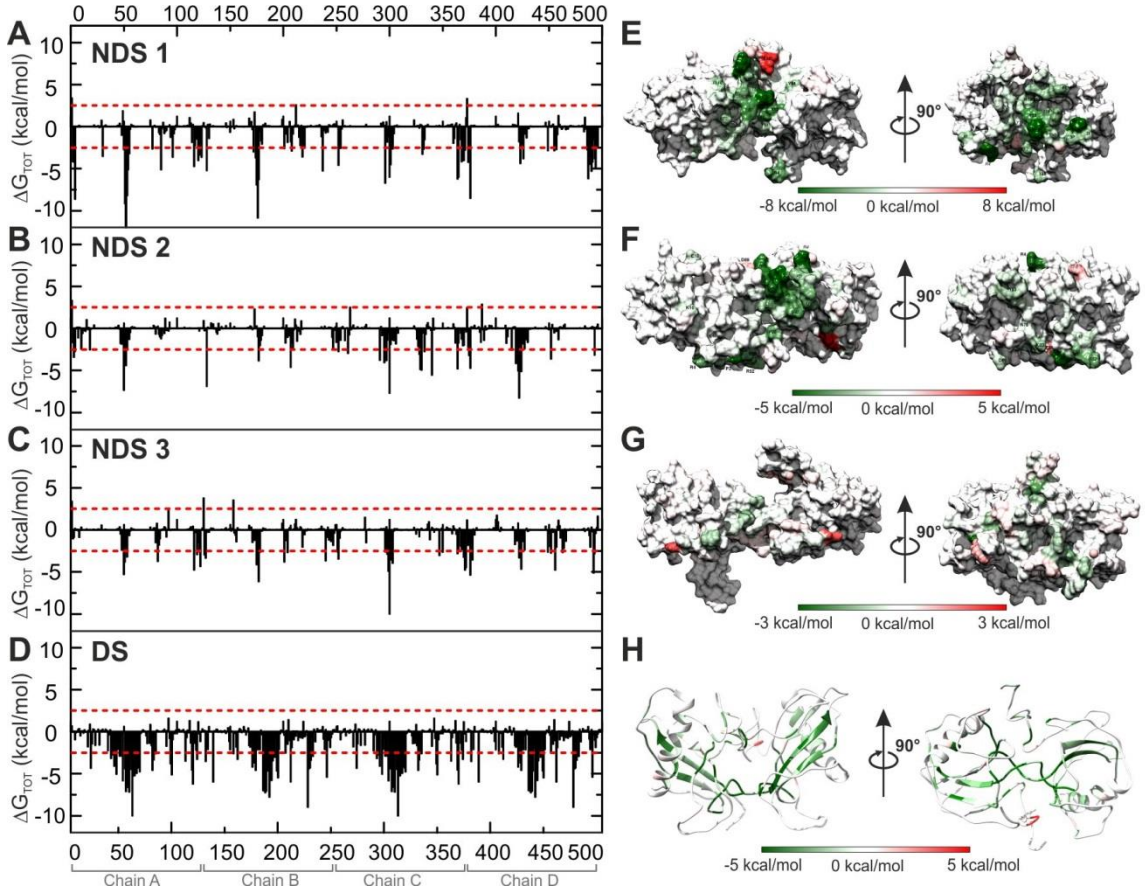

Figure 2.5: MM-GBSA calculated residue contributions to total binding energy of SQT-1C complexes formed by NDS and DS mechanism A-D) Per residue total free energy contribution of amino acids to the stability of A) NDS Cluster 1, B) NDS Cluster 2, C) NDS Cluster 3 and D) DS SQT-1C tetramers, calculated with the MM-GBSA method. e-h) Total energy contribution of each amino acid to the stability of tetramers projected to E) NDS Cluster 1, F) NDS Cluster 2, G) NDS Cluster 3 and H) DS SQT-1C tetramers. Both interaction interfaces are shown.

Overall, these results indicate that loop-mediated SQT-1C tetramerisation via both NDS and DS mechanisms is energetically favourable, and hence possible in principle.

However, the domain swapped mechanism results in much more stable oligomers, suggesting it is the preferred mechanism of SQT-1C oligomerisation in solution.

2.4 Discussion

The ability of scaffold proteins to retain their secondary and tertiary structure, and hence their structural, thermal and colloidal stability, upon insertion of various peptide loops needed for their target-binding function is critical for their industrial or research applications. Small frame of such scaffolds needs to absorb additional steric strains introduced by the modified loops, and resist forming alternative conformations, such as domain-swapped configuration. However, for small scaffold, these may be more difficult to achieve than for large-size mAbs. Although the general conservation of secondary structure of the scaffold proteins can be easily assessed by Far-UV CD spectroscopy, important structural perturbations upon peptide insertions may not always be reflected at the secondary structure level, and therefore detailed assessment of tertiary and quaternary structure is required.

We show here that in the presence of 19% dioxane SQT-1C exists as a monomer in solution and also crystallises as a monomer. However, in the absence of dioxane SQT-1C monomer readily forms dimers and tetramers, which co-exist in equilibrium. It is likely that the presence of high amounts of dioxane shifts the equilibrium towards the monomeric state[295]. The crystal structure reflects the dynamic nature of SQT with higher than average B factors and less than ideal geometries (Table S2.1). However, despite these shortcomings, the backbone fold and lattice interactions can be confidently discerned and have been extensively validated. The slow timescale of SQT-1C oligomerisation when starting from isolated monomeric form indicates structural rearrangement prior to oligomerisation, which is in line with a possible domain swap oligomerisation mechanism, typical for other proteins of the cystatin family[285, 296-297], despite SQT-1C being heavily modified in the loop regions in an attempt to engineer-out domain swapping[51, 56]. However, we observed that SQT-1C oligomerisation is spontaneous under standard buffer conditions and reversible, whereas the DS dimerisation of non-mutated cystatins is known to occur only at elevated temperatures, low pH or in the presence of chemical denaturants, is irreversible and leads to loss of function[57, 286, 298-299]. Furthermore, SQT-1C does not form oligomers bigger than a tetramer, whilst domain-swapped and non-domain swapped oligomerisation in the cystatin family leads to fibril formation[296, 300]. MM-GBSA calculations revealed that even though both NDS

and DS SQT-1C tetramers are in principle energetically favourable, the free binding energy of DS oligomers was much lower, suggesting that such a complex would be more stable in solution compared to its NDS analogue.

Domain swapping in proteins, including in the cystatin family, is a well-characterised process that is highly energetically favourable, but is traditionally believed to have a high activation energy due to large structural rearrangement, which can be overcome only in denaturing conditions[286, 301]. It has been shown previously, however, that certain point mutations in cystatins can significantly increase the rate of domain swapping and amylodiogenesis[87, 302-303]. The effect of point mutations in the protein core on domain swap oligomerisation has been extensively studied on various proteins, including the B1 domain of the immunoglobulin G binding protein (GB1)[304-305]. Malevanets *et al.*[306] simulated a monomer to domain swapped dimer transition of GB1 wildtype that does not form domain swapped oligomers in solution and its mutant that spontaneously oligomerises. They calculated a much flatter energy landscape than expected for this process and proposed that destabilisation of the protein core due to mutagenesis might be crucial for DS to occur. A similar principle may be applied to SQT-1C, where either backbone mutations of SQT scaffold, insertion of peptides or a combination of both results in a destabilisation of SQT structure that consequently lowers the energetic barrier for partial protein unfolding followed by domain swap, leading to a spontaneous formation of defined domain-swapped dimers which then form stable tetramers. It is reasonable to suggest that such a drastic conformational change driven by the inserted peptide relieves some of the strain introduced by an engineered loop. Consequently, conformation of the inserted peptide also drastically changes, from a hairpin to largely extended, which in turn influences its presentation for specific binding and recognition by the target protein. Further oligomerisation may cause an occlusion of the binding epitope. In the case of SQT-1C tetramers, both inserted loops appear hidden within the interaction interface. Therefore, the question of which of the distinct states, monomer, dimer or tetramer, has correct conformation for effective target binding will likely become important in determining the specific activity not only of SQT-1C, but also other engineered scaffolds which may suffer from the problems highlighted in the present case study.

In conclusion, we have used SQT-1C protein as a model system to study in detail the effect of peptide insertions on secondary, tertiary and quaternary structure and have deduced that domain-swapping driven by inserted peptides is the most probable mechanism of

forming well-defined dimers and tetramers. We also show that retained secondary structure does not necessarily mean that protein tertiary and quaternary structures are unperturbed. Therefore, protein aptamer designs should always be characterised at tertiary and quaternary structure levels, to reveal the conformation of the inserted target binding epitopes and identify which of the conformers, if several are present, are actually functionally competent. If the insertion of the peptide(s) leads to structural heterogeneity, approaches may be considered to stabilise "active" conformation[307-308]. We also suggest that routine assessment using 2D ^{1}H-^{15}N correlation NMR spectra would reveal early problematic behaviour of designs, helping to troubleshoot protein instability and loss of structure, and confirm the presence or absence of well-behaved stable protein fold. Such NMR-driven assessment in combination with routine SEC-MALS characterisation would enable even faster development of protein aptamers with high specific binding activity attributed to a defined conformation, while minimizing unwanted behaviour such as aggregation and long-term instability.

2.5 Materials and Methods

2.5.1 Plasmids

Synthesized codon-optimised gene construct encoding the SQT protein with AU1 insert in L1 and Myc insert in L2[56], named SQT-1, was obtained from GeneArt (ThermoFisher Lifetechnologies) and subcloned as two variants, with a cleavable hexahistidine tag at the N-terminus (SQT-1N) or C-terminus (SQT-1C). Full sequences are presented in the Supplemental Materials and Methods, section 2.7.3.

2.5.2 Protein expression and purification

Both protein constructs were expressed in E. coli BL21 CodonPlus RP competent cells (Agilent Technologies) in lysogeny broth (LB) supplemented with 50 µg/mL ampicillin and 34 µg/mL chloramphenicol (both Sigma Aldrich). LB was inoculated with 1% v/v overnight pre-culture. Cells were then grown at 37 °C with shaking until the optical density at 600 nm (OD_{600}) reached 0.8 at which point protein expression was induced with 0.5 mM IPTG. After overnight incubation at 37 °C cells were harvested by centrifugation at 6000 g for 15 min. Uniformly ^{15}N- and doubly ^{15}N and ^{13}C labelled samples were produced by growth in M9

minimal media supplemented with $^{15}NH_4Cl$ (99%), or both $^{15}NH_4Cl$ (99%) and ^{13}C-D-glucose (99%) (Sigma Aldrich), respectively.

Cell pellets were resuspended in denaturing buffer (20 mM NaPi, 500 mM NaCl, 6M GndHCl, pH 8.0) with 0.5% v/v Triton X-100 (Sigma Aldrich) followed by lysis by sonication with Sonopuls HD 3200 ultrasonic homogenizer equipped with TT13/F2 probe (Bandelin). The lysate was then clarified by centrifugation at 30000 g for 30 min at 4 °C. Supernatant was transferred onto Ni-NTA resin (Quiagen) in a gravity flow column and incubated for 90 min at 25 °C. The column was then washed with denaturing buffer supplemented with 10 mM imidazole. The bound material was eluted with 500 mM imidazole in denaturing buffer. Protein containing fractions were refolded by 1:10 rapid dilution in refolding buffer (20 mM NaPi, 150 mM NaCl, 5 mM EDTA, pH 7.2) followed by one step overnight dialysis into refolding buffer. Finally, the protein was purified on HiLoad 26/600 Superdex 200 pg column (GE Life Sciences) attached to ÄKTAPrime plus system (GE Healthcare Life Sciences), pre-equilibrated with refolding buffer. Both SQT variants eluted as a set of well-defined oligomers, which allowed isolation of monomeric, dimeric and tetrameric fractions, which were further concentrated using Vivaspin 20 centrifugal device with a 5 kDa molecular weight cut-off (Sartorius Stedium Biotech GmbH). Protein concentrations were estimated by absorbance at 280 nm ($\varepsilon = 14900\ M^{-1}\ cm^{-1}$). Molecular weights of SQT-1C oligomeric species were determined using size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) run at 25°C. 500 µg of SQT-1C was separated on Superdex 200 10/300GL column (GE Life Sciences) and passed through a Wyatt DAWN Heleos II EOS 18-angle laser photometer (Wyatt Technology) coupled to a Wyatt Optilab rEX (Wyatt Technology) refractive index detector. Data analysis was performed in ASTRA 6.1 software (Wyatt Technology).

Detailed explanation of SQT-1C and SQT-1N protein expression and purification protocol optimisation can be found in section 2.8.

2.5.3 X-ray crystallography

Crystal screening of SQT-1C was conducted by sitting drop vapour diffusion, by mixing 200 nL of protein at 20 mg/mL in buffer (20 mM HEPES pH 7.2, 150 mM NaCl) with an equal volume of reservoir solution and incubating at 4 and 21 °C. Despite broad screening, crystals only formed at 4°C when 38% v/v 1,4-Dioxane reservoir solution was added to the freshly-prepared tetrameric fraction of SQT-1C, giving the final concentration

of 19% dioxane in the droplet. Effect of 19% v/v dioxane on SQT-1C structure and oligomerisation state in solution was further assessed using NMR and SEC-MALS. Crystals were cryoprotected in perfluoropolyether cryo oil (PFO) prior to flash cooling in liquid nitrogen. Data were subsequently collected at io3 beamline at Diamond Light Source and scaled and merged with Xia2[309]. Preliminary phases were obtained by molecular replacement in Phaser[310] using a search model derived from Protein Data Bank (PDB) entry 3K9M. Iterative cycles of model building in Coot[311] and refinement in Phenix.refine[312] were used to generate the completed model. Validation with MolProbity[313] was integrated into the iterative rebuild and refinement process. Complete data collection and refinement statistics are presented in Table 2.1 and Table S2.1. The SQT-1C coordinates were deposited to PDB (ID 6QB2).

2.5.4 CD spectroscopy

Far-UV CD spectra were acquired on an Applied Photophysics Chirascan using a 0.01 cm path length quartz cell. The wavelength was varied from 190 to 280 nm with 0.5 nm step and acquisition time of 3 s per point. CD measurements of individual oligomeric species were performed immediately after SEC step in protein purification at a protein concentration of 1 mg/mL. Three scans were averaged and smoothed for each CD spectrum.

2.5.5 Intrinsic fluorescence

Melting temperature of SQT-1C was determined by measuring the change of intrinsic fluorescence maximum upon heating, using UNcle (Unchained labs). The temperature was varied from 20 to 90°C with a heating rate 1°C min^{-1}. Melting temperature was determined using the first derivative method.

2.5.6 NMR experiments

NMR samples were prepared by adding 5% v/v $^{2}H_2O$ to 1 mM ^{15}N- or ^{15}N,^{13}C-labelled protein solutions in refolding buffer. All NMR spectra were acquired at 25 °C on 800 MHz Bruker Avance III spectrometer equipped with 5mm triple resonance TCI cryoprobe with a temperature control unit. The spectra were acquired and processed using Bruker Topspin 3.5 and analysed using NMRFAM-SPARKY[314], Dynamics Center 2.2.4 (Bruker), and OriginPro 8.5.1 (OriginLabs).

NMR chemical shift assignment

For backbone assignment of SQT-1N, 2D ^{1}H-^{15}N HSQC and HSQC-based triple-resonance 3D HNCACB, HNCA, HNCO, HN(CA)CO and CBCA(CO)NH experiments from standard Bruker pulse program library were acquired using non-uniform sampling with multidimensional Poisson Gap scheduling strategy with sine bell weighting[315]. In addition, 3D TOCSY-HSQC and NOESY-HSQC with mixing times of 30 ms and 120 ms, respectively, were used to facilitate the assignment of SQT-1N. The backbone assignment of SQT-1N was then transferred to SQT-1C by matching peak positions and verified using 3D TROSY-based HNCA and HNCO spectra, together with 3D TOCSY-HSQC and NOESY-HSQC with mixing times of 30 ms and 120 ms, respectively.

Proton-deuterium exchange

After lyophilising the SQT-1C solution, 0.5 mL of $^{2}H_2O$ was added to SQT-1C, mixed vigorously and immediately transferred to NMR tube, and inserted into NMR spectrometer. A series of 10 BEST-TROSY experiments were acquired using pulse program b_trosyetf3gpsi.2 from standard Bruker library, with 6 min 20 s acquisition time per experiment. Decay of signal intensities was fitted to an exponential equation:

$$I(t) = A + Be^{-k_{HD}*t}$$

where A and B are arbitrary constants, $I(t)$ is the signal intensity at a given time and k_{HD} is the rate of H-D exchange. Intrinsic exchange rates (k_{ex}) were calculated based on SQT-1C structure using SPHERE[240], while protection factors (P) were calculated as k_{ex}/k_{HD} ratio.

2.5.7 Molecular dynamics simulations

Molecular system

The SQT-1C X-ray structure was used as an initial template into which the missing polypeptide stretches were added by multiple template homology modelling in I-TASSER[288-290]. SQT-1C oligomer models were obtained by ab-initio docking using the multi-body interface of HADDOCK 2.2 webserver[223, 291-292], see section 2.7.3 for full details. Three possible tetramer topologies were identified, and the lowest energy structure from each group was used in further MD simulations. SQT-1C dimers were obtained by separating tetramers into two subunits. SQT-1C domain swapped tetramer and dimer models were obtained by homology modelling using PDB ID 2OCT[287] as a template.

Molecular dynamics simulations

MD simulations were performed in Amber 16[293] with ff99SB force field[316]. Protonation states of individual amino acids at pH 7.2 and ionic strength of 150 mM NaCl were assigned using H++ 3.2 webserver[317-318] and kept constant throughout MD simulations. Models were placed into an octahedron box of TIP3P[319-320] water molecules with the box border of at least 10 Å away from atoms of the protein. Extra chloride or sodium ions were added to neutralize the charges of the protein complexes and achieve the final salt concentration of 150 mM NaCl. Prior to MD simulations, systems were subjected to a series of minimisation, heating and equilibration. The minimisation protocol started with 1000 steps of steepest descent minimisation followed by 4000 steps of conjugate gradient minimisation with 10 kcal mol^{-1} $Å^{-2}$ positional restraints on water molecules and backbone protein atoms. The system was then heated from 10 to 300 K during 300 ps followed by 700 ps of initial equilibration at 300 K at constant temperature and volume. This was followed by a 1 ns pre-equilibration at a constant temperature of 300 K and a constant pressure of 1 bar. The production simulations were carried out at constant temperature of 300 K, maintained using Langevin dynamics with collision frequency of 5.0 ps^{-1} in periodic boundary conditions. Electrostatic interactions were calculated by the particle mesh Ewald method[321-322] with the non-bonded cut-off set to 8 Å. The SHAKE algorithm[323] was applied to bonds involving hydrogen atoms. The simulations were run continuously for at least 120 ns with an integration step of 2 fs and snapshots written every 5 ps. For systems that did not converge within this timeframe, MD simulations were extended until convergence was reached for at least 20 ns. All trajectories were analysed using the CPPTRAJ module[324] of AMBER 16. For each model, the energy minimised structure (before the MD equilibration step) was chosen as a reference structure for RMSD and RMSF calculations.

The binding free energy calculation

The binding free energies of SQT-1C oligomers were calculated using the MM-GBSA method[325] implemented in the AMBER program. Snapshots at 4 ps intervals were extracted from the last 20 ns of the production MD runs and used for MM-GBSA analysis. In the GB calculation, variables α, β and γ were set to 0.8, 0.0 and 2.909225[326-327]. Binding free energy contributions of all residues were extracted using the energy decomposition scheme as implemented in AMBER[328]. Based on the individual contributions to the binding

free energy, amino acid residues with significantly higher than average energy contribution (> 2.5 kcal/mol) were considered as interaction hotspots. Solvent accessible surface areas (SASA) of all SQT-1C oligomer models were calculated using GETAREA webserver[329] and compared to SASA of monomeric SQT-1C.

2.6 Acknowledgements

We are grateful to Edward McKenzie and Ronald Burke from the University of Manchester Protein Production Facility for making the SQT-1N plasmid. We would also like to thank Matthew Cliff from the MIB NMR Facility for technical support with NMR spectrometers, Diana Ruiz Silvia from Biomolecular Analysis Facility for help with SEC-MALS, Derren Heyes from the MIB Biophysics Facility for technical support with the CD spectrometer, Pernille Harris from the Technical University of Denmark for co-hosting M.Z. during her secondment and discussions about crystallography data, and Jeremy Derrick from the University of Manchester for helpful comments on the manuscript. Simulations were performed at the GPU-cluster at the Department of Chemistry, Technical University of Denmark. This research was supported by EU Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement No 675074. The creation of plasmid coding SQT-1N was funded as a feasibility study “Protein biopharmaceuticals as NMR Targets” by the EPSRC Centre of Innovative Manufacturing in Emergent Macromolecular Therapies (UCL). We would also like to thank Diamond Light Source for access to beam line i03 (proposal MX17773-25).

2.7 Supplementary information

2.7.1 Supplementary Figures

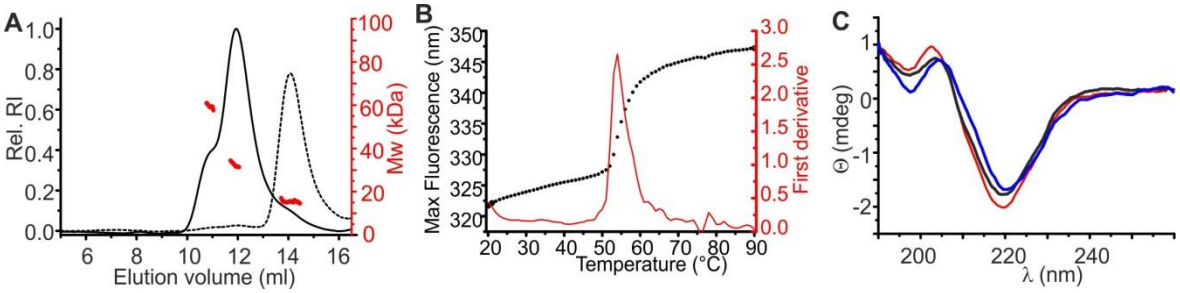

Figure S2.1: Biophysical characterisation of SQT-1C. A) SEC-MALS profiles of freshly isolated monomeric (dashed line) and dimeric-tetrameric (solid line) SQT-1C fractions. Relative refractive indexes are depicted in black, while molecular weight distribution across peaks is shown in red. B) Maximum intrinsic fluorescence of SQT-1C as a function of temperature. Melting temperature of 54 °C±1°C was determined by the first derivative method. C) Comparison of Far-UV CD spectra of isolated SQT-1C monomers (blue), dimers (black) and tetramers (red).

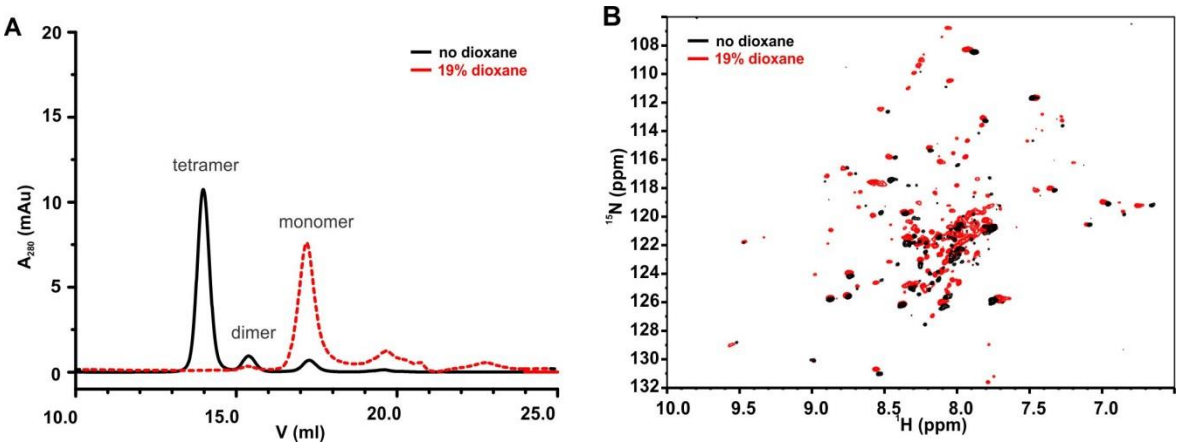

Figure S2.2: Biophysical characterisation of SQT-1C in the presence of 19% v/v dioxane. A) SEC traces of SQT-1C tetrameric fraction before (black) and after (red) addition of dioxane. B) Overlay of ^{1}H-^{15}N-HSQC spectra of tetrameric SQT-1C before (black) and after (red) addition of dioxane

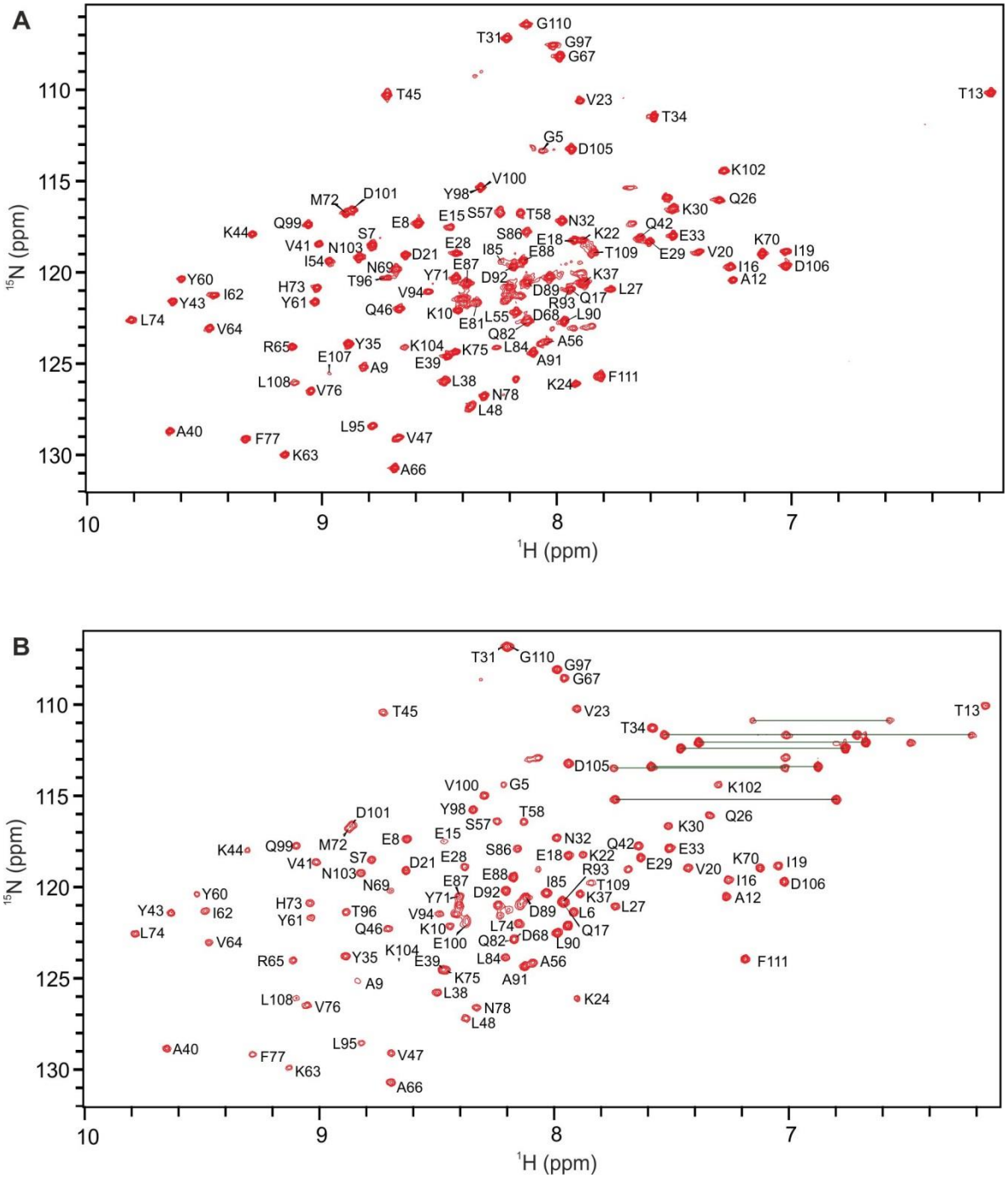

Figure S2.3: Backbone assignment of SQT-1C and SQT-1N. A) ^{1}H-^{15}N TROSY spectrum of freshly isolated monomeric SQT-1C and B) ^{1}H-^{15}N HSQC spectrum of SQT-1N recorded on 800 MHz spectrometer at 25 °C. The chemical shift assignments of backbone amide groups are annotated. Amine signals from the same sidechain groups appearing in ^{1}H-^{15}N HSQC spectrum are connected by horizontal lines

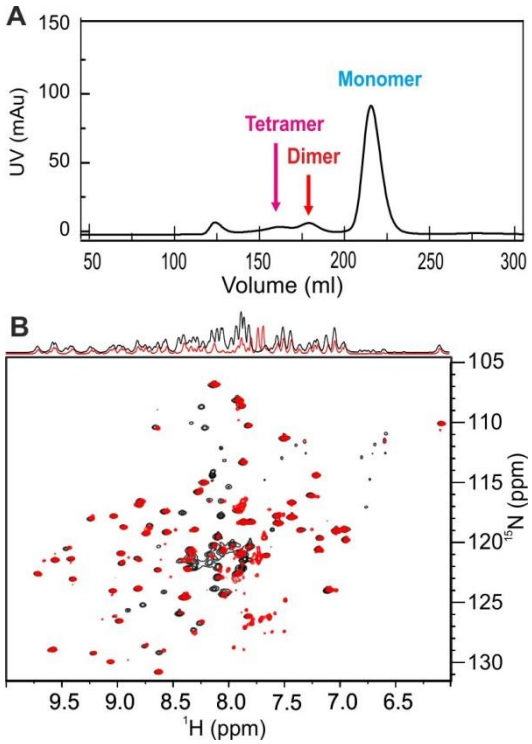

Figure S2.4: Biophysical characterisation of SQT-1N A) Purification SEC trace of SQT-1N using HiLoad 26/600 Superdex 200 pg column (GE Life Sciences). B) Overlay of ^{1}H-^{15}N-HSQC spectra of monomeric SQT-1N immediately after (black) and 48 hours after isolation (red). 1D projections of both ^{1}H-^{15}N-HSQC spectra are shown on top**.**

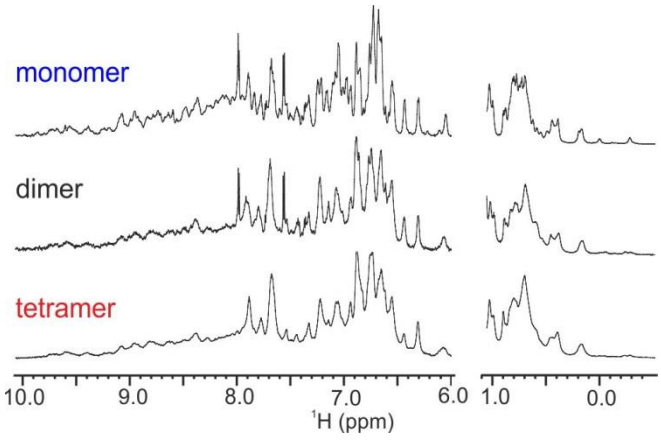

Figure S2.5: Comparison of aromatic and methyl regions of 1D ^{1}H spectra of freshly-separated SQT-1C monomers, dimers and tetramers

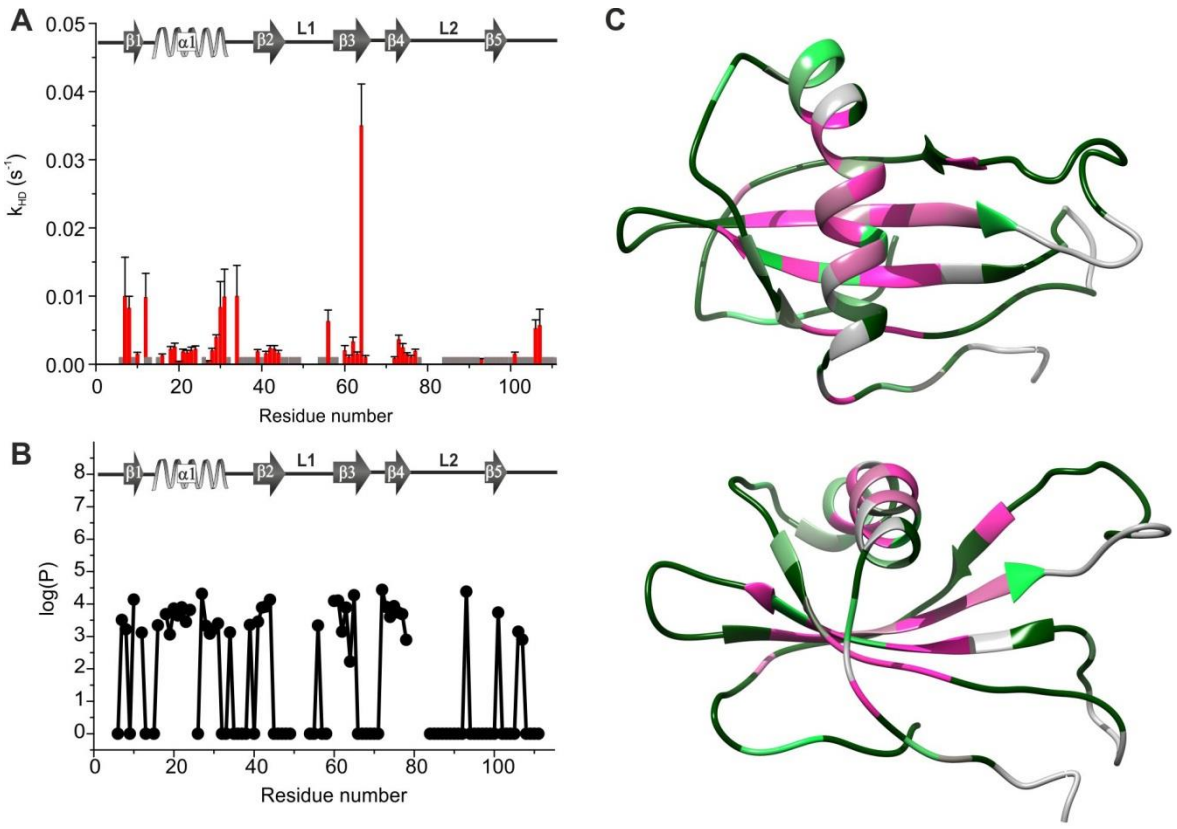

Figure S2.6: H/D exchange profile of SQT-1C. A) HD exchange rates determined by fitting signal intensities to exponential decay function plotted against residue numbers. Grey bars denote residues that were exchanged completely prior to collection of the first spectrum, while slower exchanging residues are denoted in red. B) Logarithm of protection factors (log(P)) values calculated as a ratio between observed and predicted exchange rates plotted against residue number. Value 0 was assigned to residues which exchanged completely prior to collection of the first spectrum. Gaps in data represent proline residues and unassigned amino acids. C) Top and side view of SQT-1C structure rendered by log(P) values. Residues for which protection factors could not be determined due to rapid H/D exchange are denoted in dark green, missing residues are represented in grey, while the rest of residues are rendered in green to magenta scale, where green represents more and magenta less solvent exposed residues.

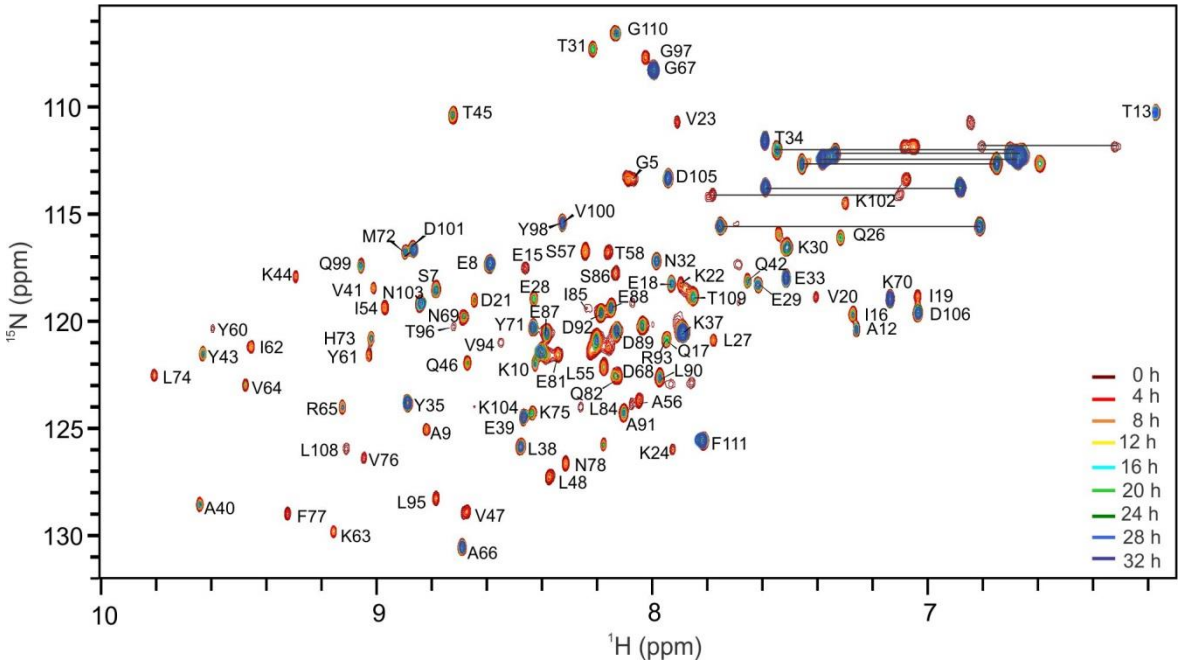

Figure S2.7: Overlay of ^{15}N-HSQC spectra of freshly prepared monomeric SQT-1C monitored over time.

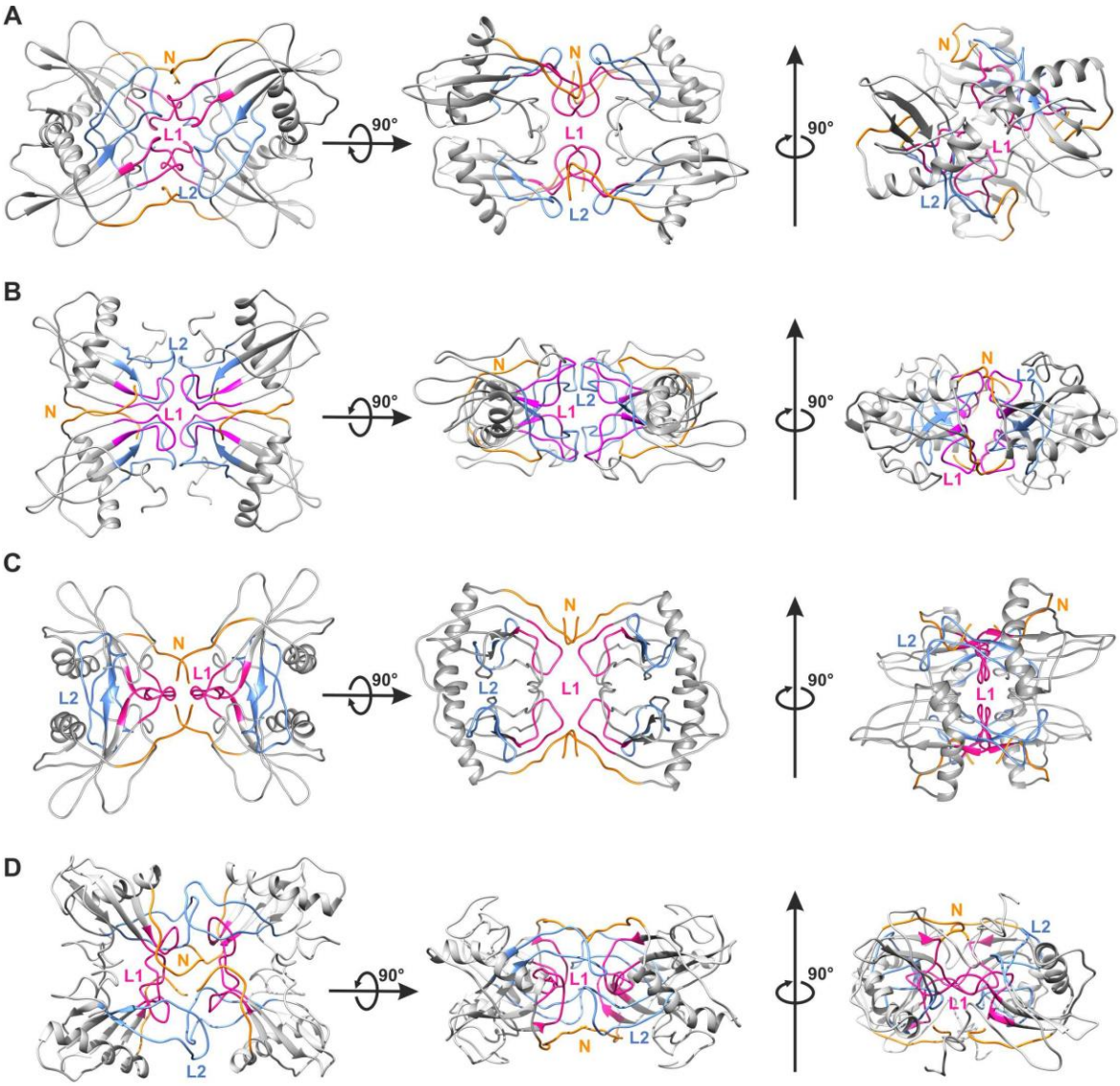

Figure S2.8: Possible SQT-1C tetramer models formed by NDS or DS oligomerisation. Representative structures for A) cluster 1, B) cluster 2 and C) cluster 3 of HADDOCK docking solutions for the NDS SQT-1C tetramers. D) Homology model of DS SQT-1C tetramer. N-terminal (N), loop 1 (L1) and loop 2 (L2) amino acid residues together with their neighbours that were identified by NMR as potentially involved in the binding interface of tetramers are shown in orange, pink and blue, respectively.

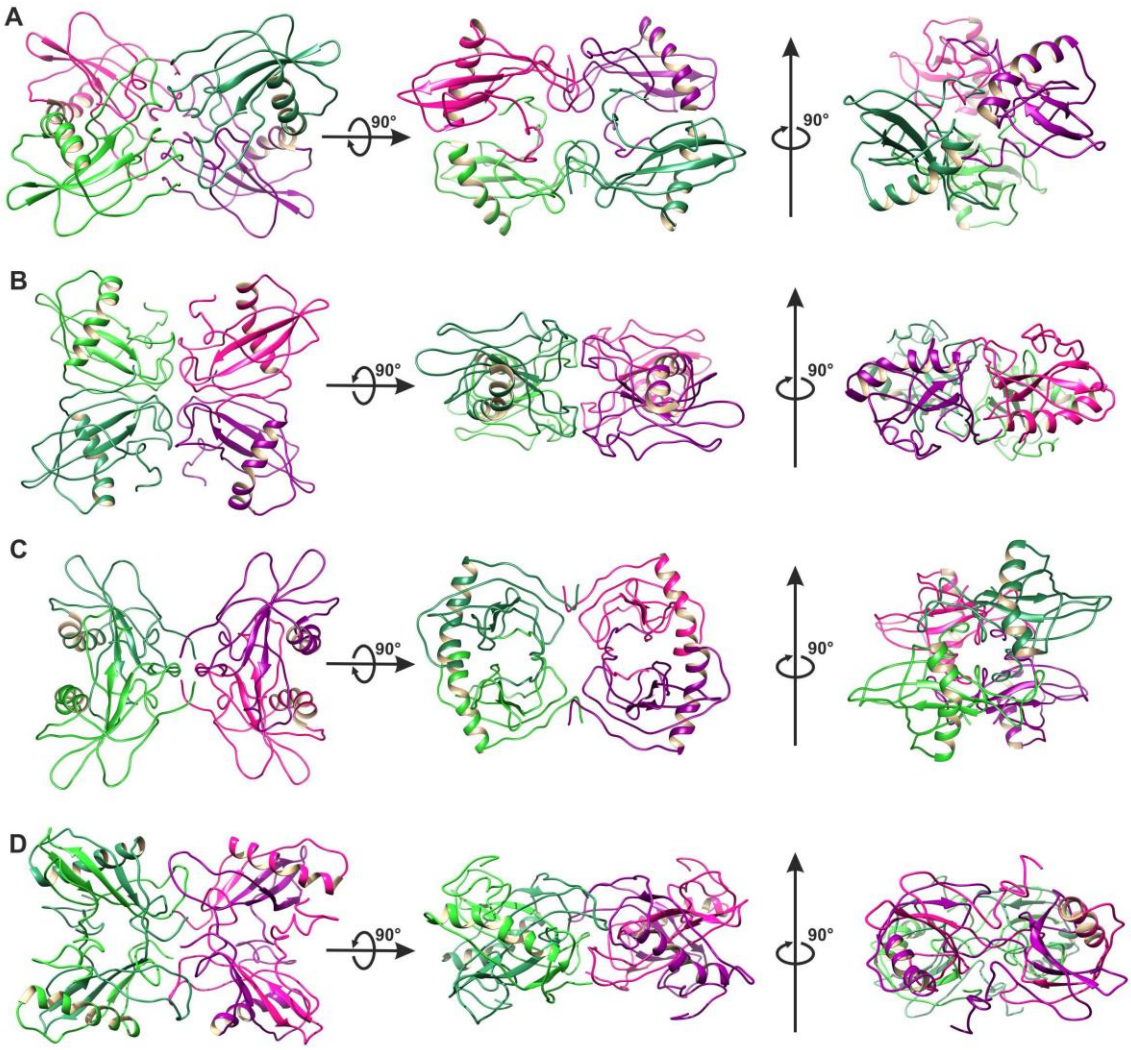

Figure S2.9: Polypeptide chain orientation in SQT-1C tetramer conformation formed by NDS and DS mechanisms. Representative structures for A) NDS cluster 1, B) NDS cluster 2 and C) NDS cluster 3 of HADDOCK docking solutions for the self-associated SQT-1C tetramers. D) Homology model of DS SQT-1C tetramer. The most probable dimers are indicated by the different shades of the same colour. Chains A, B, C and D are coloured in dark green, light green, dark magenta and magenta, respectively.

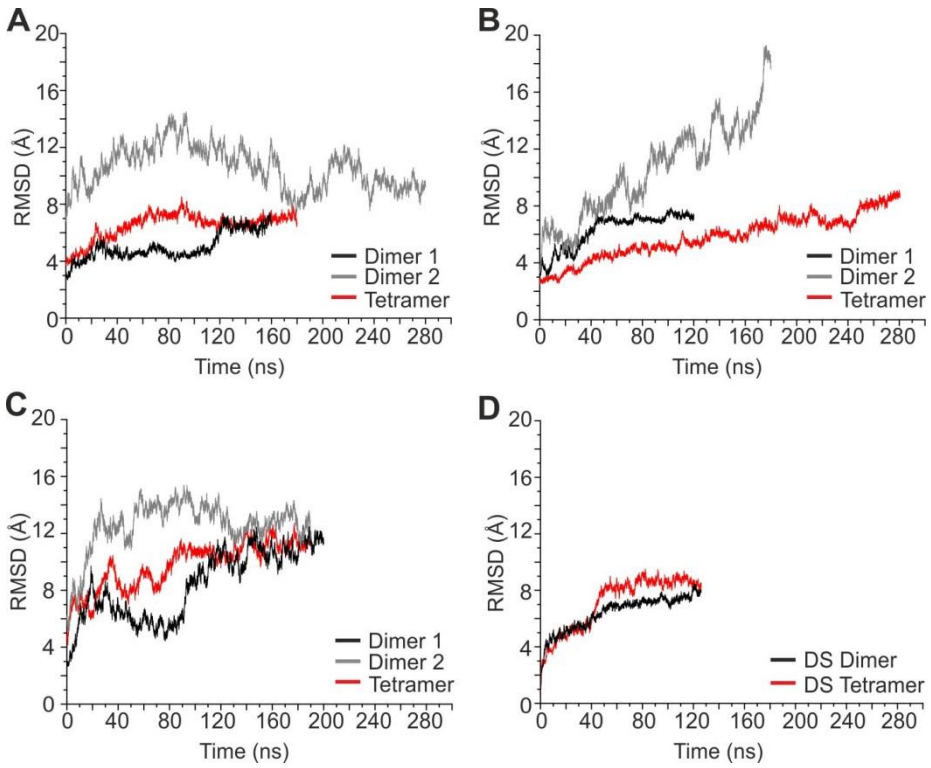

Figure S2.10: Backbone root mean square deviation (RMSD) during production runs of NDS A) Cluster 1; B) Cluster 2; C) Cluster 3; and D) DS SQT-1C oligomers. For each run, the RSMD was calculated with respect to the reference structure, which was arbitrarily chosen as the starting structure before MD equilibration and simulations. For systems that did not converge during 120 ns of a production run the simulations were extended to probe further convergence in the longer runs.

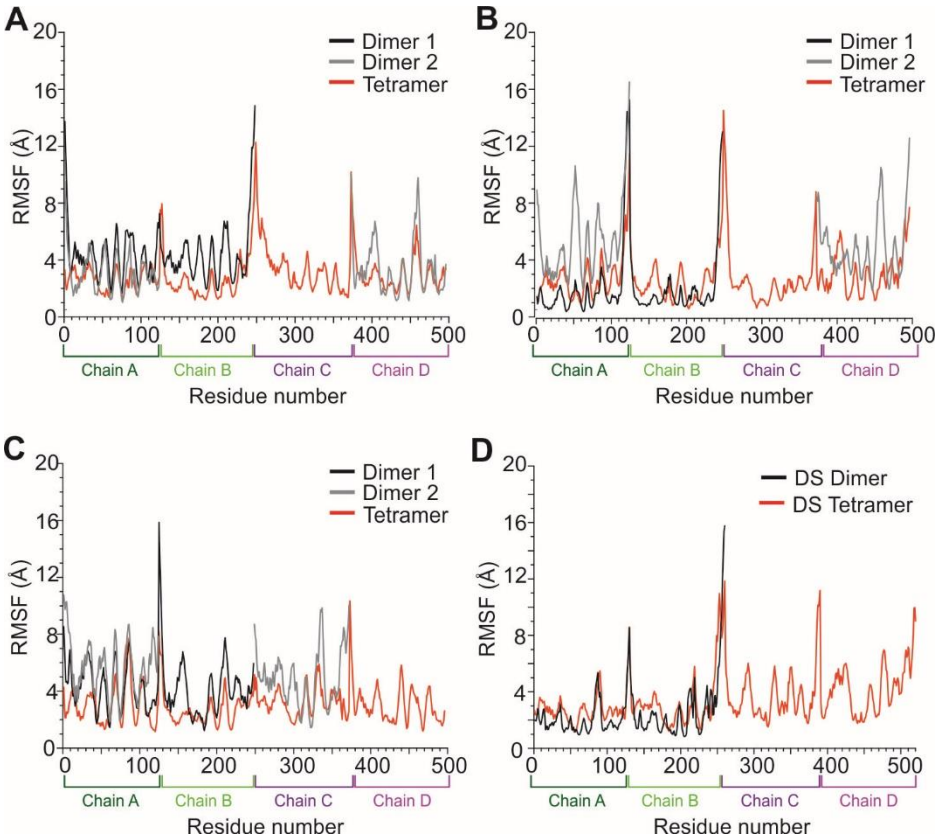

Figure S2.11: Root mean square fluctuations (RMSF) of NDS A) Cluster 1, B) Cluster 2, C) Cluster 3 and D) DS SQT-1C oligomers calculated based on heavy backbone atoms over the entire production runs

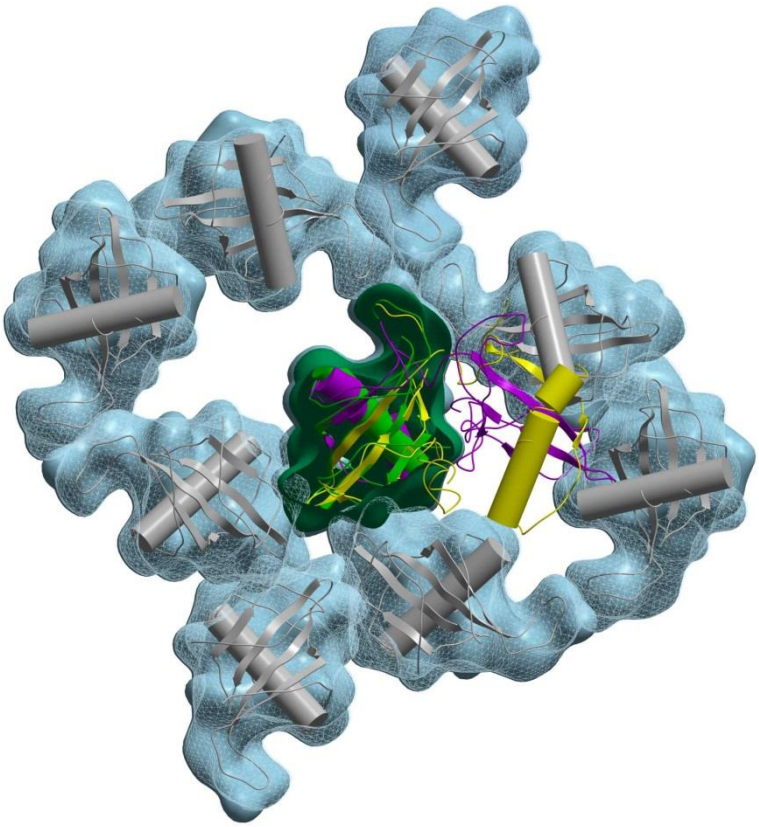

Figure S2.12: Predicted DS SQT-1C dimer is not compatible with SQT-1C crystal lattice. Two layers of the crystal lattice based upon the parent SQT-1C molecule (green) have been generated. The bottom layer is shown in blue surface representation, while the top layer is depicted with grey ribbon and mesh representation. The predicted DS SQT-1C (shown as ribbon presentation) has been superimposed on to the green central SQT-1C structure and is incompatible with the packing observed within the crystalline lattice, encroaching into two of the symmetry-related molecules.

2.7.2 Supplementary Tables

Table S2.1: Extended statistics for crystallography data collection and refinement for SQT-1C

	SQT-1C*
Data Collection	
Approximate crystal dimensions (μm)	150, 150, 100
Crystal mosaicity (°)	0.145
Wavelength (Å)	0.9
Space group	P 42_12
Unit cell dimensions	
a,b,c (Å)	96.124, 96.124, 29.8593
α,β,γ (°)	90, 90, 90
Resolution range (Å)	42.99 - 2.5 (2.589 - 2.5)
Total reflections	62236 (6263)
Unique reflections	5209 (497)
Multiplicity	11.9 (12.6)
Completeness (%)	99.79 (99.60)
$I/\sigma I$	17.39 (2.37)
R_{merge}	0.05822 (0.9239)
Wilson B-factor (Ang.^{2})	84.09
R_{meas}	0.06114 (0.9636)
R_{pim}	0.0181 (0.2712)
Refinement	
Reflections used in refinement	5207 (497)
Reflections used for R_{free}	256 (27)
CC1/2	0.998 (0.685)
*CC**	1 (0.902)
R_{work}	0.2725 (0.4411)
R_{free}	0.2892 (0.6248)
CC_{work}	0.877 (0.551)
CC_{free}	0.844 (0.290)
Number of non-hydrogen atoms	753
Protein	753
Protein residues	95
R.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.63
Ramachandran	
Favoured (%)	92.31
Allowed (%)	4.4
Outliers (%)	3.3
Rotamer outliers (%)	6.33
Clashscore	4.01
B-factors (Å^2)	
Average	104.03
Protein	104.03
Number of TLS groups	1
**Statistics for the highest resolution shell are shown in parentheses*	

Table S2.2: Relative solvent-exposed surface areas of engineered loops in SQT-1C oligomers

	Loop 1	***Loop 2***
NDS Cluster 1		
Tetramer	22	72
Dimer 1	90	96
Dimer 2	120	79
NDS Cluster 2		
*Tetramer**	NC	NC
Dimer 1	80	96
*Dimer 2**	NC	NC
NDS Cluster 3		
Tetramer	27	86
Dimer 1	96	84
Dimer 2	140	100
Domain swap		
Tetramer	0	21
Dimer	24	84
*All units are given in % relative to SASA of SQT-1C monomer. The * denotes conformations that were not stable during MD simulations; no values were calculated for these conformations*		

Table S2.3: MM-GBSA Binding free energy components of SQT-1C oligomers

	[1]ΔG_{TOTAL}	[2]ΔE_{VDW}	[3]ΔE_{EEL}	[4]ΔE_{EGB}	[5]ΔE_{SURF}
NDS Cluster 1					
Tetramer	-291±25	-473±20	-1395±123	1643±108	-66±2
Dimer 1	-47±6	-53±7	-476±53	491±52	-9± 1
Dimer 2	-63±9	-99±8	-310±58	360±56	-14±1
NDS Cluster 2					
*Tetramer**	NC	NC	NC	NC	NC
Dimer 1	-52±10	-91±7	-341±73	396±72	-15±1
*Dimer 2**	NC	NC	NC	NC	NC
NDS Cluster 3					
Tetramer	-177±19	-351±19	-31±115	251±107	-47±3
Dimer 1	-29±7	-80±7	116±56	-54±48	-10±1
Dimer 2	-34±5	-45±6	-62±28	80±26	-7±1
Domain swap					
Tetramer	-662±26	-838±30	-1002±56	1268±53	-110±2
.....Dimer	-336±13	-420±15	-501±56	640±53	-56±2

All units are given in kcal/mol

[1] ΔG_{TOTAL} *Total binding free energy*

[2] ΔE_{VDW} *van der Waals interaction energy*

[3] ΔE_{EEL} *electrostatic energy*

[4] ΔE_{EGB} *General-Born polar solvation energy*

[5] ΔE_{SURF} *General-Born non polar solvation energy*

Table S2.4: Restraints used for HADDOCK 2.2 docking. Active and passive residues used in the definition of the ambiguous distance restraints, non-crystallographic symmetry restraints (NCS) used to ensure same conformation of all chains and symmetry restraints used to enforce symmetry of the tetramer.

Restraint type	
Active residues	1-10, 46-59, 76-96
Passive residues*	11-12, 15, 18-20, 22-23, 26-28, 30, 41-45, 60-62, 68, 73-75, 97-98, 104-105, 107-112, 113-117, 120-130
NCS restraints	A=B, B=C, C=D
Symmetry restraints**	A=B, A=C, A=D, B=C, B=D, C=D
**All surface exposed residues that are within 6.5 Å of active residues* ***symmetry was imposed on the whole protein (residue ID 1-130)*	

Table S2.5: Structural statistics of the representative clusters from each NDS structural cluster of SQT-1C tetramers

	Cluster 1	**Cluster 2**	**Cluster 3**
HADDOCK score	-692 +/- 7	-653 +/- 47	-654 +/- 13
RMSD from overall lowest-energy structure (Å)	3 +/- 1	14 +/- 1	18 +/- 1
VDW energy (kcal/mol)	-339 +/- 6	-259 +/- 32	-361 +/- 19
Desolvation energy (kcal/mol)	-240+/- 24	-78+/- 16	-318 +/- 21
Electrostatics energy (kcal/mol)	-1120 +/- 71	-1941+/- 123	-507 +/- 135
Restraints violation energy(kcal/mol)	1084 +/- 80	698 +/- 69	1243+/- 156
Buried surface area ($Å^2$)	9320 +/- 354	8372 +/- 262	8421 +/- 100

2.7.3 Supplementary Materials and Methods

Plasmids

The amino acid sequence of SQT-1C construct was:

```
1                10                20                30                40                50
MIPRGLSEAK PATPEIQEIV DKVKPQLEEK TNETYGKLEA VQYKTQVLDT
                 60                70                80                90                100
YRYILASTNY YIKVRAGDNK YMHLKVFNGP EQKLISEEDL ADRVLTGYQV
                 110               120               130
DKNKDDELTG FENLYFQSLE RYLEHHHHHH
```

The amino acid sequence of SQT-1N construct was:

```
-19              -9                1                11               21               31
MGSSHHHHHH SSGLVPRGSM IPRGLSEAKP ATPEIQEIVD KVKPQLEEKT
                 41                51                61               71               81
NETYGKLEAV QYKTQVLDTY RYILASTNYY IKVRAGDNKY MHLKVFNGPE
                 91                101               111
 QKLISEEDLA DRVLTGYQVD KNKDDELTGF
```

Tetramer docking simulations

The SQT-1C non-domain swapped (NDS) tetramers were obtained by protein-protein ab-initio docking using the multi-body interface of HADDOCK 2.2 webserver[223, 291-292]. The NDS SQT-1C tetramers were built from 4 monomeric units, where the minimized structure of SQT-1C was used as a starting structure for the docking. Ambiguous intermolecular restraints (AIR) that guided the docking in HADDOCK were defined as follows: For all four chains, the residues with the I_f/I_0 *ratio* below 0.25 were selected as active residues while surface residues within 6.5Å of active residues were defined as passive. In addition, non-crystallographic symmetry restraints and six pairs of C2 symmetry restrain were used to ensure symmetrical topology of the tetramer while enabling the system to adopt either D2 or C4 symmetry (Table S2.4). Standard HADDOCK protocol was used to run the simulations. A total of 10000 structures were generated during the rigid-body docking, and the best 400 structures were subjected to semi-flexible and explicit solvent refinement.

Twenty-nine clusters of solutions were obtained with a 0.6 Fraction of Common Contacts (FCC) cut-off. Structural analysis of representative structures from each cluster revealed that SQT-1C could adopt three different tetramer topologies with similar energies and similar size of the buried surface area. Representative structures from each group with highlighted binding interfaces are shown in Figure S2.8; the orientation of monomeric units within each complex is shown in Figure S2.9 while their full statistics of docking can be found in Table S2.5. NDS SQT-1C tetramers in Cluster 1 exhibit D2 symmetry (Figure S2.8 and Figure S2.9). First interaction surfaces forms between C-terminal residues and L2. The second interaction surface forms through association of L1 loops. The tetramer is further stabilised by contacts between N-terminal residues and L2. Cluster 2 exhibits C4 symmetry with contacts between L1 and L2, forming both interaction surfaces. Additionally, the complex is stabilised by interactions between L2 and N-terminal residues (N). (Figure S2.8b and Figure S2.9b) Similarly to Cluster 1, Cluster 3 also exhibits D2 symmetry, with the interaction between L2, C-terminal residues and L1 forming one of the binding interfaces, while the other comprises only of contacts between N-terminal residues of binding partners (Figure S2.8c and Figure S2.9c).

2.8 Optimisation of protein expression and purification protocols

As described in section 2.5.1, synthesized codon-optimised gene construct encoding the SQT protein with AU1 insert in loop 1 and Myc insert in loop 2[56], named SQT-1, was obtained from GeneArt (ThermoFisher Lifetechnologies) and subcloned as two variants, with a cleavable hexahistidine tag at the N-terminus (SQT-1N) or C-terminus (SQT-1C). As the initial tested purification protocol yielded insufficient protein amounts, expression and purification protocols were optimised for both SQT-1N and SQT-1C.

2.8.1 Optimisation of SQT-1N expression

In the first attempt, SQT-1N was expressed in E. *coli* BL21 T7 express competent cells in M9 minimal media and purified using the previously reported procedure[56]. Total protein yield from E. *coli* BL21 T7 express culture was approximately 10 mg per 1L of M9 minimal media which was too low for NMR analysis and far less than previously reported[56]. Therefore, the optimisation of SQT-1N expression was carried out in M9 minimal media as described in Materials and Methods, section 2.8.7. In the first round of optimisation, E. *coli* BL21 (DE3), BL21 T7 Express (DE3), BL21 Rosetta™ 2(DE3) and BL21 CodonPlus RP+ strains were transformed with pHis vector containing SQT-1N sequence. Cultures were grown at 37°C with shaking, with OD_{600} of each culture recorded regularly throughout incubation to follow the bacterial growth. In all cases, SQT-1N expression was induced with 1 mM IPTG at OD_{600} 0.6. 1 ml samples of cell cultures were retained prior to and 20 h after induction for SDS-PAGE analysis of protein expression levels (Figure S2.13). Rosetta™ 2 (DE3) exhibited the slowest growth and reached the lowest optical density among the tested E. *coli* strains, but it had the highest total expression yield of 15 mg per 1L of M9 media.

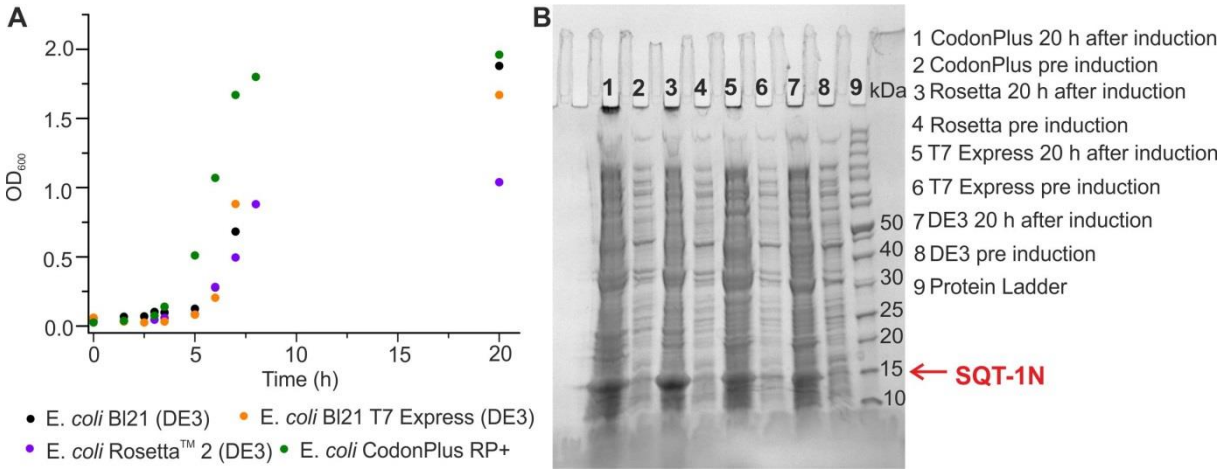

Figure S2.13: A) E. *coli* culture growth in M9 minimal media. The Optical density at 600 nm (OD_{600}) of the E. *coli* BL21 (DE3), Rosseta™ 2 (DE3), BL21 T7 Express (DE3) and CodonPlus RP+ throughout incubation at 37 °C. B) SQT-1N expression yield in different E. *coli* strains. Total expression yield in E. *coli* BL21 (DE3) (8,7), Rosseta™ 2 (DE3) (6,5), BL21 T7 Express (DE3) (4,3) and CodonPlus RP+ (2,1) cell cultures was analysed and compared by SDS-PAGE analysis of samples taken prior to and 20 hours after induction.

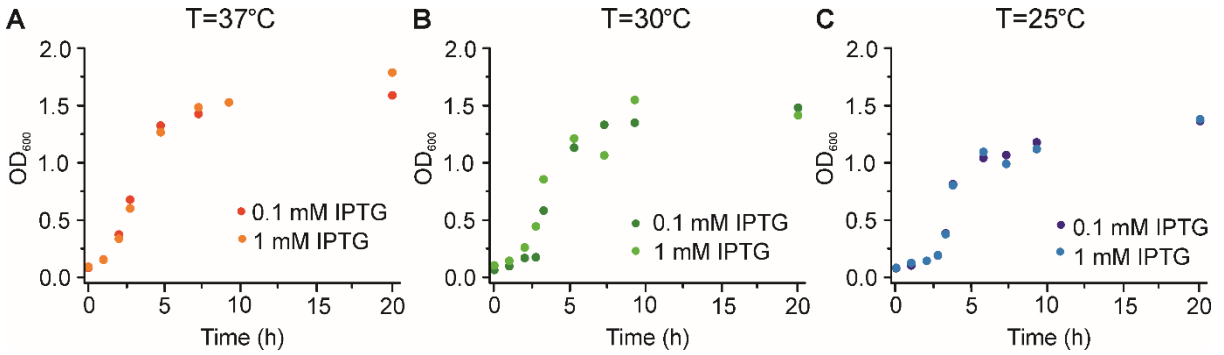

Figure S2.14: Growth of E. *coli* Rosetta™ 2 (DE3) at different conditions. Cultures were grown at A) 37 °C, B) 30°C and C) 25°C, and induced with either 0.1 mM or 1 mM IPTG.

In the second round of SQT-1N expression optimisation, the optimum temperature of growth and optimum IPTG concentration were determined. Rosetta™ 2 (DE3) cultures were grown at 37, 30 or 25 °C and induced with either 0.1 or 1 mM IPTG (Figure S2.14). The fastest growth and highest optical density of Rosetta™ 2 (DE3) cultures was observed at 37 °C with the similar plateau of growth reached at both IPTG concentrations. Cultures grown at 30 °C reached similar optical density that those grown at 37 °C, while those growth

at 25 °C resulted in significantly lower optical density. To compare total expression yields in different growing conditions, samples of each cell culture taken prior to and 20 h after induction were analysed by SDS-PAGE (Figure S2.15A). The highest total expression yield of SQT-1N was observed in the Rosetta™ 2 (DE3) cultures grown at 30°C and induced with 0.1 mM IPTG.

To check whether SQT-1N is fully soluble in the cytoplasm or retained in inclusion bodies, samples of Rosetta™ 2 (DE3) cultures obtained 2, 4, 6 and 20 hours after induction were treated with BugBuster Protein Extraction Reagent following manufacturer's instructions to obtain the soluble protein fraction and isolate inclusion bodies. Samples were than prepared for SDS-PAGE analysis according to manufacturer's protocol (Figure S2.15B). The concentration of SQT-1N in soluble fraction reached its maximum 4 hours after induction and did not change significantly with further incubation. On the other hand, concentration of SQT-1N in inclusion bodies increased with time, reaching its maximum 6 hours after induction and did not change significantly with further cell growth. 20 h after induction, approximately 50% of SQT-1N was soluble in the cytoplasm, while the rest was packed into inclusion bodies. These results suggested that further purification should be done under solubilisation conditions to recover SQT-1N from inclusion bodies.

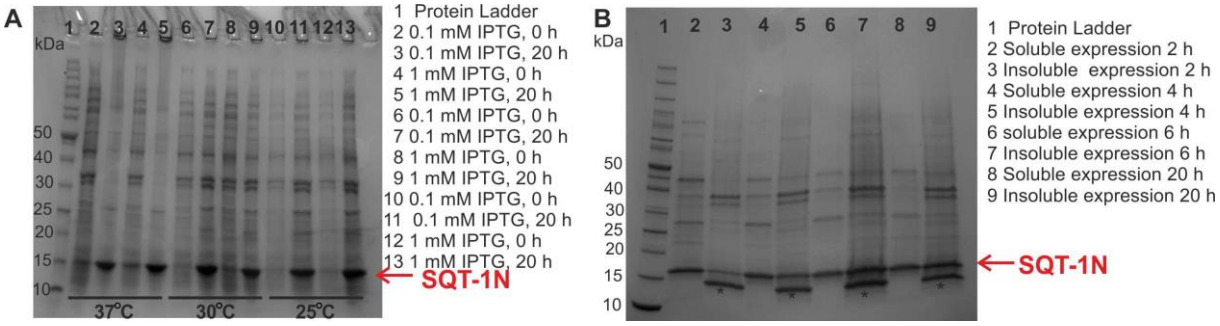

Figure S2.15: A) SDS PAGE of total SQT-1N expression in Rosetta™ 2 (DE3) cells in different growth conditions. Samples of cell cultures grown in different conditions were retained prior to (denoted as 0 h) and 20 h after induction. B) SDS PAGE of soluble and insoluble SQT-1N expression in of Rosetta™ 2 (DE3) cells 2, 4, 6 and 20h after induction. Bands denoted with * correspond to lysozyme that was used during inclusion bodies purification

2.8.2 Optimisation of SQT-1N purification

Three different protocols for SQT-1N purification were tested; first using standard solubilisation with 6 M guanidine hydrochloride (GndHCl) followed by refolding step, second applying mild solubilisation in Tris-HCl buffer with addition of Triton X-100, which eliminated the need of the refolding step, and third using native purification protocol published by Stadler *et. al.*[56] In all three cases, the first step in purification was Ni^{2+} affinity chromatography using Ni-NTA agarose as described in section 2.8.7. Samples of unbound flow through, washing and 6 elution fractions were analysed using SDS-PAGE to determine efficiency of purification with Ni-NTA agarose (Figure S2.16). In all cases 90 min incubation was sufficient for binding of SQT-1N to Ni-NTA, negligible amount of SQT-1N was present in the wash fraction, and most of the SQT-1N was eluted from the resin after elution with 6 ml of elution buffer.

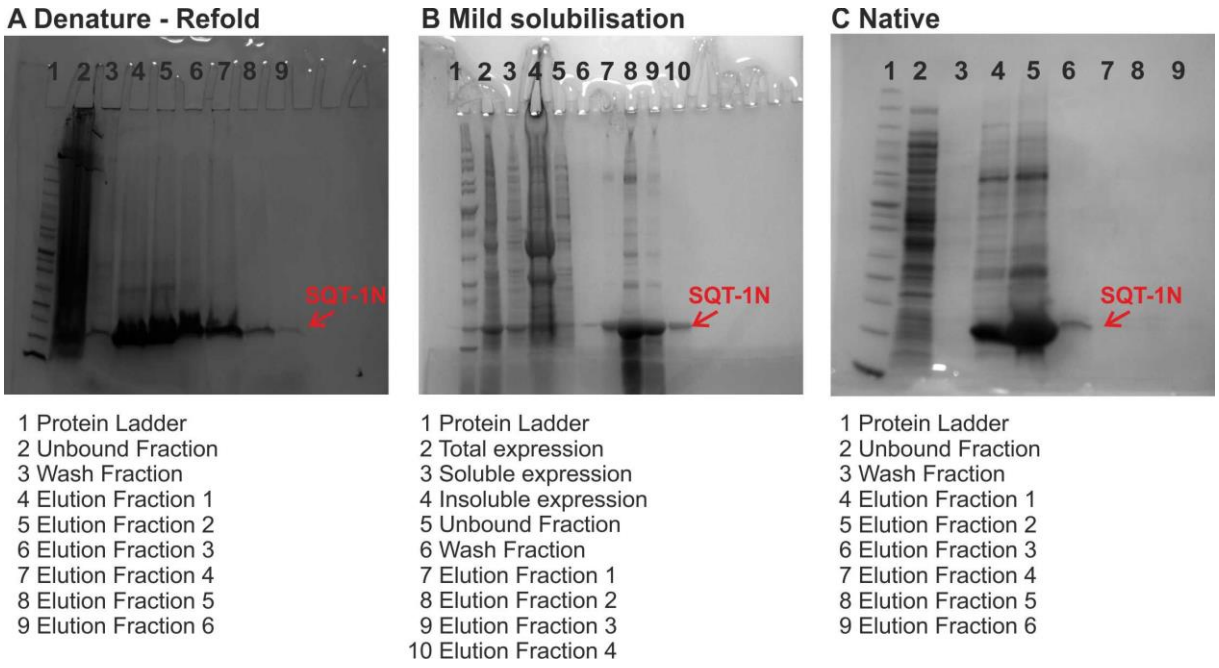

Figure S2.16: Analysis of Ni2+-affinity chromatography efficiency for A) Denature-Refold, B) Mild solubilisation and C) Native purification of SQT-1N

The second step in SQT-1N purification was size exclusion chromatography (SEC). In the mild solubilisation and native purification protocols, fractions containing SQT-1N after Ni^{2+} affinity chromatography were joined, centrifuged to remove any insoluble aggregates and injected directly onto the gel filtration column while in the denaturing protocol, eluted SQT-1N was first refolded by rapid dilution, dialysed against refolding buffer overnight, concentrated to 6 mL and centrifuged before injection onto the SEC

column. The next day, refolded SQT-1N was concentrated to 6 ml, briefly centrifuged to get rid of any aggregates and injected onto the SEC column pre-equilibrated into refolding buffer. It is important to note that a considerable amount of visible aggregates was present after centrifugation during denature-refold purification while no precipitation was observed during native and mild solubilisation purification. Traces obtained during SEC from native and denature-refold purifications are shown in Figure S2.17. In the denature-refold purification protocol, the majority of SQT-1N eluted as a monomer, with minor traces of oligomer formation. On the contrary, in mild solubilisation and native purification protocols majority of SQT-1N was eluted in oligomeric forms. These results suggested that the denature-refold purification protocol should be used for purification of SQT-1N.

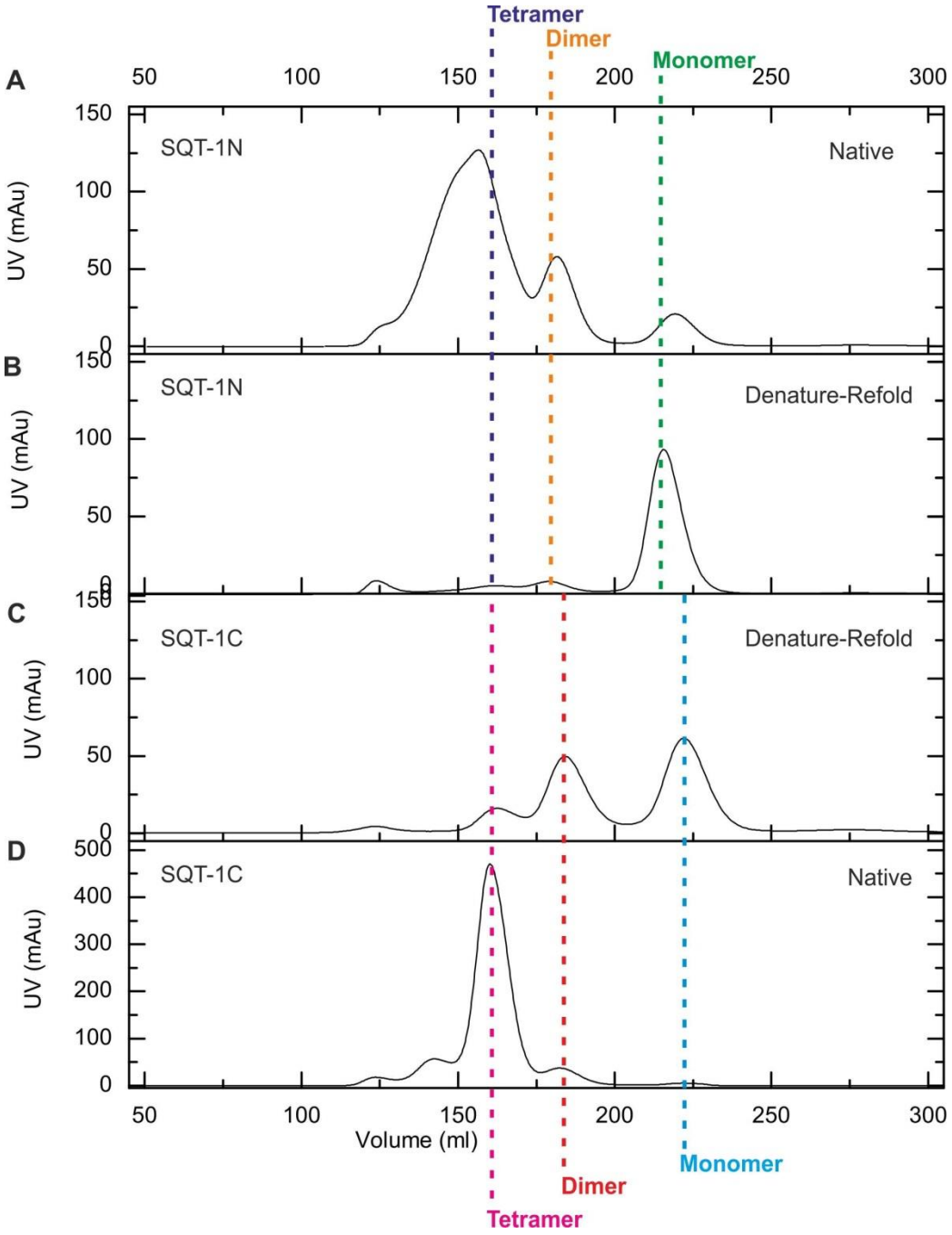

Figure S2.17: SEC purification traces for both SQT-1 variants. SEC traces of SQT-1N purified by A) Native and B) Denature-Refold protocols, and SQT-1C purified by C) denature-refold and D) native purification protocol. Purification in panel A was achieved using HiLoad 26/600 Superdex 200 pg column while HiLoad 26/600 Superdex 75 pg column was used in other purification protocols.

2.8.3 Assessing the stability of SQT-1N

Stability and folding of SQT-1N was assessed by NMR spectroscopy. Sharp, well-resolved peaks in amide-aromatic and methyl region in ^{1}H NMR spectrum of SQT-1N indicated that SQT-1N is completely folded. To assess the effect of pH on SQT-1N folding ^{1}H NMR spectra of unlabelled SQT-1N that was buffer exchanged into 20 mM Na-phosphate buffers, 150 mM NaCl, 50 mM ArgGlu, pHs 6.1, 7.2 and 7.4 were acquired (Figure S2.18A). Upon change of pH, minor changes in ^{1}H spectra were observed. However, there was insufficient evidence that changing pH would significantly improve the NMR spectrum. Hence, we decided to keep the pH at 7.2.

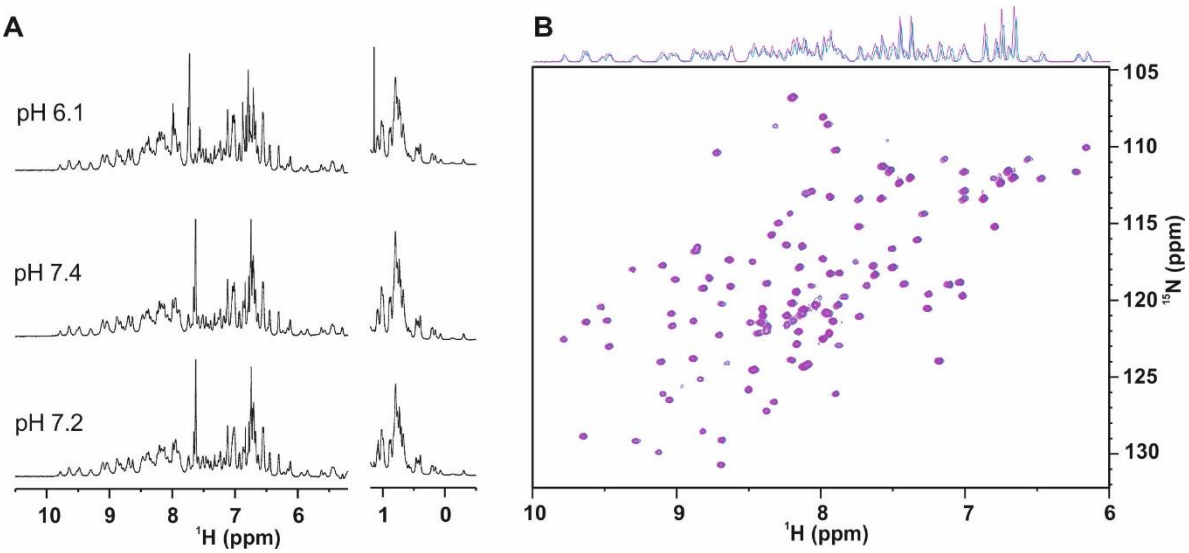

Figure S2.18: A) Amide-aromatic (left) and methyl-aliphatic (right) regions of ^{1}H NMR spectra of SQT-1N at different pH. B) ^{15}N HSQC of SQT in refolding buffer with 150 mM NaCl (pink) and without any NaCl (blue). 1D projections of the ^{1}H dimension are plotted above HSQC spectra. All spectra were recorded on the 800 MHz Bruker spectrometer at 25 °C. The concentration of all samples was 0.15 mM in their respective buffers, 5% $^{2}H_2O$ was added to each sample.

To analyse the effect of NaCl on the stability of SQT structure, one aliquot of 0.15 mM ^{15}N SQT-1N sample was buffer exchanged into refolding buffer without salt while the other aliquot was kept in the original refolding No significant changes were observed between ^{15}N HSQC spectra of both samples. However, comparison of 1D projections in the ^{1}H dimension revealed lower intensity of signals in the sample without salt (Figure S2.18B). To determine the short-term stability of SQT-1N with and without NaCl, ^{15}N HSQC spectra of both samples were acquired across 5 days. During this time no significant changes were

observed in ^{15}N HSQC spectra of the SQT-1N in refolding buffer with salt while the loss of signal intensities and peak broadening were noted after 2 days in the sample without NaCl indicating partial aggregation of SQT-1N (Figure S2.19A). The SQT-1N sample in the absence of salt ultimately precipitated in 5 days. This data suggested that NaCl stabilises SQT-1N and should be thus present in refolding and storage buffers.

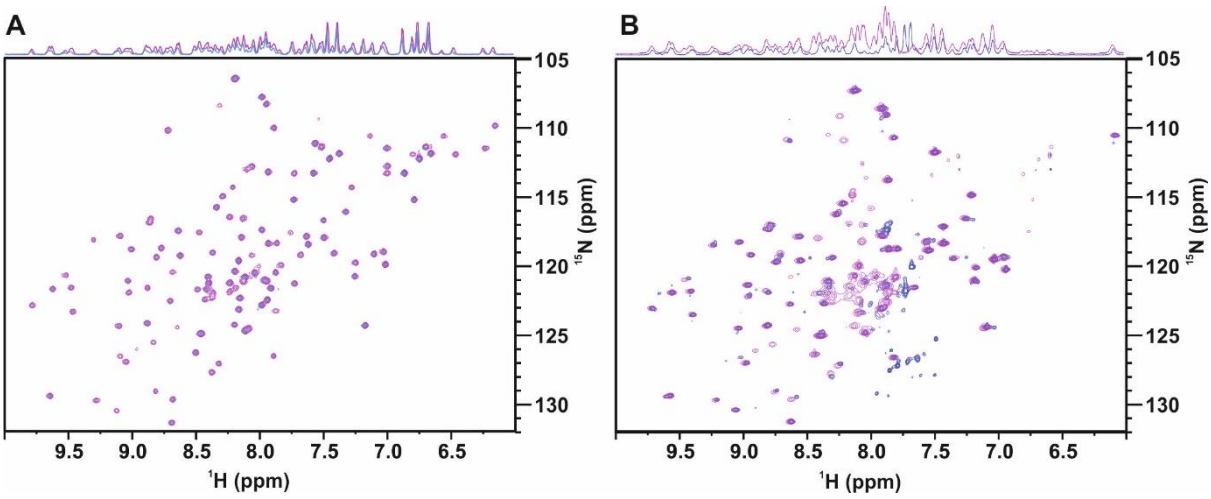

Figure S2.19: A) ^{15}N HSQC spectra of SQT in the absence of NaCl immediately (pink) and after 2 days (blue). B) ^{15}N HSQC spectra of SQT immediately (pink) and 2 days after (blue) concentrating. All spectra were recorded on the 800 MHz Bruker spectrometer at 25 °C. The concentration of samples in A and B was 0.15 and 5 mM, respectively, 5% $^{2}H_2O$ was added to each sample

As we wanted to study the stability of SQT-1N in highly concentrated solutions, needed for industrial applications, the SQT-1N was gradually concentrated to 5 mM. Upon concentration, chemical shifts perturbation and peak broadening were observed in ^{15}N HSQC spectrum compared to a spectrum of low concentration sample, which indicated oligomerisation. Almost complete oligomerisation of SQT-1N was observed after 2 days at high protein concentration (Figure S2.19B).

2.8.4 Optimisation of SQT-1C expression

First E. *coli* BL21 T7 Express (DE3), BL21 Rosetta™ 2(DE3) and BL21 CodonPlus RP+ strains were transformed with pET21a(+) vector containing SQT-1C sequence. Cultures were grown in 50 ml of Minimal M9 media at 30°C and 37°C in triplicates with shaking, where OD_{600} of each culture was recorded regularly throughout incubation to determine bacterial growth. SQT-1C expression was induced with 0.1 mM, 0.5 mM or 1 mM IPTG at OD_{600} 0.6. 1 ml samples were retained prior to and 16 h after induction for analysis of expression levels (Figure S2.20). All strains reached the plateau of growth at OD_{600} 1.95 regardless of the incubation temperature, but the highest total expression yield was observed in CodonPlus RP+ strain induced with 0.1 mM IPTG grown at 37 °C; therefore, these growth conditions were used going forward.

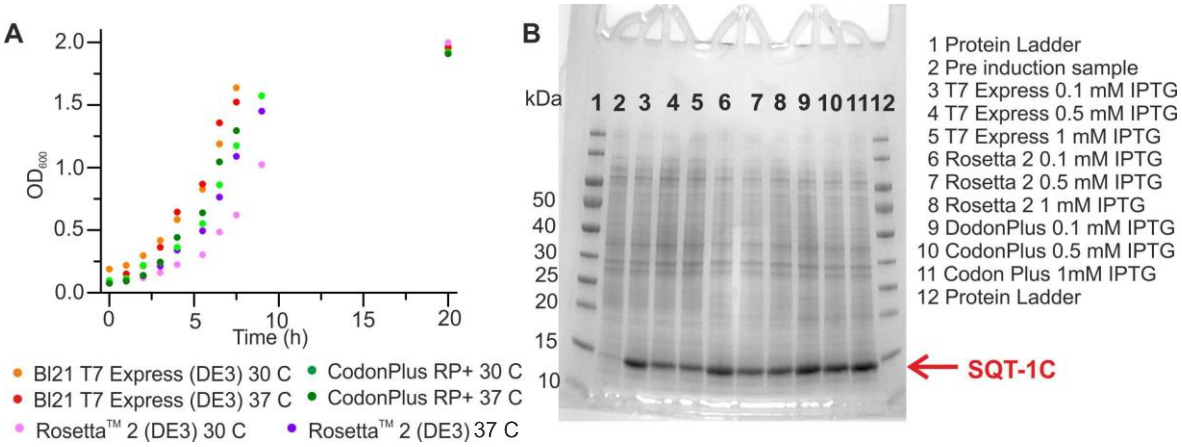

Figure S2.20: A) E. *coli* culture growth in M9 minimal media. The optical density at 600 nm (OD_{600}) of the E. coli Rosseta™ 2, BL21 T7 Express (DE3) and CodonPlus RP+ measured throughout incubation at 30 and 37 ℃ to determine culture growth of individual E. coli strains B) Total SQT-1C expression yield in different E.coli strains at 37 ℃ Total Expression yield in E. *coli* T7 Express (DE3), Rosseta™ 2 and CodonPlus RP+ cell cultures grown was analysed and compared by SDS-PAGE analysis of samples taken prior to and 16 hours after induction.

2.8.5 Optimisation of SQT-1C purification

Native and denature-refold purification protocols for SQT-1C purification were tested. Both protocols consisted of two same two steps as SQT-1N purification. The efficiency of purification with Ni-NTA agarose was analysed by SDS-PAGE (Figure S2.21).

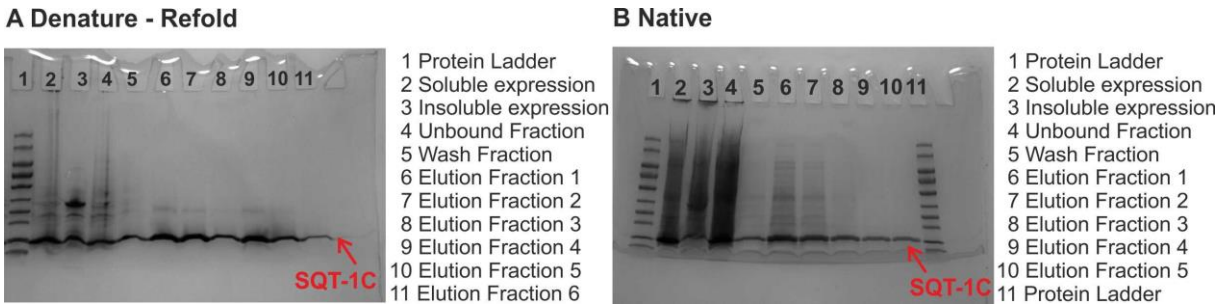

Figure S2.21: Analysis of Ni^{2+}-affinity chromatography efficiency for A) the Denature-Refold and B) the Native purification of SQT-1C

In both protocols, a significant fraction of SQT-1C did not bind to the column, which indicated that either incubation time should be longer or the capacity of the column was exceeded. The majority of bound SQT-1C eluted from the column in 6 ml of elution buffer. In native purification, elution fractions containing SQT-1C were joined, centrifuged to remove any aggregates and injected onto the SEC column. In denature-refold purification, elution fractions containing SQT-1C were first refolded by rapid dilution with refolding buffer followed by overnight dialysis in refolding buffer before injecting onto the SEC column. It is important to note that no visible aggregation was observed in neither of the tested purification protocols. SEC traces from both purifications are shown in Figure S2.17. Monomeric SQT-1C eluted from the column at 220 ml while SQT-1C dimer and tetramer eluted at 180 ml and 160 ml, respectively. In the denature-refold purification protocol, approximately half of SQT-1C was eluted as a monomer, with the other half forming dimers and tetramers. The final yield of monomeric SQT-1C in denature-refold purification was 50 mg per 1L of M9 minimum media. In the native purification protocol majority of SQT-1C eluted as a tetramer with barely any monomers present, which was a clear indication that denature-refold protocol should be used for purification of SQT-1C.

2.8.6 Assessing folding and stability of SQT-1C

The accurate size of SQT-1C oligomers present in solution was determined by SEC-MALS (Dawn Heleos II, Wyatt). The peak eluting at 14 ml had a molecular mass of 15.39 kDa, which corresponds to monomeric SQT-1C. Molecular mass of peaks eluting at 11 and 12 ml was determined to be 59.54 and 32.49 kDa, respectively, which are in agreement with dimeric and tetrameric species of SQT-1C. To assess whether different forms of SQT-1C are in equilibrium with one another, fractions containing monomers and dimers were re-run

on SEC-MALS immediately and 7 days after concentration. SQT-1C remained in its monomeric form after 1 h incubation, but the equilibrium between mono-, di- and tetramers was established after 7 days. Ratio between species was the same as in the initial sample run. In the rerun of dimeric SQT-1C fraction after an hour, peak characteristic for tetramers was already present, indicating that dimer and tetramers are in fast equilibrium. Transition of dimer into higher-ordered species was further observed with NMR spectroscopy. ^{15}N HSQC of dimeric SQT-1C was recorded immediately and 48 h after purification. In this time, almost all dimers associated/aggregated into higher-order species as indicated by loss of signal intensity and line broadening (Figure S2.22). The stability of monomeric SQT-1C was assessed by regularly recording ^{15}N HSQC spectra over two weeks. In that time, approximately 50% of SQT-1C was lost due to aggregation. (Figure S2.22).

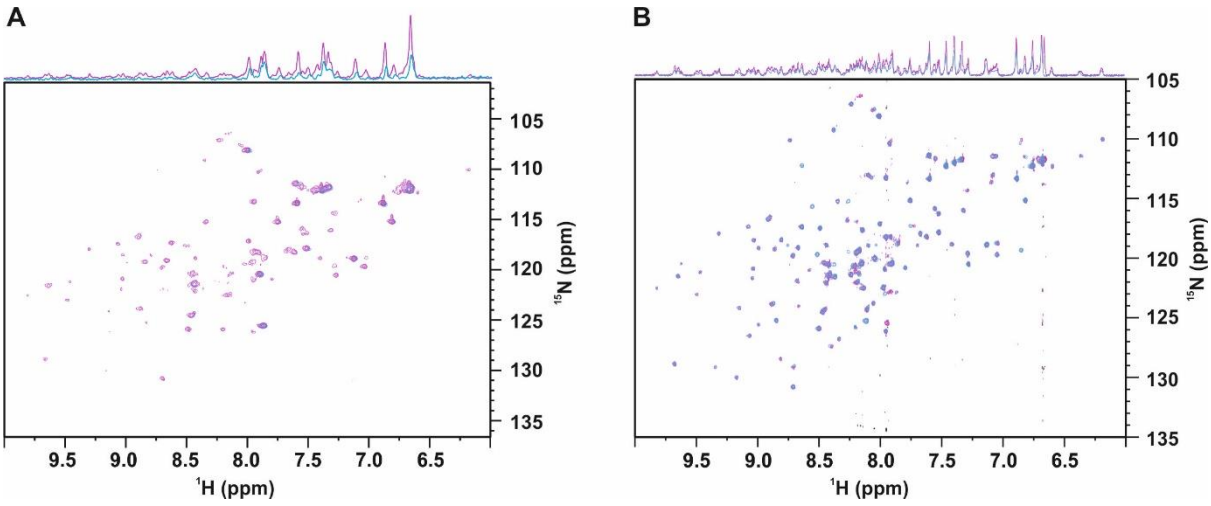

Figure S2.22: A) SQT-1C dimer immediately (pink) and 2 days (blue) after purification B) monomeric SQT-1C immediately (pink) and 2 weeks (blue) after purification

To determine optimum buffer conditions for SQT-1C, we tested the effect of NaCl and pH on SQT-1C folding. First, NMR titration of desalted SQT-1C sample with the gradual addition of NaCl to the final concentration 300 mM was performed (Figure S2.23). Initially, peaks in amide-aromatic region of ^{1}H NMR spectra sharpened and slightly increased in intensity with increasing NaCl concentration. The opposite effect was observed when the concentration of NaCl exceeded 100 mM. Therefore, it was concluded that NaCl should be used in the buffer at concentrations lower than 100 mM. Next, the effect of pH on SQT-1C was tested by gradual NMR titration from pH 7.5 to pH 6.5. Slightly narrower linewidths and a small increase of some signals in aromatic and amide region were observed

with decreasing pH. In ^{15}N HSQC three distinctive behaviours of peaks were noted during pH titration (Figure S2.23): Firstly, chemical shifts of the majority of peaks did not change and their intensities remained the same throughout the titration. Secondly, less than 10 peaks exhibited chemical shift perturbation on the fast exchange scale. Thirdly, peaks noted with dashed arrows appear to be in slow chemical exchange and their perturbations most probably occurred due to pH-induced local conformational changes. It was additionally observed that SQT-1C sample at pH 6.5 precipitated overnight. pH 7.2 was chosen for further work as a compromise between spectral quality and stability of the SQT-1C.

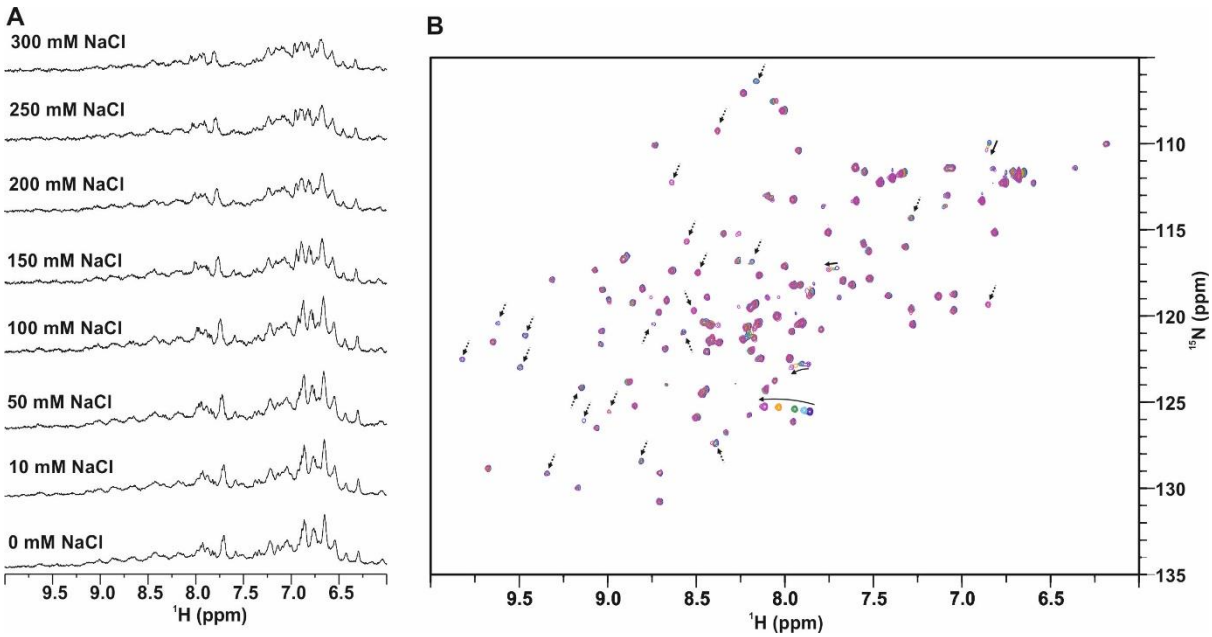

Figure S2.23: NMR assessment of SQT-1C stability. A) Amide and aromatic regions of ^{1}H NMR spectra of SQT-1C at different concentrations of NaCl. B: 15N HSQC spectra of SQT-1C at pHs 7.5 (dark blue), 7.2 (light blue), 7.0 (green), 6.8 (orange) and 6.5 (magenta). Solid and dashed arrows denote peaks in fast and slow exchange, respectively.

2.8.7 Materials and Methods

Transformation

All competent cells were transformed using a heat-shock protocol: 50 µl of competent cells were first incubated with 100 ng of appropriate plasmid on ice for 30 min, then placed into heat block at 42°C for 45 s and immediately transferred back to ice. After the addition of 0.5 ml SOC outgrowth media (NEB) cells were incubated at 37°C for an hour and then plated on LB agar plates with appropriate antibiotic (Table S2.6) and incubated at 37 °C overnight.

Table S2.6: Table of used competent cells and antibiotics

E. *coli* Competent cells	**Antibiotics on LB Agar plates**	**Antibiotics used in liquid media**
DH5α (Invitrogen)	Amp 100 µg/ml	Amp 50 µg/ml
BL21 (DE3) (TransGen Biotech co. Ltd.)	Amp 100 µg/ml	Amp 50 µg/ml
BL21 T7 Express (NEB)	Amp 100 µg/ml	Amp 50 µg/ml
Rosetta™ 2 (DE3) (Merck Millipore)	Amp 100 µg/ml, CAM 34 µg/ml	Amp 50 µg/ml, CAM 34 µg/ml
BL21 CodonPlus RP (Agilent technologies)	Amp 100 µg/ml, CAM 34 µg/ml	Amp 50 µg/ml, CAM 34 µg/ml

Protein expression optimisation

Protein expression optimisation was done in M9 minimal media. Transformed cells were grown in 50 ml of M9 minimal media with respective antibiotics (Table S2.6) to OD_{600} = 0.600 at 25, 30 or 37 °C, and protein expression induced with either 0.1, 0.5 or 1mM IPTG. 1ml samples were taken regularly pre and post induction to follow cell growth and analyse SQT-1 expression by SDS-PAGE. After overnight growth the cells were harvested by centrifugation at 6000 xg for 15 min and stored at -20 °C.

After the optimisation, SQT-1N was expressed in RosettaTM (DE3) competent cells (Novagen), grown at 37 °C to OD600 0.6, and induced with 1 mM IPTG. After overnight incubation at 30 °C, cells were harvested by centrifugation at 6000 g for 15 min and stored

at -20 °C. Following optimisation, the SQT-1C variant was expressed in BL21-CodonPlus RP+ competent cells (Agilent technologies), grown at 37°C to OD_{600} = 0.6, and induced with 0.5 mM IPTG. After overnight incubation at 37°C, cells were harvested by centrifugation at 6000 g for 15 min and stored at -20 °C.

Protein purification optimisation

Aliquot of cell pellet was resuspended in 20 ml protein of respective lysis buffer (see Table 3). Cells were disrupted by sonication with Sonopuls HD 3200 ultrasonic homogenizer (Bandelin) equipped with TT13/F2 probe (Bandelin) on ice for 15 min with 15 s on/ 45 s off setting at 30% amplitude. After centrifugation of cell lysate at 48500 g for 30 min, the supernatant was transferred onto Ni-NTA column (Quiagen), that was pre-washed with distilled water and equilibrated into appropriate lysis buffer (Table S2.7), and left to incubate at room temperature for 90 min. The column was then washed with 50 ml of corresponding washing buffer and bound SQT-1N or SQT-1C eluted into six 1 ml fractions with respective elution buffer (Table S2.7) In case of native and mild solubilisation protein purification 10 µl of each fraction was withdrawn for analysis by SDS-PAGE, while 50 µl was withdrawn in denature-refold protocol. Protein concentration in each fraction was determined using Bradford reagent. In native and mild solubilisation protocol fractions containing SQT-1 were joined, centrifuged for 10 min at 14000 g and injected on size exclusion Superdex-S75 26/600 GL chromatography column (GE Healthcare life sciences), pre-equilibrated with appropriate SEC buffer (Table S2.7). In the denature-refold purification, fractions containing SQT-1 were joined and refolded by rapid dilution with refolding buffer (Table S2.7), followed by overnight dialysis into refolding buffer. Next day, the refolded protein was concentrated using Vivaspin concentrators with cut-off 3.5 kDa to the final volume of 6 ml. The sample was then centrifuged at 14000 g for 10 min to remove any aggregates and injected on size exclusion Superdex-S75 26/600 GL chromatography column (GE Healthcare Life Sciences), pre-equilibrated with refolding buffer. In all three purification protocols, fractions containing monomeric SQT-1 were collected, joined and concentrated using Vivaspin concentrators (cut-off 3.5 kDa) to the final concentration of approximately 10 mg/ml in the stock solution. 50 µl of protease inhibitors without EDTA (Roche) were added per each ml of stock solution.

Table S2.7: Buffers used in protein purification

Purification type / Buffer	Native	Mild solubilisation	Denature-Refold
Lysis	50 mM Na-phosphate buffer, 0.3 M NaCl, pH 7.4	20 mM Tris-HCl, 0.3 M NaCl, pH 8.0	20 mM Na-phosphate buffer, 0.5 M NaCl, 6 M GuHCl pH 8.0
Washing	50 mM Na-phosphate buffer, 0.3 M NaCl, pH 7.4, 10 mM imidazole	20 mM Tris-HCl, 0.3 M NaCl, pH 8.0, 10 mM imidazole	20 mM Na-phosphate buffer, 0.5 M NaCl, 6 M GuHCl pH 8.0, 10 mM imidazole
Elution	50 mM Na-phosphate buffer, 0.3 M NaCl, pH 7.4, 500 mM imidazole	20 mM Tris-HCl, 0.3 M NaCl, pH 8.0, 500 mM imidazole	20 mM Na-phosphate buffer, 0.5 M NaCl, 6 M GuHCl pH 8.0, 500 mM imidazole
Refolding	X	X	20 mM Na-phosphate buffer, 150mM NaCl, 50mM ArgGlu, 5 mM EDTA, pH 7.2
SEC	50 mM Na-phosphate buffer, 0.3 M NaCl, pH 7.4	20 mM Tris-HCl, 0.3 M NaCl, pH 8.0	20 mM Na-phosphate buffer, 150mM NaCl, 50mM ArgGlu, 5 mM EDTA, pH 7.2

SDS-PAGE

SDS-PAGE was used to check the expression of SQT-1N and SQT-1C before and after induction of expression with IPTG. The 200 µl sample of the cell culture with the lowest OD600 was taken, while the volumes of other cell cultures were adjusted so that the amount of cells for analysis on the SDS-PAGE was the same across all samples. Cell culture samples were centrifuged at 14000 xg for 5 min, the supernatant was decanted, and the pellet was resuspended in 25 µl of MQ water. 25 µl of 2x SDS loading buffer (250 mM Tris-HCl, pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 0.002% (w/v) bromophenol blue) was added before heating at 95 °C for 5 min. Purity of SQT-1N and SQT-1C samples after each purification step was also checked with SDS-PAGE analysis. During native purification, 10 µl of tested

protein fraction was retained and 10 µl of 2x SDS-loading buffer added before heating at 95°C for 5 min and loading on the gel. In the denaturing purification, 50 µl of protein samples were retained and precipitated by addition of 450 µl chilled 100% ethanol and thorough mixing. Samples were then centrifuged at 14000 xg for 10 min and the supernatant decanted. Pellet was resuspended in 25 µl of 1x SDS loading buffer before heating at 95 °C for 5 min. In all cases, 10 µl of each sample was run on 4-20% Mini-PROTEAN® precast polyacrylamide gels (Bio-Rad) together with pre-stained protein Ladder (Life technologies) in SDS-Tris running buffer at 250 V for 25 min. Protein bands were visualized using Instant Blue Coomassie solution.

NMR experiments

If not stated otherwise, 450 µl of SQT-1N/C stock solution was first dialysed into the buffer of choice using GeBaflex dialysis tube (GeBA) overnight, and then 50 µl of 2H2O and 2.5 µl of DSA standard (100 mM stock solution) were added. All NMR spectra were recorded on the 800 MHz Bruker Avance III spectrometer equipped with TCl cryoprobe a temperature control unit at 25 °C. ^{1}H spectra were acquired using standard zgepsgp Bruker pulse sequences with excitation sculpting water suppression with 64 scans. ^{1}H-^{15}N HSQC spectra were recorded using hsqcfpfwg pulse sequence from the standard Bruker library with 8 scans and 128 increments.

BLANK PAGE

3 New disulphide bond in cystatin-based protein scaffold prevents domain-swap-mediated oligomerisation and stabilises the functionally active form

This chapter was written, peer-reviewed and published as:

Matja Zalar[1] and Alexander P. Golovanov[1,*], New disulphide bond in cystatin-based protein scaffold prevents domain-swap-mediated oligomerisation and stabilises the functionally active form, ACS Omega, 2019, 4 (19), 18248-56.

[1] Manchester Institute of Biotechnology and Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester, M1 7DN, UK

* Correspondence to: A.Golovanov@manchester.ac.uk

Author contributions:

M.Z. prepared all protein samples, performed the experiments, analysed the data and wrote the first draft of the manuscript. M.Z. and A.P.G. contributed to planning the experiments. A.P.G. provided conceptual guidance and made significant contributions to the final draft of the manuscript.

3.1 Abstract

Peptide aptamers built using engineered scaffolds are a valuable alternative to monoclonal antibodies in many research applications due to their smaller size, versatility, specificity for chosen targets, and ease of production. However, inserting peptides needed for target binding may affect aptamer structure, in turn compromising its activity. We have shown previously that a stefin A based protein scaffold with AU1 and Myc peptide insertions (SQT-1C) spontaneously forms dimers and tetramers, and that inserted loops mediate this process. In the present study, we show that SQT-1C forms tetramers by self-association of dimers, and determine the kinetics of monomer-dimer and dimer-tetramer transitions. Using site-directed mutagenesis, we show that while slow domain swapping defines the rate of dimerisation, conserved proline P80 is involved in the tetramerisation process. We also demonstrate that the addition of a disulphide bond at the base of the engineered loop prevents domain swapping and dimer formation, also preventing subsequent tetramerisation. Formation of SQT-1C oligomers compromises the presentation of inserted peptides for target molecule binding, diminishing aptamer activity, however introduction of the disulphide bond locking the monomeric state enables maximum specific aptamer activity, whilst also increasing its thermal and colloidal stability. We conclude that stabilising scaffold proteins by adding disulphide bonds at peptide insertion sites might be a useful approach in preventing binding-epitope-driven oligomerisation, enabling the creation of very stable aptamers with maximum binding activity.

3.2 Introduction

Peptide aptamers are proteins that consist of short target-binding polypeptide loops embedded within a stable protein scaffold, designed to bind specifically to a defined target. Engineered protein scaffolds are typically based on small native globular proteins, modified to remove original function and include new subcloning sites for adding the interchangeable loops. To achieve desired specificity and affinity, the sequences containing the desired binding epitope(s) (typically up to 10 – 15 residues) are usually inserted instead of the original loops. In principle, peptide aptamers mimic the antibody-based molecular recognition, but typically have a much smaller frame (often ~15 kDa), less complex structure and do not require post-translational modifications, and therefore can be often produced in simpler recombinant expression systems[266]. Peptide aptamers are applied in various research tasks, including the development of combinatorial protein libraries for protein recognition[37,

[270], studies of protein function and their interactions[330], diagnostic tools[274], biosensors[43], imaging agents[276] and as biotherapeutics[277]. As such, peptide aptamers are an emerging valuable alternative to monoclonal antibodies which until now have prevailed as the 'gold standard' for affinity binding studies.

More than 50 structurally diverse non-immunoglobulin scaffolds have been reported to date[266]. While protein scaffolds are designed to be as stable as possible, insertion of modified loops may however unintentionally destabilise them, leading to aggregation and reduction in thermal stability[51], or cause larger structural rearrangements such as domain-swap oligomerisation[331]. Changes to protein tertiary and quaternary structures may influence conformation or presentation of the binding loops, thus compromising target binding.

To explore in detail the structural and functional consequences of loop insertions, we are using a model engineered protein scaffold derived from stefin A, named SQT[56]. Stefin A belongs to the cystatin superfamily of cysteine protease inhibitors, which also includes stefin B and cystatin C[332]. SQT has three possible insertion sites for peptides, namely the N-terminus, loop 1 and loop 2. While it has been shown in the original publication[56] that SQT retains secondary structure upon various peptide insertions, we have demonstrated in our previous study that an SQT variant, named SQT-1C, with AU1 and Myc peptides inserted into loop 1 and loop 2, respectively, has decreased thermal stability and poor solution behaviour[331]. Insertion of these epitopes led to the spontaneous formation of interconverting monomeric, dimeric and tetrameric species in solution, with such oligomerisation directly mediated by the inserts in the engineered loops[331]. Although the problem with domain-swap oligomerisation and destabilisation has been identified, it was not clear what the functional consequences of this oligomerisation were, and how this structural instability could be prevented.

In this study, we have further explored the kinetics and mechanism of SQT-1C oligomerisation. We determined that tetramerisation occurs through self-association of domain-swapped dimers, with the formation of these dimers being the rate limiting step. We have designed two SQT-1C variants. In the first variant, a P80G point mutation was introduced to explore the role of conserved proline 80 in tetramerisation kinetics. For the second variant, a double mutant was designed, creating a disulphide bond which locked the configuration of the inserted loop 1. This drastically stabilised the monomeric species and prevented the formation of domain-swapped dimers. Additionally, we show that

oligomerisation of SQT-1C reduces its target binding capacity, whereas the disulphide bond-stabilised monomer had the highest specific activity. We conclude that stabilising protein scaffolds by adding disulphide bonds at peptide insertion sites to stabilise the engineered loops might be a useful approach for preventing binding-epitope-driven oligomerisation, while simultaneously also improving their thermal and colloidal stability.

3.3 Results

3.3.1 SQT-1C oligomerises through monomer-dimer-tetramer pathway

As previously shown[331] monomeric SQT-1C is in equilibrium with dimeric and tetrameric species in solution; however, the exact oligomerisation pathway has not been established. To determine the kinetic model of SQT-1C oligomerisation, we have isolated monomeric, dimeric and tetrameric protein fractions (Table 3.1), and followed the re-equilibration kinetics of each fraction using size exclusion chromatography (SEC).

Table 3.1: Molecular weights of protein oligomers (in kDa) as determined by SEC-MALS

	SQT-1C	**SQT-1C^{P80G}**	**SQT-1C$^{Q46C, N59C}$**
monomer	15.3±0.5	15.3±0.7	14.3±0.5
dimer	32±1	32±1	30±1
tetramer	62±2	59±2	61±2

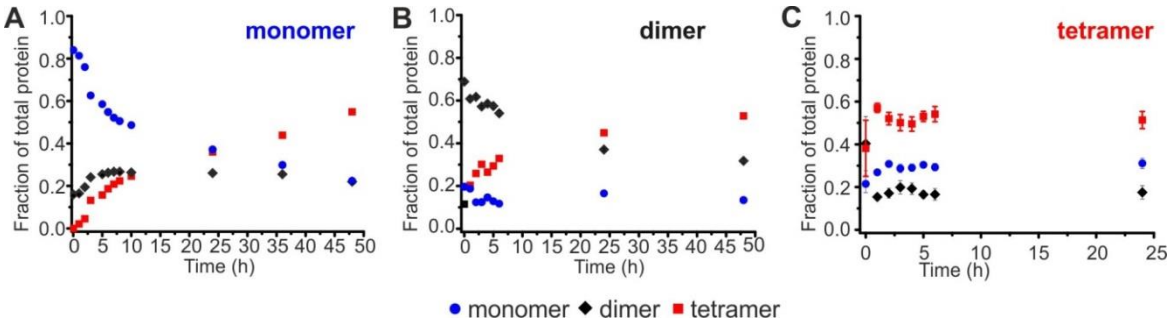

Figure 3.1: SEC analysis of SQT-1C species interconversion. Time evolution of monomers, dimers, and tetramers after incubation of 5 mg/mL SQT-1C samples at 25 °C starting from A) monomeric, B) dimeric and C) tetrameric pre-isolated species.

As shown in Figure 3.1A, monomeric SQT-1C first forms dimers, with tetramerisation occurring only after a substantial amount of dimers have accumulated in solution. This indicates that dimers act as an intermediate state on the oligomerisation pathway to the formation of tetramers. In the isolated dimer fraction (Figure 3.1B), the dimer population quickly converts to monomers and tetramers. Over time, the fraction of monomers remains constant, while the association of dimers into tetramers becomes predominant and decrease in dimer concentration together with the increase of tetramer population is observed. This further supports the observation that tetramerisation occurs by association of dimers and that dimers are only an intermediate state in the oligomerisation pathway. Finally in the tetrameric fraction, partial dissociation of tetramers into dimers and monomers occurs already during sample preparation, with the final equilibrium of predominantly tetramer population followed by monomer and then dimer reached after 2 hours of incubation (Figure 3.1C). This data clearly demonstrates that SQT-1C is in dynamic equilibrium between monomeric, dimeric and tetrameric species. Moreover, less than 1% of species with a molecular weight higher than 60 kDa were present throughout the SEC experiments, indicating that tetramers are the preferred final state with no further higher-order aggregation occurring in the time frame of the experiments.

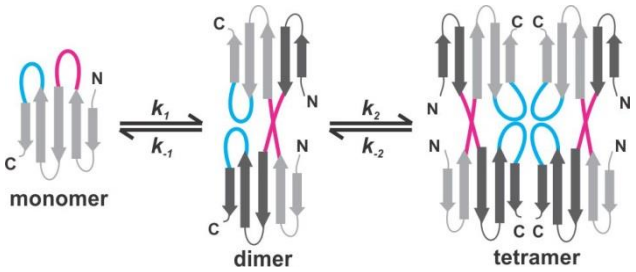

Figure 3.2: Model of SQT-1C oligomerisation. k_1 and k_{-1} are the on- and off rates of dimer formation while k_2 and k_{-2} are the on- and off rates for dimer-tetramer transition.

SQT-1C oligomerisation can be described by a sequential monomer-dimer-tetramer self-association mode[333] shown in Figure 3.2. All experimental SEC data on SQT-1C oligomerisation kinetics at various protein concentrations was successful fit to this model using DynaFit 4 software[334], obtaining a single set of global kinetic parameters. As shown in Figure 3.3, the monomer-dimer-tetramer model describes the data completely, further supporting the choice of the model.

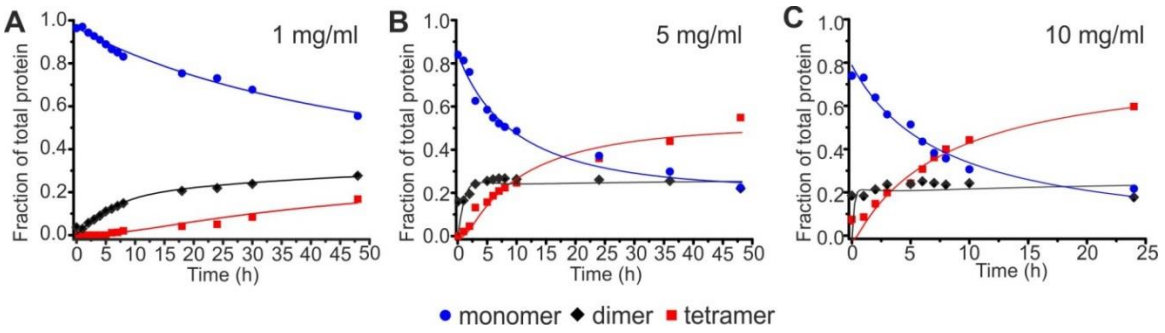

Figure 3.3: SEC analysis of SQT-1C oligomerisation kinetics starting from the monomeric form. Time evolution of monomer (blue circle), dimer (black diamond), and tetramer fractions (red square), after incubation of monomeric SQT-1C samples at 25 °C at different concentrations: A) 1 mg/mL, B) 5 mg/mL, and C) 10 mg/mL. Solid lines are global fits to a monomer-dimer-tetramer oligomerisation model.

3.3.2 Structural rearrangement of SQT-1C monomers is the rate limiting step of protein dimerisation

Table 3.2: Summary of estimated rate constants for SQT-1C oligomerisation at different temperatures. Reported errors are standard deviations obtained during data fitting in DynaFit 4 using the default settings.

SQT-1C				
T, °C	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1}) $*10^{-6}$	k_2 ($M^{-1}s^{-1}$)	k_{-2} (s^{-1}) $*10^{-6}$
20	0.018±0.003	0.1±5	0.52±0.2	9±5
22	0.02±0.005	0.1±5	8.7±7.1	0.1±5
25	0.058±0.008	0.1±0.1	1.7±0.6	2±2
27	0.096±0.01	0.1±7	1.8±0.5	3±2
30	0.47±0.12	210±40	6.4±3.7	2±0.1
33	0.53±0.34	500±2000	7.7±3.9	6±2
35	1.25±0.27	10±100	110±990	5±8

In order to further explore the kinetics of SQT-1C oligomerisation, we measured the temperature-dependence of rate constants for SQT-1C dimerisation and tetramerisation,

which were then used to estimate the apparent activation energy of these processes. Fits to experimental data are shown in Figure S3.1, while the obtained estimates of rate constants are shown in Table 3.2. It should be noted that the back rates k_{-1} and k_{-2} are poorly defined due to the lack of experimental data on the tetramer and dimer dissociation. However, the on rates are well defined as the obtained values are consistent with temperature without imposing any restraints during fitting. The on-rates for both dimerisation k_1 and tetramerisation k_2 are significantly slower than expected for a simple diffusion-limited self-association process where the rate constants typically range from 10^5 to 10^6 $M^{-1}s^{-1}$ [335]. This is consistent with structural rearrangements occurring on a slow timescale, responsible for both association steps. Furthermore, the rate for monomer to dimer (k_1) reaction is an order of magnitude slower than that of dimer to tetramer transition (k_2) indicating that the two processes are governed by different structural rearrangement mechanisms, and that dimer formation is the limiting step in SQT-1C oligomerisation pathway. Our previous molecular modelling data suggested that dimerisation proceeds via domain-swapping mechanism[331].

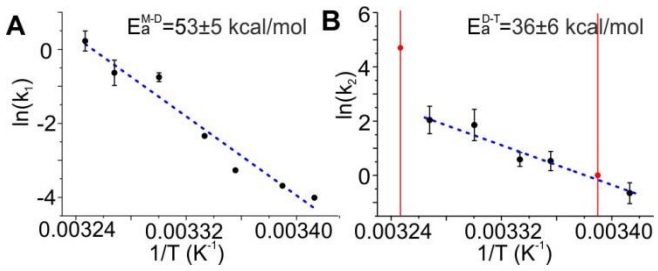

Figure 3.4: Arrhenius plot of the temperature dependence of SQT-1C A) monomer-dimer (k_1) and B)) dimer-tetramer (k_2) transition rates. Blue dashed line represents fit to Arrhenius equation. Data points coloured in red were discarded from line fitting due to excessively large errors.

Figure 3.4 shows Arrhenius plots of the temperature dependence data of SQT-1C dimerisation and tetramerisation rates k_1 and k_2. Fitting the data to the linearized Arrhenius equation allowed estimation of the apparent activation energies (E_a) for both processes. E_a for SQT-1C monomer to dimer transition, 53±5 kcal/mol, was greater than that for dimer-tetramer transition, 36±6 kcal/mol. Activation energies for SQT-1C dimerisation are consistent with previously reported values of 55±4 kcal/mol for domain swap oligomerisation of stefin B[336]. Similarly, the apparent activation energy for SQT-1C tetramerisation is similar to that of stefin B tetramerisation, 28±3 kcal/mol[287], and is consistent with slow reactions accompanied by minor, local conformational changes.

3.3.3 Rationale for SQT-1C mutant design

For the cystatin family, two distinct steps of protein association have been reported previously, both involving structural rearrangement and hence relatively slow timescale. The first one is domain swap dimerisation, where the domain swap occurs through the extension of the conserved hydrophobic five-residue 'cystatin motif' (QVVAG) in loop 1[286, 296]. In SQT, this motif has been mutated (to QVLAS) and split to accommodate peptide loop insertion into the scaffold[51, 56]; therefore this motif itself can no longer be responsible for the domain swapping (Figure 3.5). For SQT-1C our previous experiments and modelling suggest that it is the engineered loops themselves that drive domain-swap mediated dimerisation and further tetramerisation[331]. The second known step of cystatin association is a so-called hand-shaking mechanism, where the *trans to cis* isomerisation of conserved P74 in loop 2 is required for association of two domain-swapped dimers into a stable tetramer[287].

To confirm the mechanism of SQT-1C oligomerisation, and to find way of preventing it, we have created two mutants. In the first mutant, named SQT-1C^{P80G}, we have mutated the residue P80 (corresponding to P74 conserved in other cystatins) to glycine (Figure 3.5), removing the possibility of slow *trans-cis* isomerisation while allowing more conformational flexibility. Such a change could potentially either eliminate tetramerisation, as shown previously for stefin B [287] and Na^+-K^+-ATPase[337], or accelerate it due to increased flexibility of the loop 2. In the second mutant (SQT-1C^{Q46C,N59C}) we have introduced a disulphide bond across the base of loop 1, between β1 and β2 strands, by a double Q46C and N59C mutation (Figure 3.5), in an attempt to stabilise a specific topology and prevent structural rearrangement. A similar approach has been used previously on cystatin C, where prevention of domain swapping using disulphide bridges inhibited dimerisation and fibril formation[338]. The position of cysteines in SQT-1C^{Q46C,N59C} was chosen for two reasons. Firstly, their position at the end of the β2 and start of β3 should prevent opening of loop 1 in the monomer, or covalently trap the domain-swapped dimer, preventing the interconversion of monomers and dimers, allowing their purification. Secondly, these mutations are positioned outside the restriction site *Nhe*I in loop 1[56] and hence should not affect the insertion of the target-binding peptides into the SQT scaffold between L48 and A56 residues situated at the base of the loop (Figure 3.5B, D).

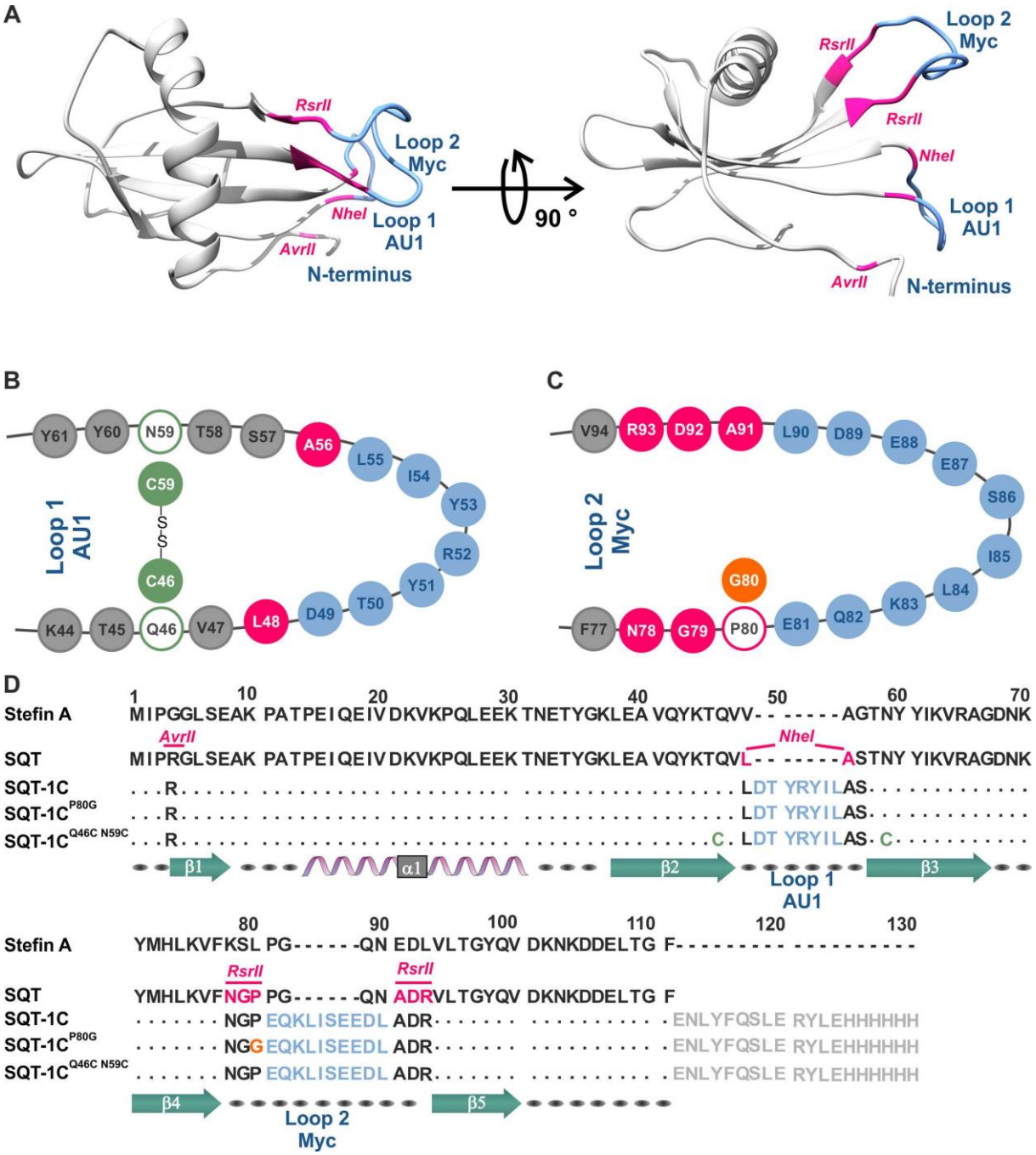

Figure 3.5: SQT-1C mutation scheme. A) Model of SQT-1C based on PDB ID 6QB2 B) Scheme of loop 1 and mutations present in SQT-1C^{Q46C,N59C} mutant. C) Scheme of loop 2 and mutation introduced in SQT-1C^{P80G} variant. D) Sequence alignment of stefin A, SQT, SQT-1C, SQT-1C^{P80G} and SQT-1C^{Q46C,N59C}. Secondary structure of SQT-1C, as determined from its structure (PDB ID 6QB2) is shown below sequences. In all panels, positions and names of restriction sites used to add functional loops to SQT scaffold are shown in pink, inserted loops are shown in blue, and unmodified regions in grey. Mutation site P80G is depicted in orange while Q46C and N59C are shown in green.

3.3.4 Proline 80 is involved in SQT-1C tetramerisation

After separation of refolded SQT-1C^{P80G} variant on SEC coupled with multi angle light scattering (SEC-MALS) three elution peaks were identified corresponding to monomer, dimer and tetramer fractions (Table 3.1 and Figure S3.2A,B). Far-UV CD spectra of freshly isolated monomeric SQT-1C^{P80G} species showed little difference to that of isolated SQT-1C, indicating that the secondary structure of SQT-1C is retained in the SQT-1C^{P80G} mutant (Figure S3.2C). Additionally, only minor chemical shift perturbations of residues located next to the mutated residues were identified in 2D ^{1}H-^{15}N HSQC spectra, further indicating that the 3D structure of SQT-1C monomer is retained in the P80G mutant (Figure 3.6A and Figure S3.2D).

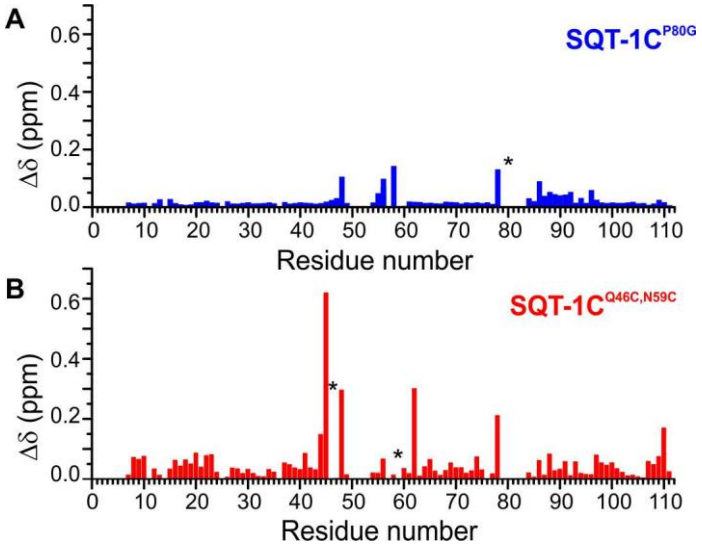

Figure 3.6: NMR chemical shift perturbation analysis of mutant variants. Per residue weighted backbone amide chemical shift perturbations for A) SQT-1C^{P80G} and B) SQT-1C^{Q46C,N59C} compared to SQT-1C show that chemical shift perturbations occur only around the mutation sites and at residues in spatial proximity of the mutated sites. Asterisks (*) denote mutation sites.

Melting temperature of SQT-1C^{P80G} was 54±1°C, compared to 56±1°C for SQT-1C, as measured by the intrinsic fluorescence peak shift upon heating, showing that the P80G mutation does not significantly affect thermal stability. The colloidal stabilities of SQT-1C and SQT-1C^{P80G}, measured by static light scattering at 266 nm as onset temperature of aggregation (T_{agg}), were also very similar, and coincided with their melting temperatures, suggesting major aggregation happening once the protein becomes thermally unfolded (Figure 3.7).

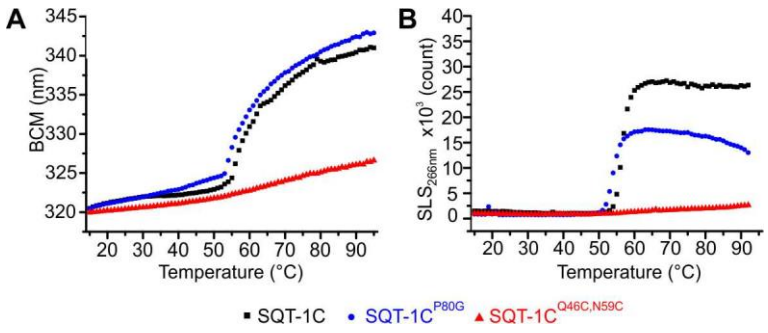

Figure 3.7: Comparison of thermal and colloidal stability of SQT-1C and mutants. A) Temperature dependence of barycentric mean (BCM) of the intrinsic fluorescence signal. B) SLS_{266nm} across the temperature ramp. In both panels, SQT-1C, SQT-1C^{P80G} and SQT-1C^{Q46C,N59C} are depicted in black, blue, and red, respectively.

To evaluate how P80G mutation affects SQT-1C oligomerisation kinetics, the temperature dependence of SQT-1C^{P80G} oligomerisation rate constants were analysed and used to estimate the activation energies of individual steps, similar to the analysis of SQT-1C kinetics. Fits to temperature-dependent data are shown in Figure S3.3, while estimated rate constants are summarised in Table 3.3. For SQT-1C^{P80G} both the dimerisation rates k_1 and their temperature dependence were similar to those of SQT-1C, as evident from Arrhenius plots (*cf* Figure 3.8A and Figure 3.4A), with similar apparent activation energy, indicating that the P80 is not involved in the dimerisation process. However, the estimated activation energy for dimer to tetramer transition was significantly lower for SQT-1C^{P80G} compared to SQT-1C (*cf* Figure 3.8B and Figure 3.4B). These experiments overall reveal that whereas P80 in SQT-1C is not involved in the dimerisation process, it is however involved in tetramerisation, and the *trans-cis* isomerisation of this residue is likely a contributing factor.

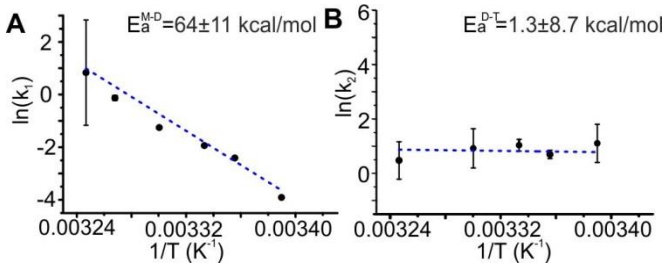

Figure 3.8: Arrhenius plot of the temperature dependence of SQT-1C^{P80G} A) monomer-dimer k_1 and B) dimer-tetramer k_2 transition rates. Blue dashed line represents fit to Arrhenius equation.

Table 3.3: Summary of estimated rate constants for SQT-1C^{P80G} oligomerisation at different temperatures. Reported errors are standard deviations obtained during data fitting in DynaFit 4 using the default settings.

SQT-1C^{P80G}				
T, °C	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1}) $*10^{-6}$	k_2 ($M^{-1}s^{-1}$)	k_{-2} (s^{-1}) $*10^{-6}$
22	0.02±0.01	0.1±10	3±2.1	6±4
25	0.09±0.01	3±3	2.1±0.3	7±1
27	0.14±0.01	5±3	2.8±0.6	10±20
30	0.29±0.02	0.1±0.2	3±2	10±3
33	0.88±0.03	0.2±3	$7*10^6$±$3*10^8$	14±3
35	2.3±2	15±13	1.6±1.1	2±3

3.3.5 Monomeric state of SQT-1C can be stabilised by addition of a disulphide bond

To stabilise the monomeric form of SQT-1C and prevent structural rearrangement leading to dimerisation, a double Q46C and N59C mutant was produced so that a disulphide bond can spontaneously form at the base of loop 1 between β2 and β3 strands. Additionally, any domain-swapped dimers formed during refolding and oxidation will be also covalently stabilised, preventing their dissociation into monomers. The SQT-1C^{Q46C,N59C} mutant was expressed, refolded, oxidised and purified as described in section 3.5.2. After separation on size exclusion column, individual monomeric, dimeric and tetrameric species of SQT-1C^{Q46C,N59C} were isolated for further analysis (Table 3.1 and Figure S3.4**).** Notably, large populations of higher molecular weight oligomers were visible on SEC-MALS trace (Figure S3.4A,B), likely formed by misfolding and crosslinking via disulphide bonds during the refolding/oxidation step. Far-UV CD spectra of freshly isolated SQT-1C^{Q46C,N59C} monomer species showed little difference to SQT-1C monomer, indicating that the secondary structure of SQT-1C is retained (Figure S3.4C). Additionally, only minor chemical shift perturbations of residues located next to the mutated residues were identified in 2D ^{1}H-^{15}N HSQC spectra, further indicating that the 3D structure of SQT-1C monomer is retained in SQT-1C^{Q46C,N59C} (Figure 3.6B and Figure S3.4D). Moreover, the CD spectra show that isolated covalently cross-linked SQT-1C^{Q46C,N59C} dimers and tetramers are structurally similar to dimers and tetramers formed by SQT-1C (Figure S3.5). In addition to structure retention, introduction of disulphide bond drastically increased the melting temperature of the monomeric species

above 95 °C, with only minor changes in intrinsic fluorescence signal observed across temperature ramps (Figure 3.7A). Additionally, only a slight increase in SLS at 266 nm over increasing temperature was observed, much less than for SQT-1C or SQT-1C^{P80G}, further indicating that this bridged mutant is colloidally stable up to very high temperatures (Figure 3.7B). Hydrogen-deuterium (H-D) exchange rates for SQT-1C^{Q46C,N59C} were lower than those for SQT-1C measured previously[331] (see Figure S2.6), but H-D exchange still occurred within minutes (Figure S3.6), suggesting that even after addition of disulphide bond the monomeric structure somewhat lacks long-lived hydrogen bond networks.

To test whether introduction of the disulphide bond between β2 and β3 sheets stabilised the monomeric species against transition into dimers and tetramers, we tested SQT-1C^{Q46C,N59C} oligomerisation kinetics across a range of temperatures and protein concentrations. While a small fraction (<5%) of dimers was present in initially isolated monomeric species due to lack of column resolution and slight overlap between the elution peaks of monomers and dimers in SEC, there was no further significant interconversion of monomeric species into dimers or higher oligomers observed over time in any of the tested conditions, even at higher concentrations (Figure S3.7). We can therefore conclude that engineered disulphide bond prevents the opening of the monomeric species through loop 1, hence preventing the domain swapping and dimerisation, and subsequent tetramerisation. Consequently, SQT-1C^{Q46C,N59C} stays in solution as a stable monomer, with greatly enhanced colloidal and thermal stability.

3.3.6 Oligomeric state of SQT-1C influences its interaction with its binding partners

One of the main, but often implicit, assumptions in the engineered scaffold design is that the inserted target-binding loops are held by the scaffold in a correct conformation optimal for their binding, and that the scaffold is stable enough to maintain this conformation throughout its preparation, storage and usage lifecycle. In the case of SQT-1C dimerisation and tetramerisation clearly changes the conformation of binding epitope within loop 1, from hairpin to extended conformation, and also changes its solvent exposure[331]. Therefore, the question arises whether this oligomerisation alters the functional competency of SQT-1C. While isolated monomeric and oligomeric fractions of SQT-1C could not be tested previously for functionality due to fast interconversion between the species, SQT-1C^{Q46C,N59C} mutant yielded stable monomers, dimers and tetramers which can now be

separated. Hence, we tested the binding efficiency of monomeric, dimeric and tetrameric species of SQT-1C^{Q46C,N59C} using an ELISA experiment with commercially available polyclonal antibodies against AU1 and Myc peptides located in loops 1 and 2, respectively.

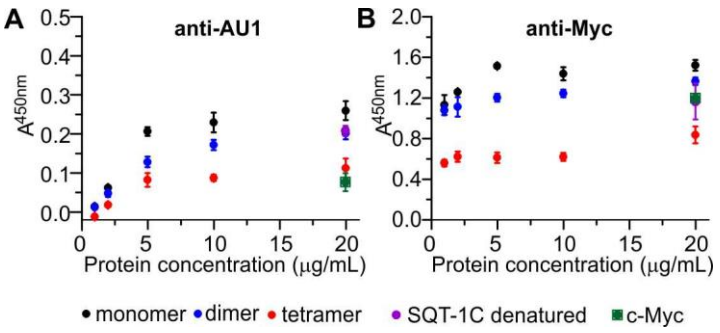

Figure 3.9: ELISA titers for SQT-1C^{Q46C,N59C} binding to A) anti-AU1 and B) anti-Myc antibodies. Monomeric, dimeric and tetrameric fractions (shown in black, blue, and red, respectively) were tested for binding efficiency, in comparison with denatured 20 μg/mL SQT-1C (violet) and 20 μg/mL Myc protein (green) used as controls. While addition of denatured SQT-1C was used as positive control for both experiments, c-Myc protein served as a negative control in AU1 binding and as positive control in Myc binding experiments.

The most efficient AU1-mediated binding (on total protein quantity basis) was observed for the monomeric protein, followed by dimer, and then tetramer, for a range of total protein concentrations (Figure 3.9A). Both dimers and tetramers have a smaller binding capacity than monomers, presumably due to a combination of several factors, namely steric clashes, partial burial and non-optimal extended conformation of loop 1 in domain-swapped oligomers. Binding efficiency of tetramer with AU1 antibody was not dissimilar to binding of this antibody to standalone Myc peptide, used as a negative control. Interestingly, efficiency of SQT-1C^{Q46C,N59C} binding to Myc peptide present in loop 2 is only slightly reduced by dimerisation, while tetramerisation significantly decreases SQT-1C^{Q46C,N59C} ability to present the Myc peptide to the respective antibody (Figure 3.9B), presumably due to burial of loop 2 within the tetramerisation interface. Control reactions performed with 20 μg/mL standalone Myc peptide (positive control) showed similar efficiency of binding to Myc antibodies as denatured SQT-1C, however SQT-1C^{Q46C,N59C} monomers exhibited even higher binding efficiency.

Binding of SQT-1C^{Q46C,N59C} monomers to both AU1 and Myc antibodies was consistently more efficient than that of denatured SQT-1C which was added to ELISA reactions as a control . It can be envisaged that upon addition and dilution, this control WT SQT-1C partially refolds and partitions into usual mixture of monomers, dimers and tetramers, so it is expected to have an appreciable binding affinity for both antibodies. These results clearly indicate that SQT-1C monomers possess the highest specific activity towards target binding, whereas for domain-swapped dimers, and particularly tetramers, the specific binding activity is significantly reduced. This finding provides a rationale for stabilising a specific form of an engineered protein scaffold (in the case of SQT-1C this is the monomeric form) to achieve maximum specific activity, as well as to improve its thermal and colloidal stability and prevent domain swapping. These several beneficial effects can be achieved simultaneously by introduction of a single disulphide bond at the base of target-binding loop which otherwise drives domain swapping and oligomerisation.

3.4 Discussion

The ability of engineered protein scaffolds to retain their structural, thermal and colloidal stability upon insertion of various peptide loops needed for their target-binding function is crucial for their research and industrial applications. As such, the small frame of scaffolds needs to absorb additional steric strains introduced by the inserted loops, and to maintain the correct conformation. We have shown that protein scaffold SQT-1C forms domain swap dimers that further associate into stable tetramers, in a similar way to oligomerisation pathway of other proteins in the cystatin family[287, 296, 336]. Slow kinetics and high apparent activation energy of SQT-1C dimerisation are consistent with large structural rearrangement needed for dimer formation. Interestingly, the apparent activation energy for SQT-1C dimerisation is roughly twice smaller than that reported for stefin A[286], and is similar to the activation energy in the nucleation phase of fibrillation reaction for stefin B, the less stable of the two stefins[339]. While monomer-dimer and dimer-tetramer transitions occur on a similar timescale to domain swap dimerisation and subsequent tetramerisation and fibrillation of other members of the cystatin family, SQT-1C oligomerises at room temperature, while oligomerisation of other members of the cystatin family, including stefin A[286], stefin B[336], and cystatin C[88, 285, 296] occurs only at elevated temperature, in the presence of organic solvents or in the presence of denaturants. This suggests that insertion of the

specific peptides in SQT-1C significantly lowers the stability of the protein, making it more likely to form domain swapped dimers.

Domain swap dimerisation in cystatin family has been previously reported to occur through extension of the conserved hydrophobic five-residue 'cystatin motif' (QVVAG) in Loop 1 as a consequence of frustration of this hairpin hinge region[340-341]. It has been shown recently that this motif, when engineered into the hinge of a β-hairpin, causes domain swapping of otherwise non-domain swapped proteins[90]. On the other hand, it has been established that mutations in this hinge region can slow down or even completely eliminate domain swap oligomerisation of cystatins[87-88, 296]. Even though V48L mutation has been introduced into this particular motif in SQT to introduce *NheI* restriction site for insertion of peptides between L48 and the following alanine into the scaffold and to prevent domain swap oligomerisation[51, 56], SQT-1C still forms domain swap dimers at room temperature[331]. As we have shown previously by measuring the temperature of unfolding T_m and studies of H-D exchange rates[331], SQT-1C structure somewhat lacks a stable hydrogen bonding network. It is likely that structural frustration introduced by AU1 peptide in the loop 1 destabilises the construct, with the frustration relieved by fully extending AU1 peptide conformation. This is achieved in the domain-swapped dimers, which associate further into tetramers. Introducing the disulphide bond stabilising loop 1 in a hairpin configuration greatly increases thermal and colloidal stability of SQT-1C^{Q46C,N59C}, and essentially prevents monomer-dimer transition, allowing isolation of a stable monomeric form, which also exhibits the highest specific activity towards binding antibodies for both loops 1 and 2.

In proteins from the cystatin family, tetramerisation occurs via a so called hand-shaking mechanism, where the *trans* to *cis* isomerisation of conserved P74 in loop 2 of these proteins drives association of two domain-swapped dimers into a stable tetramer[287]. In SQT-1C, the correspondent residue is P80, which was mutated here to a glycine to remove contribution from proline isomerisation, producing SQT-1C^{P80G} mutant. We show that P80G mutation significantly reduces the activation energy needed for tetramer formation. This suggests that *trans-cis* isomerisation of P80 in SQT-1C may be one of the transitions needed for loop 2 to adopt conformation favourable for tetramerisation and engage in interaction with the neighbouring chain. Overall the tetramers of SQT-1C^{P80G} mutant are structurally similar to those formed by SQT-1C. This is in contrast to previous observations in stefin B, where mutation of the conserved proline disrupted the typical pathway of oligomerisation leading to fibril formation. Instead, amorphous aggregates were formed without clear

monomer-dimer-tetramer transition[287]. This highlights the subtle differences between the engineered protein scaffold, and its native ancestors.

Engineering disulphide bonds into the protein core is generally a well-established method to increase protein stability[121]. A disulphide bond was successfully introduced previously in cystatin C to prevent its dimerisation and eliminate fibril formation[338]. Here, we have introduced a disulphide bond between β2 and β3 sheets of SQT by a double Q46C and N59C mutation in an attempt to stabilise the monomeric form and prevent interconversion of monomeric species into dimers and tetramers. However, the effect of this disulphide bond on SQT protein scaffold was quite dramatic, not only locking the structure in monomeric form and preventing oligomerisation, but also raising the melting temperature and onset temperature of aggregation above 95 °C. This increase surpasses the 79.9 °C melting temperature of the original "empty" SQT scaffold itself[56]. From our functional binding experiments, we found that this monomeric form had the highest specific binding activity, compared with dimers and tetramers, suggesting that monomers ensure the best presentation of the target binding epitopes. This further implies that in the case of SQT scaffold, its major degradation pathway, formation of soluble dimers and tetramers, is detrimental to its functional activity. As it can be anticipated that addition of target-binding loops in other small engineered protein scaffolds may introduce similar strains on the core structure, leading to domain swapping and subsequent oligomerisation, we propose that adding disulphide bonds at the base of ligand binding loop(s) may increase scaffold stability and maximise its specific target-binding activity.

3.5 Materials and Methods

3.5.1 Plasmids

Synthesized codon-optimised gene constructs of SQT-1C and two mutants, SQT-1C^{P80G} and SQT-1C^{Q46CN59C} were obtained from GeneArt (ThermoFisher Life Technologies), and subcloned into pET21a+ vector with a cleavable hexa-histidine tag as previously described[331].

3.5.2 Protein expression and purification

All three SQT-1C variants were expressed as previously described[331]. While SQT-1C and SQT-1C^{P80G} were purified as previously reported[331], cell pellets of SQT-1C^{Q46C,N59C}

were resuspended in denaturing buffer (20 mM NaPi, 500 mM NaCl, 6M GndHCl, 5 mM TCEP, pH 8.0) with 0.5% v/v Triton X-100 (Sigma Aldrich). Resuspended pellets were then lysed by sonication with Sonopuls HD 3200 ultrasonic homogenizer equipped with TT13/F2 probe (Bandelin) and clarified by centrifugation at 30000 g for 30 min at 4 °C. Supernatants were transferred onto Ni-NTA resin (Quiagen) in a gravity flow column and incubated for 90 min at 25°C. After incubation, columns were washed with respective denaturing buffers supplemented with 10 mM imidazole. The bound material was eluted with 500 mM imidazole in denaturing buffer. Refolding and oxidation of SQT-1C^{Q46CN59C}, enabling disulphide bond reshuffling, was achieved by 1:10 v/v rapid dilution where 1 mM reduced GSH and 0.25 mM oxidized GSH were added to refolding buffer (20 mM NaPi, 150 mM NaCl, 5 mM EDTA, pH 7.2), followed by overnight dialysis into refolding buffer without GSH. Finally, the proteins were purified on Superdex 200 26/600 HiLoad column (GE Life Sciences), pre-equilibrated with refolding buffer. All SQT-1C variants eluted as a set of well-defined oligomers, allowing isolation of monomeric, dimeric and tetrameric fractions. Isolated species were then re-concentrated to the desired concentration using Vivaspin 20 centrifugal devices with a 5 kDa molecular weight cut-off (Sartorius Stedium Biotech GmbH). Protein concentrations were estimated by absorbance at 280 nm (ε = 14900 M^{-1} cm^{-1}). Molecular weights of protein species were determined using size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) run at 25 °C. 200 µg of protein samples were injected on Superdex 200 10/300GL column (GE Life Sciences) and passed through a Wyatt DAWN Heleos II EOS 18-angle laser photometer (Wyatt Technology) coupled to a Wyatt Optilab rEX (Wyatt Technology) refractive index detector. Data analysis was performed in ASTRA 6.1 software (Wyatt Technology).

3.5.3 CD spectroscopy

Far-UV CD spectra of individual oligomeric species were acquired on an Applied Photophysics Chirascan using a 0.01 cm path length quartz cell, immediately after separation in SEC column at protein concentration of 1 mg/mL. The wavelength was varied from 190 to 280 nm with 0.5 nm step and acquisition time of 3 s per point. For each CD spectrum, three scans were averaged and smoothed.

3.5.4 Static light scattering and intrinsic fluorescence

Static light scattering (SLS) and intrinsic fluorescence measurements were conducted simultaneously using an UNcle (Unchained Labs) across a temperature ramp from 20 °C to 90 °C with heating rate 1 °Cmin^{-1}. Data was processed using the UNcle analysis

software, as per manufacturer's recommendations. Melting temperatures (T_m) of all three SQT-1C variants were determined using temperature dependence of the first derivative of the barycentric mean (BCM) of fluorescence intensity, following standard instrument procedure. SLS at 266 nm was used as an indicator of protein colloidal stability, where the onset of aggregation temperature (T_{agg}) was defined as the temperature at which the measured scattering signal reaches 10% of its maximum value.

3.5.5 Monitoring SQT-1C oligomer transitions by SEC

Oligomerisation kinetics of all SQT-1C variants was measured using SEC, with a Superdex 200 10/300GL column (GE Life Sciences) attached to an Agilent 1100 Series HPLC system (Agilent Technologies) at 25 °C, and pre-equilibrated in refolding buffer. Elution of samples was detected at 280 nm. Isolated monomer fractions of SQT-1C fractions at 5 mg/mL were incubated for 24 hours at 20, 22, 25, 27, 30, 33, 35 and 40 °C with 10 µL sample aliquots injected onto the SEC column every 90 min. Kinetic data was obtained as a series of single independent runs, and kinetics at 25°C was measured twice to check for data reproducibility. For concentration dependent analysis of monomer-dimer-tetramer transition, isolated monomers were incubated at 25°C at 1, 5, and 10 mg/mL, and the kinetics of oligomerisation were analysed as described above. All peaks in SEC traces were integrated and finally the ratio between different oligomers was calculated. Data analysis was performed in ChemStation (Agilent Technologies) and OriginPro 9.1 (OriginLab). The experimental SEC data on oligomerisation kinetics at various concentrations and temperatures were then fit to the monomer-dimer-tetramer model using DynaFit 4 software[334]. Apparent activation energies for monomer-dimer and dimer-tetramer transitions were estimated by fitting the data to the linearized Arrhenius equation: $\ln(k) = \ln(A) - E_a/(RT)$, where k is rate constant for each oligomerisation step, A the pre-exponential factor, R a gas constant, T is absolute temperature and E_a the apparent activation energy.

3.5.6 NMR experiments

NMR samples were prepared by adding 5% v/v ^2H_2O to 1 mM ^{15}N- labelled protein solutions in 20 mM sodium phosphate buffer, 150 mM NaCl, pH 7.2. All NMR spectra were acquired at 25 °C on 800 MHz Bruker Avance III spectrometer equipped with 5mm triple resonance TCI cryoprobe and temperature control unit. The spectra were acquired and processed using Bruker Topspin 3.5 and analysed using NMRFAM-SPARKY[342] and Dynamics Center 2.2.4 (Bruker). Backbone assignment of SQT-1C has been previously described[331] (BMRB ID 27757). The backbone assignment of SQT-1C was then transferred

to SQT-1C^{P80G} and SQT-1CQ^{46C,N59C} 2D ^{1}H-^{15}N HSQC spectra by matching peak positions. Assignment of shifted cross-peaks of SQT-1CQ^{46C,N59C} was additionally verified using 3D TROSY-based HNCA and HNCO experiments from standard Bruker pulse program library. Proton-deuterium (H-D) exchange rates of SQT-1C^{Q46C,N59C} were measured as previously described[331]. The weighted chemical shift changes of backbone amide groups ($\Delta\delta_{NH}$) due to point mutations in SQT-1C^{P80G} and SQT-1C^{Q46C,N59C} were calculated as $\Delta\delta_{NH} = \sqrt{0.5[(\Delta\delta_H)^2 + (0.1 * \Delta\delta_N)^2]}$, where $\Delta\delta_H$ and $\Delta\delta_N$ were chemical shift changes in proton and nitrogen dimension, respectively.

3.5.7 ELISA

To examine the ability of SQT-1C^{Q46C,N59C} oligomers to present inserted peptides to target antibodies, plastic Maxisorb plates (Nunc) were coated with 1, 2, 5, 10 and 20 μg/mL of SQT-1C^{Q46C,N59C} monomers, dimers and tetramers in phosphate buffered saline pH 7.4 (PBS) overnight at 4 °C with shaking. c-Myc protein at 20 μg/mL was used as a positive control for anti-Myc antibody binding, and as negative control for anti-AU1 antibody binding, whereas 20 μg/mL denatured SQT-1C was used as another positive control for both antibodies. Additionally, bovine serum albumin (BSA) was used as a negative control. Protein concentrations were measured by UV absorbance at 280 nm wavelength. C-Myc protein sample was kindly provided by Dr Matthew Cliff and Prof Jon Waltho (University of Manchester). All samples were measured in triplicates. Plates were blocked with 2% (w/v) BSA (SigmaAldrich) in PBS at 25 °C for 2 h. Plates were then incubated with either goat anti-AU1 primary antibody (ab3400, Abcam) diluted 1:2000, or goat anti-Myc tag primary antibody (ab9132, Abcam) diluted 1:25000 for 2h at 25°C, followed by incubation at 25 °C for 1 hour with rabbit anti-goat secondary antibody labelled with horseradish peroxidase (ab6741, Abcam) diluted 1:50000 with PBS. Between incubation steps, plates were washed using 0.05% (v/v) Tween 20 in PBS. After incubation with secondary antibody, plates were incubated with the 3,3',5,5'-tetramethybenzidine (TMB) substrate (Abcam) for 15 min at 25°C. The reaction was stopped by addition of 450 nm Stop Solution for TMB Substrate (Abcam) and the absorbance was read at 450 nm using a multiwall plate reader CLARIOstar (BMG LABTECH). Data was processed and analysed using MARS (BMG LABTECH) and OriginPro 9.5.1 (OriginLab).

3.6 Acknowledgments

We would like to thank Dr Robin Curtis and Prof Dr Alan Dickson, and their groups for letting us use their research facilities. We are especially grateful to Prof Dr Jon Waltho for useful discussions and his comments on the work. We would also like to thank Dr Matthew Cliff from the MIB NMR Facility for technical support with NMR spectrometers, Dr Tom Jowitt from Biomolecular Analysis Facility for technical support with SEC-MALS, Dr Derren Heyes from the MIB Biophysics Facility for technical support with the CD spectrometer, and Dr Tim Eyes from SynBioChem for advice on ELISA experiments. This work was supported by EU Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement No 675074.

3.7 Supplementary information

3.7.1 Supplementary Figures

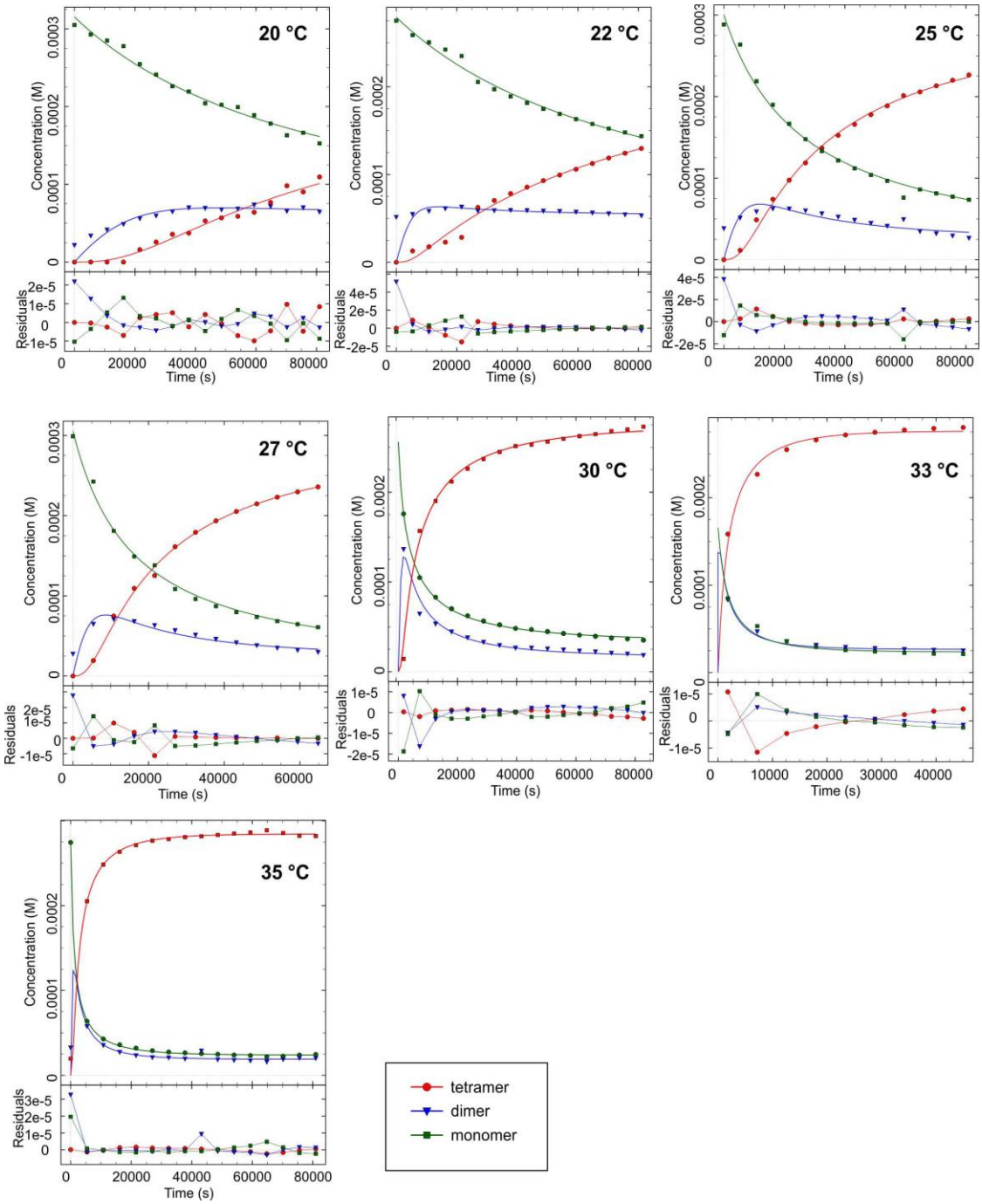

Figure S3.1: Analysis of SQT-1C oligomerisation kinetics. Fits of SQT-1C oligomerisation data at various temperatures to the monomer-dimer-tetramer kinetic model were obtained by DynaFit 4 software. Data points are shown as symbols while solid lines represent fits to the model. Attempts to fit the data to alternative models have been unsuccessful.

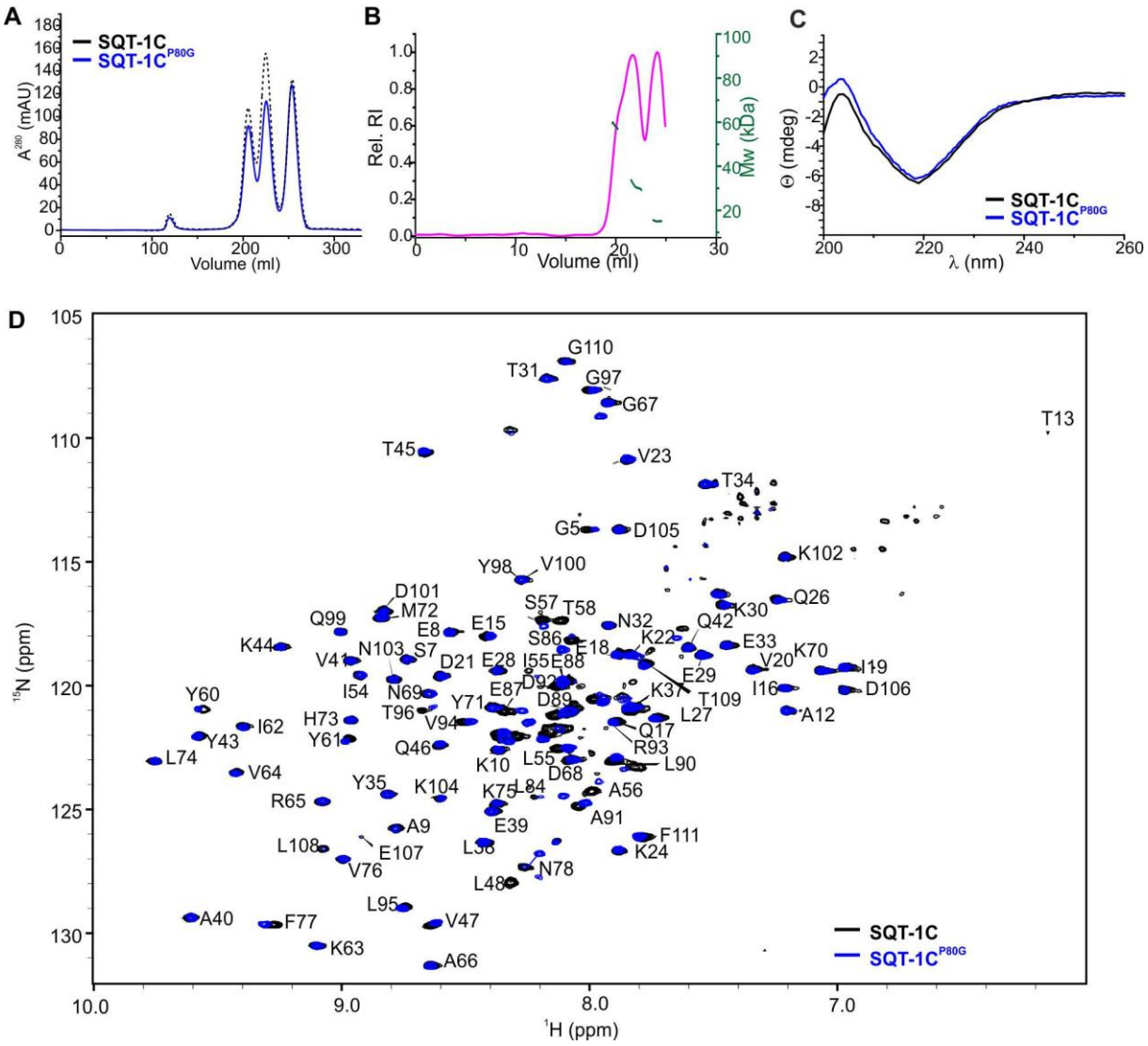

Figure S3.2: Biophysical characterisation of SQT-1C^{P80G}. A) Purification traces of SQT-1C^{P80G} (blue) and SQT-1C (black). B) SEC-MALS trace of SQT-1C^{P80G}. Relative refractive indexes (Rel. RI) are depicted in magenta, while molecular weight distribution across peaks is shown in green. C) Comparison of Far-UV CD spectra of freshly isolated monomeric fractions of SQT-1C (black) and SQT-1C^{P80G} (blue). D) Overlay of ^{15}N HSQC spectra of monomeric SQT-1C (black) and SQT-1C^{P80G} (blue).

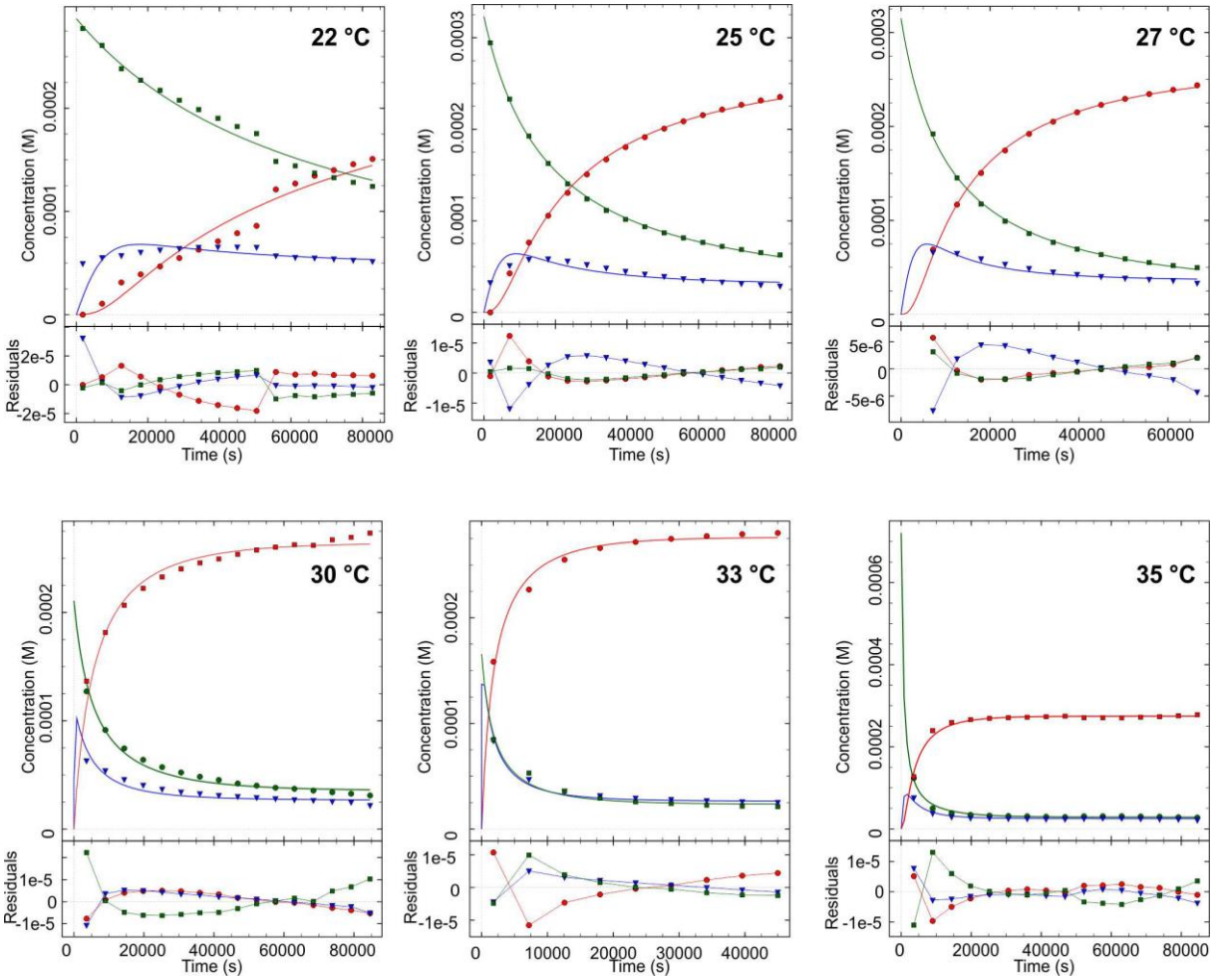

Figure S3.3: Analysis of SQT-1C^{P80G} oligomerisation kinetics. Fits to the monomer-dimer-tetramer kinetic model to SQT-1C^{P80G} oligomerisation data at various temperatures were obtained by DynaFit 4 software. Data points are shown as symbols while solid lines represent fits to the model.

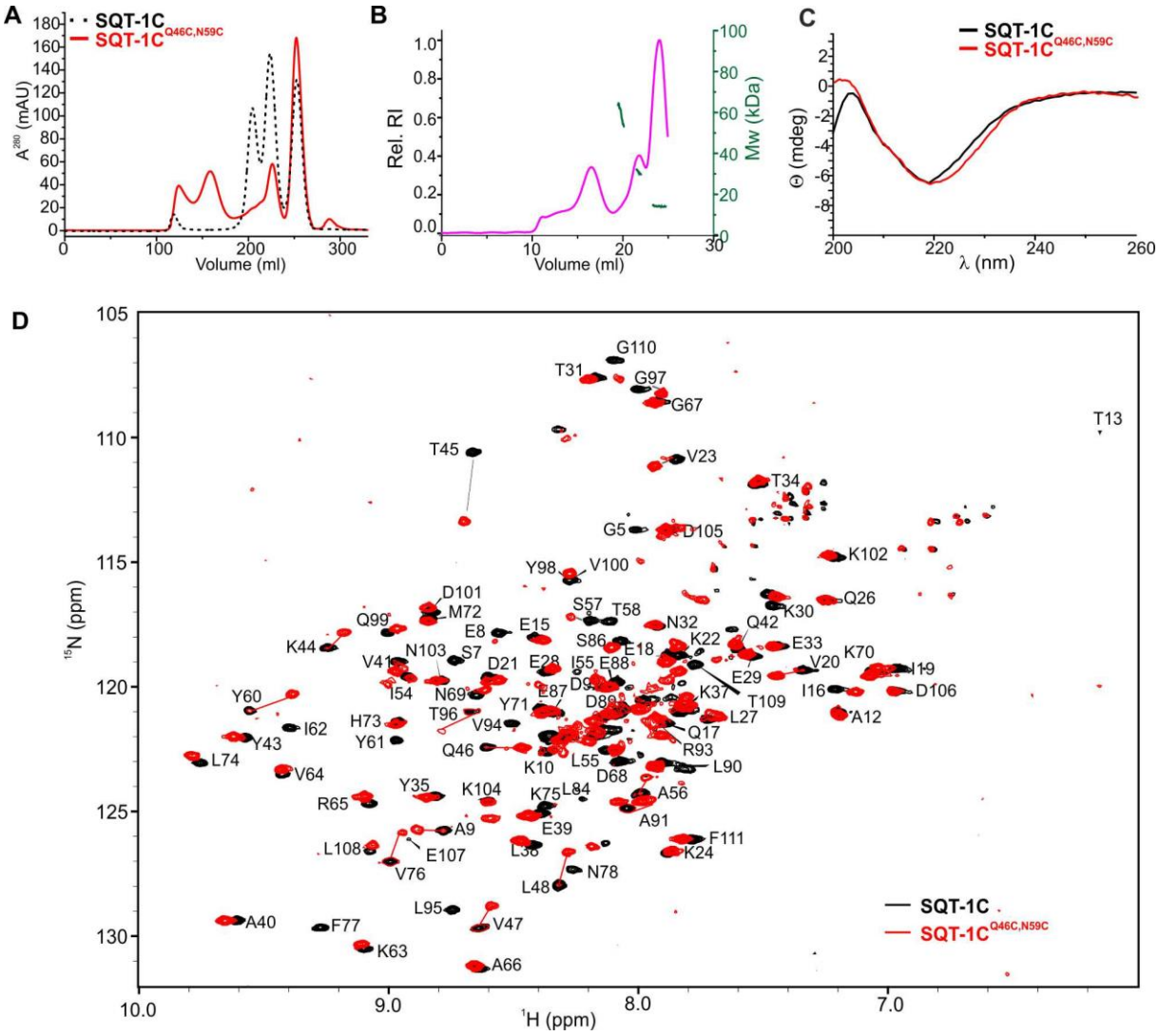

Figure S3.4: Biophysical characterisation of SQT-1C^{Q46C,N59C}. A) Purification traces of SQT-1C^{Q46C,N59C} (red) and SQT-1C (black). B) SEC-MALS trace of SQT-1C^{Q46C,N59C}. Relative refractive indexes (Rel. RI) are depicted in magenta, while molecular weight distribution across peaks is shown in green. C) Comparison of Far-UV CD spectra of freshly isolated monomeric fractions of SQT-1C (black) and SQT-1C^{Q46C,N59C} (red). D) Overlay of ^{15}N HSQC spectra of monomeric SQT-1C (black) and SQT-1C^{Q46C,N59C} (red).

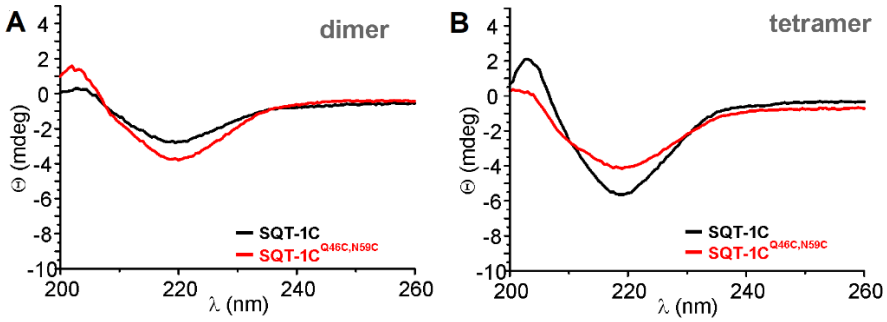

Figure S3.5: Comparison of structures of dimers and tetramers formed by SQT-1C and SQT-1C^{Q46C,N59C}. A) Comparison of Far-UV CD spectra of freshly isolated dimeric fractions of SQT-1C (black) and SQT-1C^{Q46C,N59C} (red). B) Comparison of Far-UV CD spectra of freshly isolated tetrameric fractions of SQT-1C (black) and SQT-1C^{Q46C,N59C} (red).

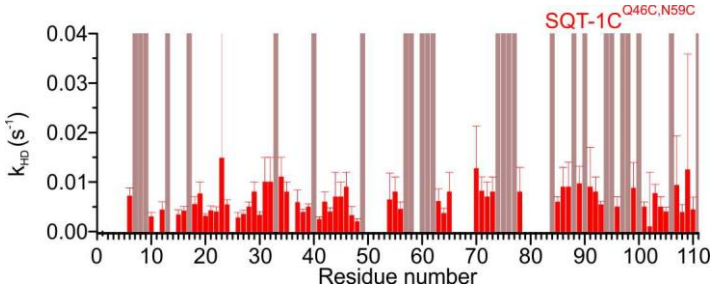

Figure S3.6: Hydrogen-deuterium (H-D) exchange rates of SQT-1C^{Q46C,N59C}. H-D exchange rates were determined by fitting signal intensities to exponential decay function plotted against residue numbers. Grey bars denote residues that were exchanged completely prior to collection of the first spectrum (i.e., within 5 minutes), while slower exchanging residues are denoted in red.

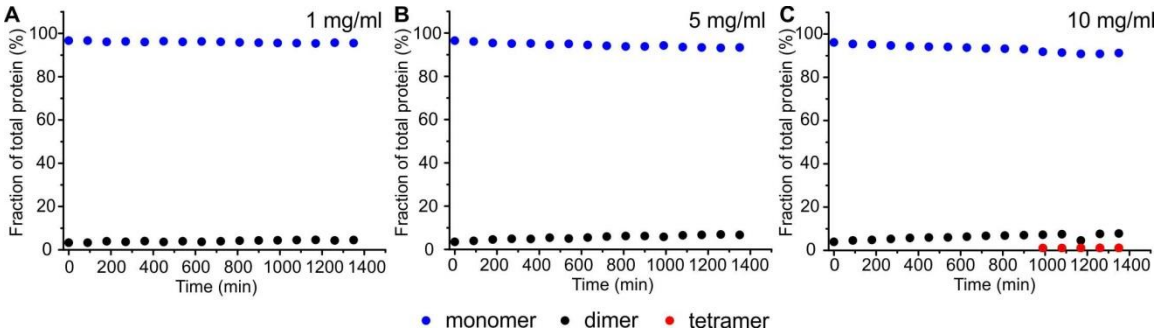

Figure S3.7: SEC analysis of SQT-1C^{Q46C,N59C} oligomerisation kinetics starting from the monomeric form. Time evolution of monomer (blue), dimer (black), and tetramer fractions (red), after incubation of monomeric SQT-1C^{Q46C,N59C} samples at 25 °C at different concentrations: A) 1 mg/ml, B) 5 mg/ml, and C) 10 mg/ml.

BLANK PAGE

4 Binding of excipients is a poor predictor for aggregation kinetics of biopharmaceutical proteins

This chapter was written, peer-reviewed and published as:

Matja Zalar[1], Hristo L. Svilenov[2], Alexander P. Golovanov[1*], Binding of excipients is a poor predictor for aggregation kinetics of biopharmaceutical proteins, European Journal of Pharmaceutics and Biopharmaceutics, 2020,151:127-136.

[1]Manchester Institute of Biotechnology and Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

[2]Ludwig-Maximillian Universität München, Department of Pharmacy, Pharmaceutical Technology and Biopharmaceutics, Butenandstr. 5, Munich 81377, Germany

* Correspondence to: A.Golovanov@manchester.ac.uk

Author contributions:

M.Z. prepared all samples, performed the experiments, analysed the data and wrote the first draft of the manuscript. H.L.S. assisted at setting up protein stability screen and provided guidance at data analysis. M.Z. and H.L.S. conceived the presented work and planned the experiments. A.P.G provided conceptual guidance. All authors contributed to paper writing.

4.1 Abstract

One of the major challenges in formulation development of biopharmaceuticals is improving long-term storage stability, which is often achieved by addition of excipients to the final formulation. Finding the optimal excipient for a given protein is usually done using a trial-and-error approach, due to the lack of general understanding of how excipients work for a particular protein. Previously, preferential interactions (binding or exclusion) of excipients with proteins were postulated as a mechanism explaining diversity in the stabilisation effects. Weak preferential binding is however difficult to quantify experimentally, and the question remains whether the formulation process should seek excipients which preferentially bind with proteins, or not. Here, we apply solution NMR spectroscopy to comprehensively evaluate protein-excipient interactions between therapeutically relevant proteins and commonly used excipients. Additionally, we evaluate the effect of excipients on thermal and colloidal protein stability, on aggregation kinetics and protein storage stability at elevated temperatures. We show that there is a weak negative correlation between the strength of protein-excipient interactions and effect on enhancing protein thermal stability. We found that the overall protein-excipient binding per se can be a poor criterion for choosing excipients enhancing formulation stability. Experiments on a diverse set of excipients and test proteins reveal that while excipients affect all of the different aspects of protein stability, the effects are very much protein-specific, and care must be taken to avoid apparent generalizations if a smaller dataset is being used.

4.2 Introduction

Protein therapeutics pose unique challenges throughout their production and formulation due to their high molecular weight, structural complexity as well as chemical and physical instabilities[9, 21-22, 68, 143]. Therapeutic proteins are prone to self-association and aggregation that can lead to the formation of potentially immunogenic particles during bioprocessing and storage, thus compromising the quality and safety of the final product[343-344]. The common consensus is that protein aggregation often proceeds through partially unfolded reactive species that associate into irreversible aggregates[80], however the mechanistic studies of aggregation pathways and how co-solutes and excipients affect these are scarce and often limited to model proteins such as insulin[345-347], Aβ peptide[348-349], sickle haemoglobin[350], lysozyme[351-352] or human serum albumin[353-355].

To prevent aggregation and improve the long-term stability of biopharmaceutics, excipients are often added at relatively high concentrations to the final formulation[70, 77, 108] to mitigate protein-protein interactions or to reduce the population of partially unfolded reactive species, thus preventing the formation of aggregates. The pool of commonly used excipients is relatively small and includes amino acids, sugars, polyols, polymers, surfactants, salts and buffering agents[7, 133]. Recently significant efforts were made to develop and characterise novel excipients that bind to protein aggregation hotspots and thus prevent aggregation[215, 356-357]. The most frequently used excipients are sugars and sugar polyols that have been shown to stabilise the native state of the proteins and reduce the population of partially unfolded reactive species[135-138, 358]. Amino acids, especially arginine salts, are often utilised, as they can reduce the aggregation and viscosity of protein solutions[147, 149, 209, 359-361], while surfactants can prevent the denaturation of proteins induced by interfaces[362-363].

To explain different effects of excipients on protein physical stability, Timasheff's theory, which postulates preferential binding/exclusion of excipients to/from protein, is often used[364]. For example, excipients traditionally considered as 'protein stabilisers' such as glycerol and sucrose are assumed to be preferentially excluded from the surface of protein molecule[365]. Although this theory may explain some of the observed effects, it does not allow to predict the effects for a given protein, perhaps because some experimental assessment is needed of whether the excipient actually preferentially binds to a protein (albeit weakly) or not. As protein-excipient interactions are usually transient and weak, with dissociation constants K_d in higher mM range[209], the standard methods for quantifying protein-ligand interactions like ITC, SPR or fluorescence are not applicable[94, 366-367]. Moreover, it has been

also shown that an increase in thermal stability itself does not necessarily correlate with other aspects of physical and long term stability[126, 368-370]. This raises the question of whether there is any correlation between experimentally-measured protein-excipient interaction ('preferential binding' or 'preferential exclusion') and various protein stability characteristics.

Due to difficulties mentioned above, obtaining stable protein formulation still requires time- and material-consuming combinatorial screening. Such screenings usually rely on convenient measures of thermal stability parameters, i.e. by measuring the protein apparent melting temperature (T_m) or the onset temperature of aggregation (T_{agg}), which correlate well with the rates of aggregation at elevated temperatures[371] but often fail to predict aggregation rates at realistic storage conditions[157, 368, 372-374]. A recent systematic study on a set of pharmaceutically relevant proteins showed that correlations between results obtained by orthogonal methods are only weak and showed the need for developing combined risk scores that provide a more reliable measure of protein stability[375].

While high-throughput screening methods enable fast characterisation of formulations in practice, there is a need to gain more fundamental understanding of aggregation mechanisms and the impact of co-solutes, including excipients, on different aspects of protein stability. Protein degradation and aggregation can be evaluated with various biophysical methods such as size exclusion chromatography, dynamic and static light scattering, UV-Vis, FT-IR, and CD, however these methods fail to detect subtle differences in structure induced by protein-excipient interactions[6]. Solution NMR spectroscopy can detect structural changes but, importantly, can also detect weak protein-excipient interactions through range of experiments, such as saturation transfer difference (STD)[251], Carr-Purcel-Meiboom-Gill (CPMG)[247] and WaterLOGSY[261-262]. These experiments report on different aspects of weak binding and are frequently used as a combination in fragment-based drug discovery. NMR has been successfully applied recently to detect specific protein-surfactant interactions[216, 376], to evaluate protein interactions and stability[209, 219] and to analyse anion binding to a protein[117].

Here, we comprehensively characterise protein-excipient interactions with four different solution NMR spectroscopy approaches using a set of seven pharmaceutically relevant model proteins, with diverse molecular surface properties, and a representative set of eleven excipients typically used in protein formulation. We propose the use of a weighted

empirical interaction parameter to evaluate very weak interactions between proteins and excipients, and distinguish between non-interacting and interacting excipients. Using orthogonal biophysical techniques, we also evaluate how excipients affect the protein thermal and colloidal stability and assess the impact on protein aggregation kinetics at elevated temperature. Finally, we assess the effect of excipients on protein storage stability. We show that excipients affect both the nucleation and aggregate growth stages of protein aggregation, with nucleation rate weakly correlating with colloidal stability, and aggregate growth rate weakly correlated with protein thermal stability. We also show that there is only a very weak correlation between the empirical protein-excipient interaction parameter and the overall effect this excipient has on protein thermal stability. To our knowledge, this is the first study attempting to explore directly and systematically a link between weak protein-excipient interaction and protein formulation stability.

4.3 Results

4.3.1 Experimental design

In this study, we have used four IgG1 monoclonal antibodies, namely, PPI03, PPI04, PPI10 and PPI13, IgG2 monoclonal antibody PPI17, a bispecific antibody PPI08 and a human serum albumin conjugated to neprilisyn (HSA-NEP), PPI18. They were formulated at low ionic strength in 10 mM sodium phosphate buffer, pH 6.5 together with eleven commonly used excipients. Selected excipients were five amino acids or their combinations, namely, arginine hydrochloride (ArgHCl), sodium glutamate (NaGlu), arginine glutamate (ArgGlu), glycine (Gly), Proline (Pro), two most frequently used sugars trehalose and sucrose, two sugar polyols (glycerol and sorbitol) and two surfactants, namely Tween 20 and poloxamer 407. Chemical structures of these excipients are shown in Figure 4.1. Samples of protein in buffer alone and with NaCl added were used as controls. To thoroughly evaluate effect of excipients on different aspects of protein stability, we have assessed the protein-excipient interactions using NMR, studied protein thermal stability using high-throughput fluorometric analysis of thermal protein unfolding with nanoDSF and light back-scattering, measured protein aggregation kinetics using light back-scattering at elevated temperatures, and performed an accelerated stability study.

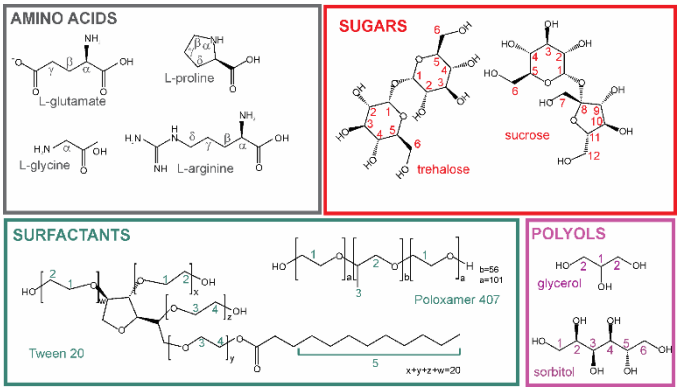

Figure 4.1: Chemical structures of excipients used in this study together with annotations of NMR-visible atom positions.

4.3.2 Characterisation of protein-excipient interactions using different NMR spectroscopy approaches

First, protein-excipient interactions were assessed using NMR spectroscopy. A combination of four ^{1}H experiments, namely simple 1D spectrum, transverse T_2 relaxation filter (CPMG filter), saturation transfer difference (STD) and WaterLOGSY spectra, were acquired for each titration point. Chosen NMR experiments were complementary to each other, capturing different aspects related to the same protein-excipient interaction event. This approach was used to increase the chance of identifying real interactions rather than false-positives and false-negatives. While STD and WaterLOGSY experiments rely on transfer of magnetization from protein or protein-bound water molecules to ligand by nuclear Overhauser effect (NOE)[251, 262], a CPMG experiment measures changes in T_2 relaxation times as small excipient molecule binds to large protein[196]. The ratio of excipients' signal intensity in ^{1}H 1D spectra in the presence and in the absence of proteins can report on binding when compared to the reference spectra of the excipient in the absence of protein, due to differential signal broadening[377-379] that occurs upon protein-excipient interaction.

In the first instance, a simple qualitative characterisation of protein-excipient interactions was performed (hit identification), to classify excipients into interacting and non-interacting. First, STD amplification factors (AF^{STD}) and ratios of excipients' signal intensities in the presence and in the absence of protein were calculated as described in section 4.5.5. Control experiments were run on excipient samples in the absence of protein to prevent direct irradiation of excipients and eliminate false positive signals due to excipient cluster formation. If any excipient-specific effect was observed, these were subtracted from

the effects in the presence of proteins. To account for uncertainties in peak position, imperfection of baseline correction and differences in shimming, 5% permissible error threshold was chosen for ^{1}H and CPMG filter experiments, while a 10% error threshold was used for WaterLOGSY experiments. It was assumed that the changes above these thresholds were caused by excipient binding. Excipient was identified as an interacting hit if positive AF^{STD} were calculated for its signals or if the calculated ratio for a particular experiment was smaller than 0.95 (0.9 for WaterLOGSY). The exemplary data for ArgHCl that was identified as a hit in the presence of PPI10 is shown in Figure 4.2 while the whole data set is presented in Table S 4.1. The results of the hit identification screen are summarised in Figure 4.3. It should be noted that in principle, NMR experiments used report not only whether an excipient interacts with the protein, but also reflect the strength of the interactions. For example, differential line broadening, as observed in decreased signal intensity in ^{1}H spectra, is proportional to the amount of bound excipient and the strength of protein-excipient interaction[380]. Similar correlations have also been observed for CPMG and WaterLOGSY experiments, while AF^{STD} are directly proportional to the strength of the interaction, proximity of excipient molecule to the protein and residence time, which allows for direct comparison of the interactions and allows for evaluation of the binding mode of excipients with the proteins[259].

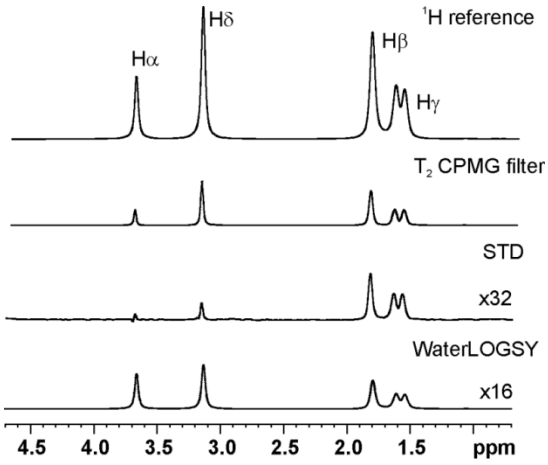

Figure 4.2: Example of using different NMR approaches for the evaluation of protein-excipient interactions. ^{1}H reference, T_2 CPMG filter, STD and WaterLOGSY spectra of ArgHCl in the presence of PPI10. Peak assignments are shown next to the corresponding peaks on the reference spectrum. In order to show the difference peaks more clearly, scaling of STD and WaterLOGSY spectra were increased 32 and 16-fold, respectively, relative to the reference spectrum.

From the qualitative analysis we can see that charged excipients (ArgHCl, ArgGlu and NaGlu) and surfactants (poloxamer 407 and Tween 20) bind to majority of the tested proteins, while the other interactions are more protein-specific. In general, different NMR experiments often identify different excipients as hits and thus can be considered complementary with each other. In this screen WaterLOGSY proved to identify the largest amount of positive hits, however we should note that only a decrease in signal intensity compared to the control sample$\left(\frac{I^W}{I^W_{REF}}\right)$ was observed, as the binding of excipients was too weak to completely invert the NOE signal. Therefore, some of the identified hits might be false positives. Differential broadening observed in the ^{1}H spectra proved to be useful in determining potential binders, however, the linewidth in this experiment might be influenced not only by the interaction with the protein but also by the quality of shimming and changes in viscosity upon addition of protein, which might also result in false positives. The ^{1}H T_2 CPMG filter experiment was the least sensitive to detect protein-excipient interactions in chosen experimental conditions, where excess of excipients was used and thus the observed signal was dominated by non-bound ligand from the bulk solution.

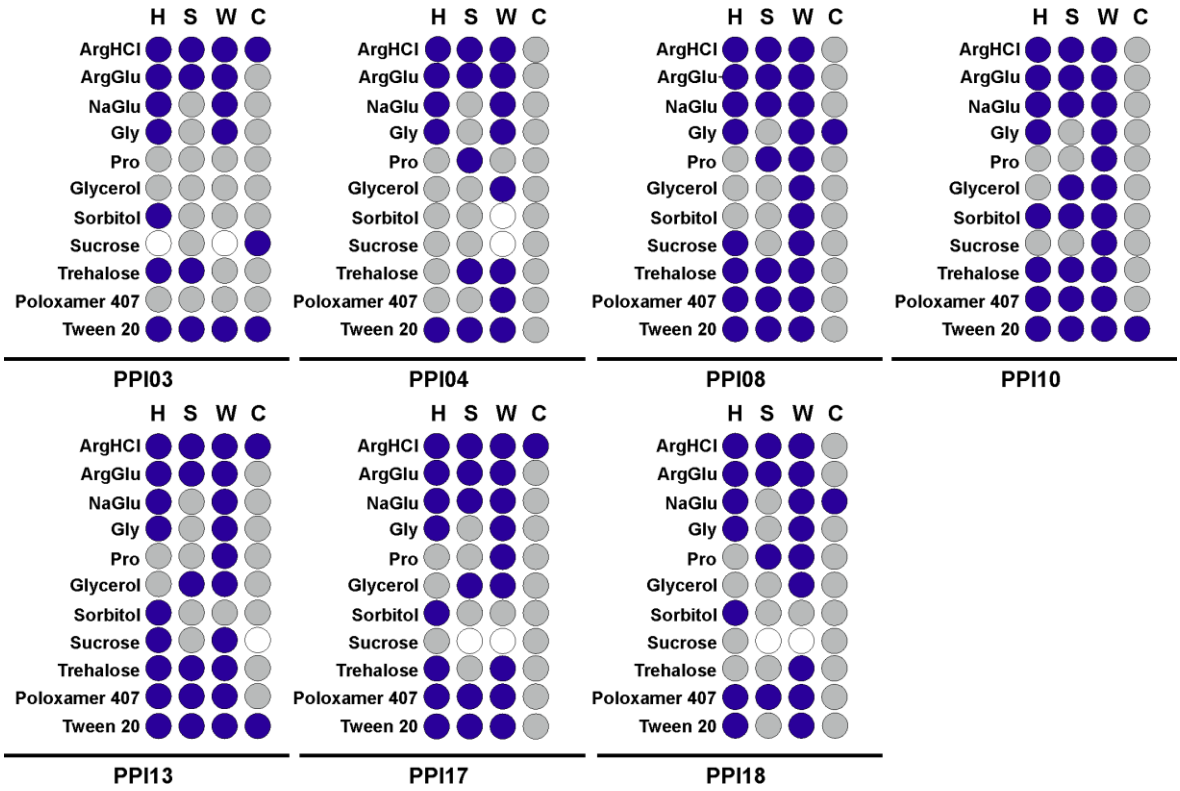

Figure 4.3: Qualitative screening of protein-excipient interactions by NMR. Qualitative evaluation of protein excipient interactions as detected by ^{1}H (H), STD (S), WaterLOGSY (W) and CPMG (C) NMR methods. Positive hits for excipient binding identified by individual method are denoted in blue, non-hits in grey while data points excluded from analysis due to experimental error are indicated with empty circles.

4.3.3 Assessing protein-excipient interactions using NMR-derived parameter I^N

Since different NMR methods appear to identify excipients as binders somewhat inconsistently, some sort of scoring is needed for more reliable identification of binders, and to rank them relative to each other. Here we introduce a weighted empirical protein-excipient interaction parameter (I^N) that reflects all four NMR observables at once. In the calculation of the interaction parameters equal weights were given to all four experiments but the numerical values obtained can be used as a proxy for interaction "strength" in a broad sense (Eq. 4.3). Interaction parameter also provides an empirical cut-off to distinguish between excipients which interact or not with proteins. The interaction was considered significant if I^N was greater than an arbitrary threshold of 0.25. This threshold corresponds to a situation where at least one of the four NMR approaches used identified an excipient as the strongest binder of all tested, or several approaches identified excipient as an intermediate binder. The

theoretically-possible value of I^N=1 corresponds to a situation when all four approaches identified excipient as the strongest binder. As seen from Table 4.1, charged amino acids bind to all the proteins, except to PPI03, but the interaction, reflected by I^N, does noticeably vary between proteins. Interestingly, proline and glycine interact with the majority of the tested proteins, while sugars in sugar polyols generally do not interact with the tested protein, except in a few selected cases. Poloxamer 407 binds to PPI08, PPI10, PPI13 and PPI18, while Tween 20 is the strongest binding excipient in the test set, interacting with all of the tested proteins relatively strongly. It is interesting to note that PPI08, PPI10, PPI13 and PPI18 interact with more than half of the tested excipients.

Such analysis provides a deeper insight into protein-excipient interactions, allowing for crude scoring and ranking of excipients based on the interaction parameter I^N. As the next step we evaluated effect of these excipients on protein thermal stability, kinetics of aggregation and accelerated stability study using other biophysical techniques, to relate these with the values of protein-excipient interaction parameter I^N.

Table 4.1: NMR-derived excipient-protein interaction parameters I^N

Excipients	**PPI-03**	**PPI-04**	**PPI-08**	**PPI-10**	**PPI-13**	**PPI-17**	**PPI-18**
ArgHCl	-	0.48	0.50	0.27	0.60	0.62	0.33
ArgGlu	-	0.46	0.57	0.29	0.29	0.39	0.30
NaGlu	-	0.30	0.52	0.33	-	0.49	0.27
Glycine	0.29	-	0.39	0.30	0.42	-	0.29
Proline	-	0.33	0.43	0.41	0.41	0.40	0.43
Sucrose	-	-	-	0.43	-	-	-
Trehalose	-	-	-	-	-	-	0.66
Glycerol	-	-	0.26	-	-	-	-
Sorbitol	-	-	-	0.40	-	-	-
Poloxamer 407	-	-	1.00	0.27	0.34	-	0.35
Tween 20	0.65	0.52	0.56	0.53	0.50	0.46	0.79
**interaction parameter could not be determined as the excipient does not have NMR observable signals - Protein-excipient interaction was considered insignificant, with I^N<0.25*							

4.3.4 Effect of excipients on protein thermal unfolding and aggregation

The effect of excipients on thermal stability was studied in terms of the apparent melting temperature (T_m), and onset temperatures of aggregation (T_{agg}) for proteins used in

this study. All proteins exhibited two unfolding transitions during thermal unfolding as measured by the change in intrinsic fluorescence intensity ratio ($F_{350/330}$), except for the conjugated protein, PPI18, which showed three transitions (Figure S4.1) Simultaneously, the back scattering data was collected and the onset temperature of aggregation was determined and was used together with the difference between the apparent first melting transition temperature and onset temperature of aggregation (T_m-T_{agg}) as a measure of colloidal stability (Figure 4.4). The full dataset on thermal stability is presented in Table S4.2, while the effect of excipients as compared to the control in buffer alone is shown in Figure S4.2. In general, we can observe that excipients have a bigger effect on the onset temperature of aggregation, and a smaller one on individual apparent temperatures of melting transitions. We can also see that for PPI04, PPI08, PPI17 and PPI18 aggregation precedes or coincides with T_{m1}, indicating low colloidal stability of these proteins, while for PPI03, PPI10 and PPI13 aggregation occurs well beyond it, indicating higher colloidal stability of the partially unfolded states.

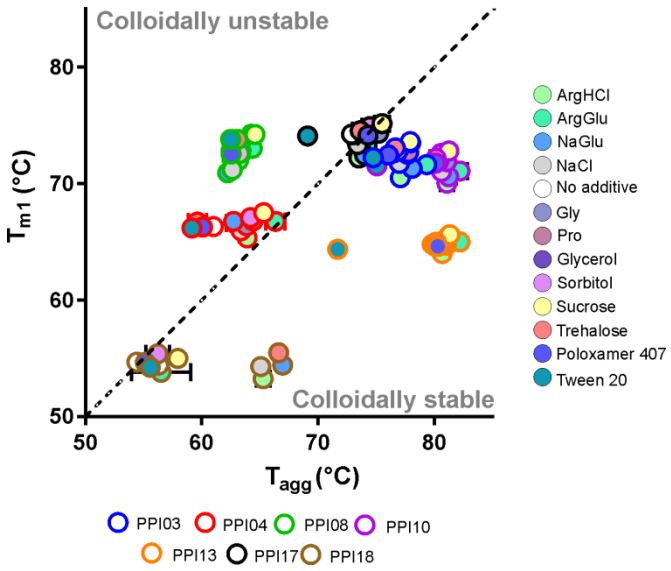

Figure 4.4: Relation between the first apparent melting transitions and aggregation onset temperatures. Outline of the data points indicates the protein while the fill colour corresponds to excipients. The black dashed line shows a guideline for x=y linear correlation.

As evident from Figure 4.4 and Table S4.2, charged amino acids (ArgHCl and NaGlu) decrease the T_{m1}, however when used in equimolar mixture (ArgGlu), they have a much smaller effect on all of the melting transitions. Nonpolar amino acids used in this study

(Gly, Pro) do not significantly affect the T_m values and only marginally improve the T_{agg}. Addition of trehalose, sucrose and sorbitol increased both apparent melting temperatures and the aggregation onset temperatures. Glycerol and poloxamer 407 did not significantly affect the unfolding temperatures and only marginally affected the aggregation onset temperatures, while Tween 20 did not affect the apparent melting temperatures but significantly decreased onset temperature of aggregation of all proteins (Figure 4.4 and Table S4.2). In general, we observed that charged excipients improved the apparent colloidal stability of the proteins, while the neutral ones either did not affect or lowered the apparent colloidal stability of the proteins. The surfactants had the most profound negative effect on the protein colloidal stability, lowering it significantly for all of the tested proteins, except for PPI18.

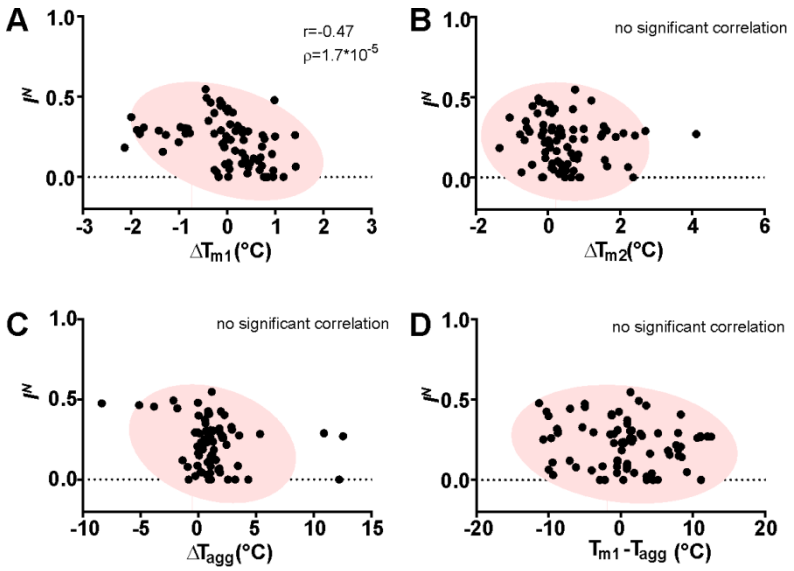

Figure 4.5: Effect of protein-excipient interactions on protein thermal stability. A) Correlation between the effect of excipients on protein first melting transition and NMR-derived excipient-protein interaction parameter (I^N) B) Correlation between the effect of excipients on protein second melting transition and NMR determined interaction parameter (I^N). C) Correlation between the effect of excipients on aggregation onset temperatures and NMR-derived interaction parameter (I^N) D) Correlation between T_m-T_{agg} values and NMR-derived excipient-protein interaction parameter (I^N) Experimental data points are shown in black, while the 95 % confidence ellipse obtained during calculation of Spearman correlation coefficient (r) is shown in pale red.

To see how protein-excipient interactions influence protein thermal and colloidal stability, we have explored the relationship between all of the parameters describing thermal and colloidal stability of the protein with NMR derived interaction parameter I^N (Figure 4.5). Interestingly, we observed a weak negative correlation only between I^N and T_{m1}, while no significant correlation was observed between other parameters. Stronger-interacting excipients were found to be mostly reducing T_{m1} and destabilising proteins.

4.3.5 Effect of excipients on protein aggregation kinetics

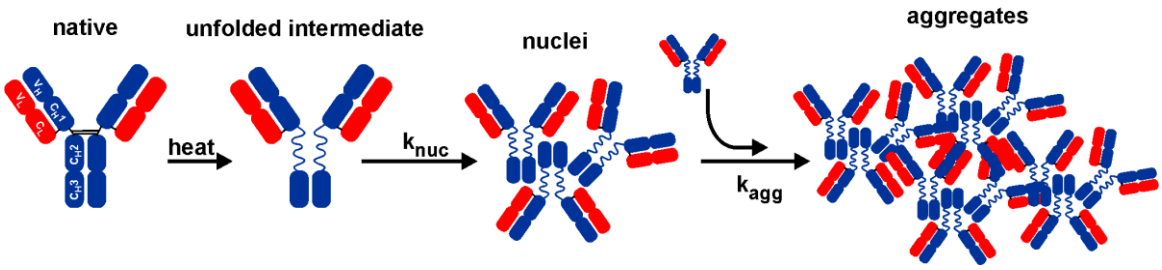

Figure 4.6: Scheme of the proposed aggregation mechanism

Next, the effect of excipients on protein aggregation kinetics was evaluated. As the formation of aggregates under typical storage conditions (4°C) is very slow, it was not practical to study the mechanism and kinetics of aggregation under these conditions within a realistic project timescale. The common consensus is that protein aggregation occurs through interactions of partially unfolded species [80]. While the kinetics of this process is temperature dependant the aggregation itself is expected to proceed through the same mechanism. We have therefore monitored the effect of excipients on protein aggregation kinetics at temperatures above the T_{m1} so that a uniform population of partially unfolded species was obtained. The process of aggregation was monitored for 12 h at this temperature, as detected by the back-scattering signal, as aggregates form in solution. It should be noted that in this type of turbidity measurement-based assay smaller aggregates cannot be detected which results in the typical lag-phase before the rapid growth of aggregates is observed. Such an assay allows for a quick assessment and selection of excipients that slow down non-native aggregation. The acquired time dependant data was fitted to a simple two-step Finke-Watzky polymerisation mechanism model (Eq. 4.1)[103], that defines the aggregation process

in terms of rate of nucleation (k_{nuc}) and rate of aggregate growth (k_{agg}). The overall scheme of the simplified two-step polymerisation mechanism of protein aggregation is shown in Figure 4.6 and example of experimental data is shown in Figure S4.4.

Effect of excipients on the rate of nucleation (k_{nuc}) and aggregate growth (k_{agg}) was evaluated relative to the reference sample without excipients added (Figure S4.5). In general, ArgHCl, ArgGlu and NaGlu slow down rate of nucleation to a similar extent as NaCl. Glycine, proline, glycerol and sorbitol do not significantly affect the nucleation stage of protein aggregation but they significantly slow down the rate of aggregation of PPI04, and increase the rate of nucleation for PPI17, most likely due to protein specific effects. Sucrose in general decreases the rate of nucleation, apart for PPI17 where it significantly increases it. Trehalose on the other hand does not significantly affect the rate of nucleation step. Interestingly, both surfactants speed up nucleation.

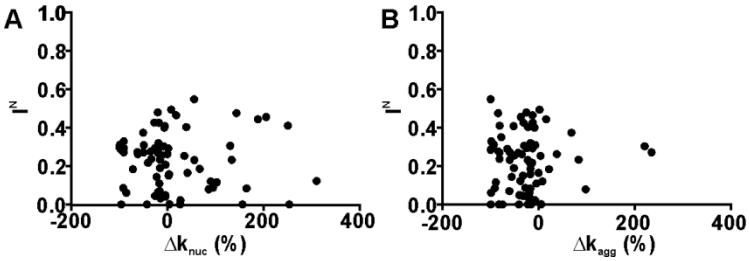

Figure 4.7: Excipient-protein interaction parameter (I^N) does not correlate with protein aggregation kinetics. A) Correlation between the effect of excipients on rate of nucleation (k_{nuc}) and NMR determined interaction parameter (I^N) B) Correlation between the effect of excipients on rate of aggregate growth (k_{agg}) and NMR determined interaction parameter (I^N). No significant correlations were observed

Effects of excipients on the rate of aggregate growth are on the other hand more universal (Figure S4.5). All non-ionic excipients slow down the apparent rate of aggregate growth. Excipients that increase the ionic strength of the protein solution generally decrease the rate of aggregate growth, except for PPI08 and PPI13. In general, we observed very little correlation between the effects of excipients on protein thermal stability and rate of aggregate growth, and even less for the nucleation step (Figure S4.5), indicating that the rate of aggregate growth depends primarily on the fraction of the unfolded species in solution. This behaviour is typical if the aggregate growth proceeds by amorphous addition of partially

unfolded protein monomers, while other mechanisms prevail in the nucleation stage. Additionally, no significant correlation was observed between the NMR-derived excipient-protein interaction parameter I^N (determined at lower temperature) and the effect of the excipient on the aggregation kinetics at temperature above T_{m1}, indicating that general binding of excipients to proteins is not a contributing factor in protein aggregation kinetics, especially at elevated temperatures (Figure 4.7).

4.3.6 Effect on excipients on proteins in an accelerated stability study

To comprehensively analyse the effect of excipients on different aspects of protein stability, an accelerated stability study was performed, during which the stability of all studied proteins was assessed with size-exclusion chromatography (SEC), dynamic light scattering (DLS) and measurements of total soluble protein concentration. The measurements were carried out at 0, 2 and 4 weeks of storage at elevated temperature and the results are summarised in Figure S4.6. Monomer recovery of PPI03, PPI10, PPI13, PPI17 and PPI18 could be attributed to formation of soluble high molecular weight species (HMW) and to a lesser extent to fragmentation (LMW). On the other hand, the majority of monomer loss in the case of PPI04 and PPI08 was due to formation of larger insoluble aggregates that were filtered out prior to SEC analysis. Generally, addition of charged excipients (ArgHCl, ArgGlu, NaGlu and NaCl) depleted the amount of soluble aggregates, while the addition of sugars slightly increased the amount of soluble aggregates present in solution, but those were generally smaller and more uniformly sized than aggregates formed in the presence of charged excipients. While addition of proline, glycine and sorbitol did not significantly influence the formation or the size of aggregates in accelerated stability study, except in selected cases, addition of glycerol resulted in increased number of soluble aggregates that were generally smaller than those formed in the control conditions. Overall, these data with wide range of excipients and wide range of proteins, suggest no simple rules and no noticeable correlations, underlying the necessity to run extensive screening experiments to find best excipient for each particular protein target of interest.

4.4 Discussion

In the study presented here we have comprehensively assessed protein-excipient interactions in a range of pharmaceutically relevant formulations using a combination of several NMR approaches, and examined the effect of excipients on various aspects of protein stability using more traditional screening methods. We show that differential line broadening, CPMG, STD NMR, and WaterLOGSY, used here to assess these weak

interactions, show slightly different picture of excipient binding, despite reporting on the same process. In our study, WaterLOGSY identified the biggest number of apparent positive hits for binders, possibly due to unspecific effects, while CPMG was the least sensitive one, because the observed relaxation rate was dominated by the large excess of non-bound ligand. The observed NMR effects in different experiments are influenced to different extents by off and on rates of binding, diffusion and residence times, different sensitivity to binding, and are affected by experimental and measurement artefacts to different extents. These effects mean that binding may be detected by one experiment, but not another, thus highlighting the need to use a consensus of multiple NMR approaches. A similar strategy is frequently employed in fragment-based screening where different types of NMR experiments typically identify different hits, which are then typically tested further with complementary biophysical methods to identify 'true hits'[381-383]. In protein formulation similar approach cannot be easily implemented as there are no complementary non-NMR techniques readily available to test for such a weak binding. Therefore, an alternative could be to quantify excipient-protein binding in terms of the empirical parameter I^N, which we have introduced here and which allows not only for quick identification of binding but also for rough ranking of the strengths of the interaction for different proteins and excipients. Additionally, artefacts originating from some features of the small molecule can affect the NMR results, causing false positive and false negative hits. For example, exchangeable protons always produce a false positive signal in WaterLOGSY experiments, while self-association of the small compound itself may produce false positive results in all three types of the experiments. Therefore, it is important to use a consensus parameter such as I^N to average out such a 'noisy' data.

We show here that, under low-salt conditions chosen for the study, sucrose and polyols have a fairly universal effect on protein thermal stability, while the effects of other excipients are more protein specific. Excipients either increase or decrease the melting temperatures compared to the control samples. It has been established previously that if the small molecule binds to the native state of the protein and stabilises it, the melting temperatures increase, while preferential interaction with partially or fully unfolded state decreases the thermal stability[384-385]. As it is generally accepted that if excipient binds to the native state of the protein, this will stabilise the protein and increase its T_m value, the observed destabilisation of proteins here can be explained by excipient binding to partially-unfolded state. Therefore, maximising overall protein-excipient binding itself is unlikely going to be a successful

strategy: instead, binding to the folded form needs to be maximised. At the moment there is no clear way to discriminate, using NMR or other ligand-observed techniques, whether an excipient binds to the folded, or to the partially-unfolded state of a protein. Therefore, the excipient binding as detected by NMR in general cannot be used itself for predicting whether an excipient will be stabilising or destabilising in terms of melting temperature, neither it can be used to predict the change in colloidal stability. However, there is potential to further explore binding of excipients to different non-native states of the protein by using denaturants, temperature or high pressure to obtain a non-native protein states, methods that are often employ to study protein unfolding/folding pathways[386-387].

When we look at the wide range of proteins and excipients, there is a lack of obvious correlations or "hard rules" for different types of excipients as judged by the accelerated stability study and screen of aggregation kinetics. What may be beneficial for one protein may be detrimental for another one. Different parameters used as a measure of stabilising effects in protein formulation, e.g. general monomer loss, polydispersity index, rate of aggregation, apparent hydrodynamic radius, all reflect on different attributes of formulation and provide largely orthogonal views, not correlated with each other in any obvious way. In our study, no excipient showed universally stabilising effect in the comprehensive accelerated stability screen used.

Additionally, we have observed that apparent trends that may be observed for one protein or a group of similar proteins disappear in a larger dataset, which raises a question of how relevant are case studies previously published on simple single model proteins (often, widely-available lysozyme and BSA), in terms of the effect of various excipients and conditions, and how such isolated case studies can help to understand the behaviour and response of other proteins to addition of excipients. The complexity observed here raises a question of how simple theoretical methods can capture the variety and complexity of the formulation space. New, more complex methods need to be developed to take into account much more parameters possibly considering not only the structures of the folded states, but also of partially-unfolded states, protein interactions between folded and partially-unfolded states, and interactions of excipients with both of these. One obvious difficulty for measuring weak binding is that the ligand needs to be in significant excess, which creates dynamic range problem, and overall fraction of bound ligand also becomes small and difficult to measure. New NMR methods may need to be developed to assess the very weak ligand binding to proteins, given that the need for these has now been identified.

4.5 Materials and Methods

4.5.1 Sample preparation

All the proteins used in the study, which included PPI03, PPI04, PPI10, PPI13 (monoclonal IgG1 antibodies), PPI17 (IgG2 monoclonal antibody), PPI08 (bispecific antibody), and PPI18 (a human serum albumin conjugated to neprilisyn, HSA-NEP), were provided by the PIPPI consortium. Protein sequences and basic characterisation data will soon be available on the PIPPI-data database. Protein molecular weights, isoelectric points and extinction coefficients at 280 nm are summarised in Table 4.2.

Table 4.2: Molecular weights, isoelectric points and extinction coefficients of proteins in the dataset

Protein	**PPI03**	**PPI04**	**PPI08**	**PPI10**	**PPI13**	**PPI17**	**PPI18**
MW (kDa)	144.8	146.2	204.4	144.2	148.9	145.1	146.7
pI	8.4	9.0	9.0	8.9	9.1	7.8	5.8
ε_{280} (ml mg^{-1} cm^{-1})	1.435	1.755	1.57	1.533	1.66	1.31	1.04

All proteins were extensively dialysed at 20 °C using Slide-A-Lyser™ cassettes (Thermo Fisher Scientific) with 10 kDa molecular weight cut-off against excess of 10 mM sodium phosphate buffer with pH 6.5 until components of original formulation were no longer visible in 1D ^{1}H NMR spectra. Protein concentrations were determined by measuring UV-absorption at 280 nm using NanoDrop 2000 (Thermo Fisher Scientific). Protein stock solutions with concentration 5 mg/mL were then prepared. Stock 2x solutions of excipients (ArgHCl, ArgGlu, NaGlu, Gly, Pro, sorbitol, glycerol, sucrose, trehalose, poloxamer 407, Tween 20, and NaCl) were prepared in 10 mM sodium phosphate buffer, pH 6.5 and mixed in 1:1 ratio with protein stock solutions to obtain final formulations that are presented in Table 4.3. All formulations were sterile filtered with 0.22 μm cellulose acetate filter (Millipore). The solutions were then aseptically filled into pre-sterilised 0.5 ml Protein LoBind tubes (Eppendorf), with 330 μL in each tube.

Table 4.3: Final composition of each of the formulation tested

Formulation 1	**Formulation 2**	**Formulation 3**	**Formulation 4**
2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 100 mM ArgHCl	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 100 mM Arg 100 mM Glu	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 100 mM NaGlu	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 200 mM Pro
Formulation 5	**Formulation 6**	**Formulation 7**	**Formulation 8**
2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 200 mM Gly	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 100 mM NaCl	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 200 mM Glycerol	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 200 mM sucrose
Formulation 9	**Formulation 10**	**Formulation 11**	**Formulation 12**
2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 200 mM trehalose	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 0.05% poloxamer 407	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5 0.05% Tween 20	2.5 mg/ml protein 10 mM sodium phosphate buffer pH 6.5

4.5.2 Initial assessment of protein stability

Apparent melting temperatures (T_{m1}, T_{m2}, T_{m3}) of the first, second and third thermal melting (unfolding) transitions and aggregation onset temperatures (T_{agg}) of proteins were determined with nanoDSF using a Prometheus NT.48 (NanoTemper Technologies). Protein samples were filled into standard glass capillaries (NanoTemper Technologies) and thermal unfolding was monitored by measuring intrinsic protein fluorescence intensity at 330 and 350 nm after excitation at 280 nm across a temperature ramp, from 25 to 100 °C, with a heating rate of 1 °C/min. All measurements were performed in triplicates. Analysis of the fluorescence intensity ratio at these two wavelengths (F_{350}/F_{330}) plotted against temperature was performed using the PR.ThermControl V2.1 software (NanoTemper Technologies) where inflection points of thermal unfolding were determined using the first derivative method, and the inflection points were reported as apparent melting temperatures. Simultaneously, the back-reflection intensity of a light beam that passes through the sample capillary was measured and analysed using the PR.ThermControl V2.1 software

(NanoTemper Technologies). The aggregation onset temperature (T_{agg}) was determined from the increase in the back-reflection intensity over the temperature ramp, using the same software.

4.5.3 Isothermal protein aggregation kinetics

Protein aggregation kinetics were monitored using the Prometheus NT.48 (NanoTemper Technologies) at elevated temperature. Incubation temperatures were chosen so that the proteins were only partially unfolded and were 60 °C for PPI04, 65 °C for PPI08 and 18, and 75°C for PPI03, 10, 13 and 17. Isothermal time-dependent aggregation curves were fitted to a Finke-Watzky mechanism of nucleation, followed by aggregate growth[104]. The data was fitted using the equation **Error! Reference source not found.**, where *Y(t)* is t he back-reflection intensity, *t* is time, y_f is maximal scattering intensity, k_{nuc} is the apparent rate of nucleation, and k_{agg} is the apparent rate of aggregate growth [106].

$$Y(t) = y_f - \frac{\frac{k_{nuc}}{k_{agg}} + y_f}{1 + \frac{k_{nuc}}{y_f k_{agg}} e^{\left[\frac{k_{nuc} + k_{agg} * y_f}{t}\right]}} \qquad \text{(Eq. 4.1)}$$

The effect of excipients on the rates of nucleation and aggregate growth were compared to k_{nuc} and k_{agg} values measured in the reference sample where no excipients were added to the formulation buffer.

4.5.4 Accelerated stability study

Based on initial assessment of protein thermal and colloidal stability, PPI04 and PPI08 were stored at 40 °C and other proteins at 50 °C, for the duration of the accelerated stability study. At each time point (0, 2 and 4 weeks of storage) three samples from different tubes were used for the analysis of each formulation using size-exclusion chromatography (SEC) and dynamic light scattering (DLS).

Dynamic light scattering

A 5 µL aliquot of each formulation was pipetted into 1536 well LoBase plate (Aurora Microplates Inc.) in duplicates. After the plate was centrifuged for 2 min at 2000 rpm using Heraeus Megafuge 40 centrifuge with M-20 well plate rotor (ThermoFisher Scientific), 5 µL of silicon oil (Sigma Aldrich) was added, and the centrifugation step was repeated. The

samples were measured with DynaPro plate reader III (Wyatt Technology) at 25 °C. For each well, 10 acquisitions of 5 s were collected and processed using the Dynamics V7.8 (Wyatt Technology) software. Cumulant analysis was used to derive the polydispersity index (PDI) and apparent coefficient of self-diffusion (*D*). The apparent hydrodynamic radius (R_h) was calculated from the Stokes-Einstein equation using the measured *D* from DLS and sample viscosity. The viscosity of samples, needed for R_h calculations, was measured by a falling ball viscometer, AMVn (Anton Paar GmbH) in triplicates.

Size exclusion chromatography

A 50 µL sample of each formulation was injected onto a Superdex 200 10/300GL column (GE Life Sciences) attached to a Dionex Summit 2 system (ThermoFisher). The elution of the protein was detected at 240 nm. Running buffer contained 50 mM sodium phosphate buffer, pH 7.4 with 500 mM NaCl. The chromatograms were integrated with Chromeleon V7 (ThermoFisher) and the relative area of the monomer, low molecular species (LMW) i.e. fragments, and high molecular species (HMW) i.e. small soluble aggregates was calculated. All measurements were done in triplicates.

4.5.5 NMR experiments

Samples for NMR were prepared by addition of 5% v/v ^2H_2O to 300 µL of each protein formulation and transferred to 3 mm NMR tubes (Wilmad). All NMR spectra were acquired at 25 °C on 800 MHz Bruker Avance III spectrometer equipped with 5 mm triple resonance TCI cryoprobe and temperature control unit. The spectra were acquired and processed using Bruker Topspin 3.5 (Bruker). Additional analysis was done in OriginPro9.1 (OriginLabs) and NMRFAM-Sparky[342]. Assignment of excipient signals was achieved using a combination of ^1H-^{13}C HSQC and ^1H-^{13}C HSQC-TOCSY experiments, acquired on the reference samples of excipients.

For each sample, a set of 1D ^1H, STD, WaterLOGSY and ^1H T_2 CPMG filter experiments were acquired. ^1H spectra were acquired using a standard Bruker pulse sequence with excitation sculpting with gradients for water suppression (zgesgp) with 64 scans. For STD NMR studies a standard Bruker stddiffesgp.3 pulse sequence was used with the interleaved acquisition of on- and off-resonance spectra with 32 scans. On- and off-resonance saturation frequencies were 0.175 ppm and 20 ppm, respectively. Saturation time

of 3 s was used, and 20 ms spin lock filter was applied to eliminate protein signals. STD spectra were obtained by subtracting on-resonance from off-resonance spectrum. Appropriate control experiments without addition of proteins were performed to confirm no direct irradiation of excipients. STD amplification factors were calculated using equation (Eq **4.2)** where AF^{STD} is the STD amplification factor, I_{off} the intensity of excipient signal in the off-resonance spectrum, I_{on} the intensity of signal in the on resonance spectrum, and *[L]* and *[P]* represent total concentrations of excipients and mAbs, respectively.

$$AF^{STD} = \frac{I_{off} - I_{on}}{I_{off}} * \frac{[L]}{[P]} \qquad \text{(Eq 4.2)}$$

For WaterLOGSY experiments, ephogsygpno.2 pulse sequence from standard Bruker library was used with mixing time of 3s and 64 scans acquired. To determine the reference spectra, appropriate control experiments of each excipient in the absence of protein were acquired. Changes in transverse relaxation rates of protons were monitored using two-point Carr-Purcell-Meiboom-Gill (CPMG) experiments, with the number of echoes 1 and 128. The fixed echo time to allow elimination of diffusion and J-modulation effect was 1.4 ms. Temperature compensation was used to ensure equal heating of the sample over the course of the experiment. All spectra were recorded with 16 scans. Ratio between signal intensities in both spectra was calculated and compared to reference intensities of excipients alone, without the presence of protein.

NMR-derived excipient-protein interaction parameter I^N

Effects of excipients in each type of the experiments were normalised by the maximum measured effect and summed across all four experiments to obtain the overall excipient-protein interaction parameter I^N using (Eq 4.3), where AF^{STD} is the STD amplification factor, AF^{STD}_{max} is the maximum STD amplification factor across all of the measurements, $\frac{I^W}{I^W_{REF}}$, $\frac{I^H}{I^H_{REF}}$, and $\frac{I^C}{I^C_{REF}}$ ratios between excipients signals observed in WaterLOGSY, ^{1}H and CPMG experiments, respectively, in the presence of protein ($I^{W,H,C}$) and its absence ($I^{W,H,C}_{REF}$), while $\left(\frac{I^{W,H,C}}{I^{W,H,C}_{REF}}\right)_{max}$ are the maximum observed ratios among all excipients.

$$I^N = 0.25 * \frac{AF^{STD}}{AF^{STD}_{max}} + 0.25 * \left[1 - \frac{\left(\frac{I^W}{I^W_{REF}}\right)}{\left(\frac{I^W}{I^W_{REF}}\right)_{max}}\right] + 0.25 * \left[1 - \frac{\left(\frac{I^H}{I^H_{REF}}\right)}{\left(\frac{I^H}{I^H_{REF}}\right)_{max}}\right] + 0.25 * \left[1 - \frac{\left(\frac{I^C}{I^C_{REF}}\right)}{\left(\frac{I^C}{I^C_{REF}}\right)_{max}}\right] \quad \text{(Eq 4.3)}$$

The values of the NMR derived protein-excipient interaction parameter I^N therefore vary between 0 and 1, where 0 means the excipient does not interact with the protein, and 1 denotes the strongest interaction among the tested combinations. An arbitrary threshold of 0.25 was chosen as a cut-off, where excipients, scoring lower than 0.25 were determined as false positives.

4.6 Acknowledgments

We would like to thank Jack Bramham for his help with the automation of data analysis. We would also like to thank Dr Matthew Cliff from the NMR Facility in Manchester Institute of Biotechnology for technical support with NMR spectrometers. We are especially grateful to Prof. Dr. Gerhard Winter for hosting M.Z. at his laboratory and providing access to his research facilities at LMU. This work was supported by EU Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement No 675074.

4.7 Supplementary information

4.7.1 Supplementary Figures

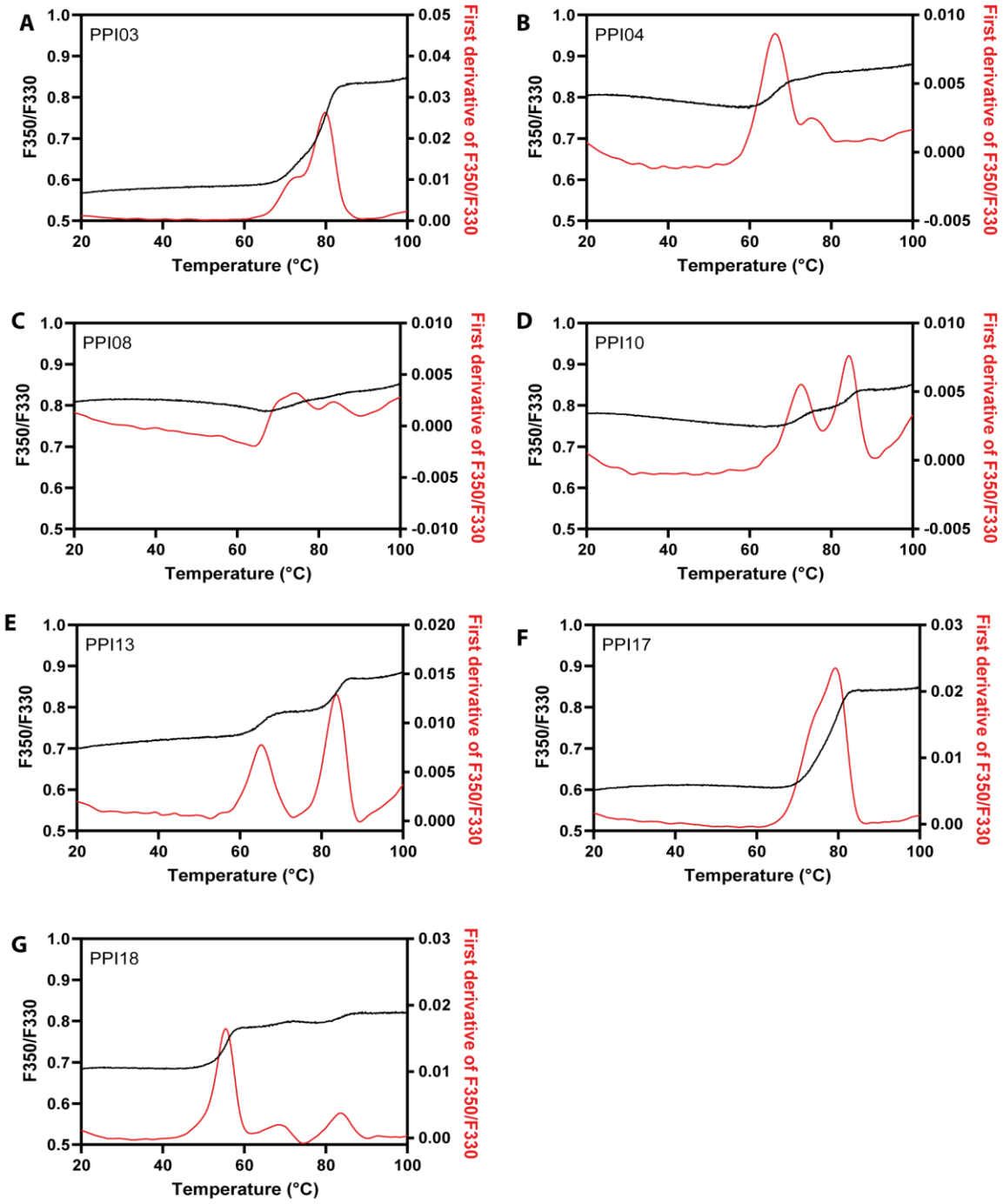

Figure S4.1: Thermal protein unfolding. A) PPI03, B) PPI04, C) PPI08, D) PPI10, E) PPI13, F) PPI17, G) PPI18. Raw data is shown in black, while the first derivative from which the apparent melting temperatures were determined are shown in red.

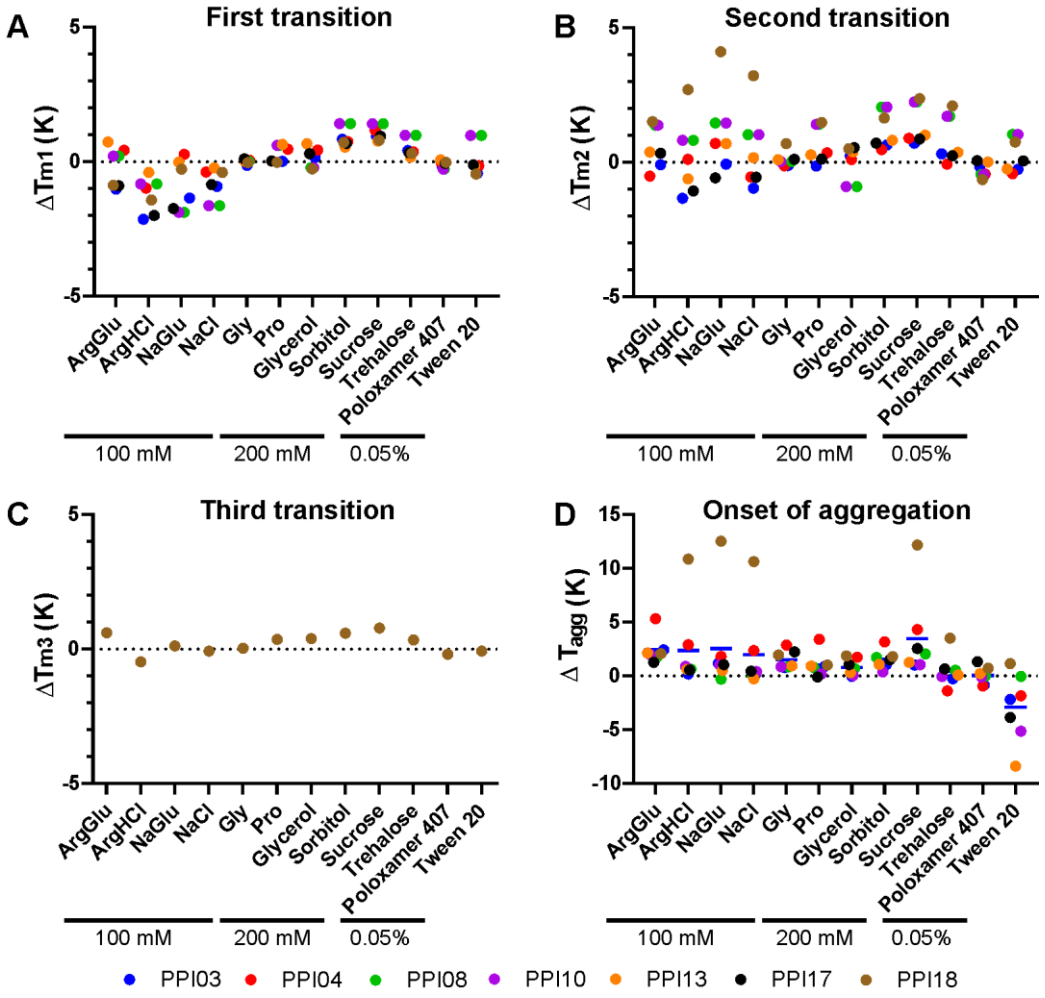

Figure S4.2: Effect of excipients on thermal and colloidal stability of proteins from the dataset. Thermal stability was represented by melting temperatures for the A) first T_{m1}, B) second T_{m2} and C) third T_{m3} melting transitions, and the D) aggregation onset temperature (T_{agg}). The changes of the temperature are shown relative to the protein in buffer alone. The blue horizontal lines show the average effect of each excipient.

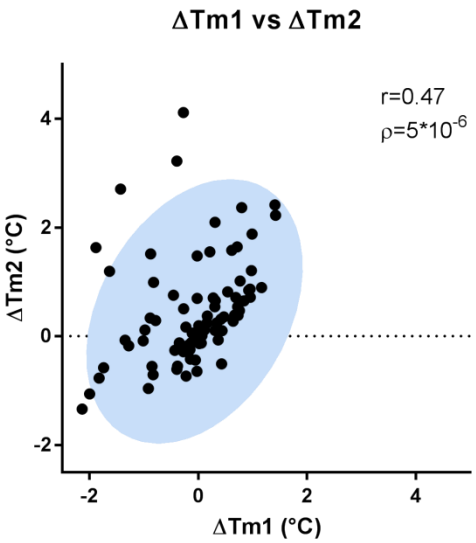

Figure S4.3: Correlation between first and second melting transitions. Data points are shown in black circles, while 95% confidence ellipse obtained during the calculation of Spearman correlation coefficient (r) is shown in blue

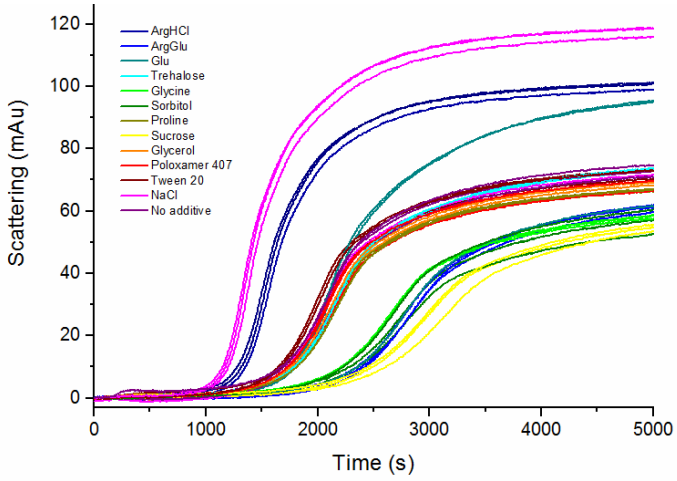

Figure S4.4: Isothermal aggregation of PPI03 in the presence of excipients at 70°C. Data was run in triplicates.

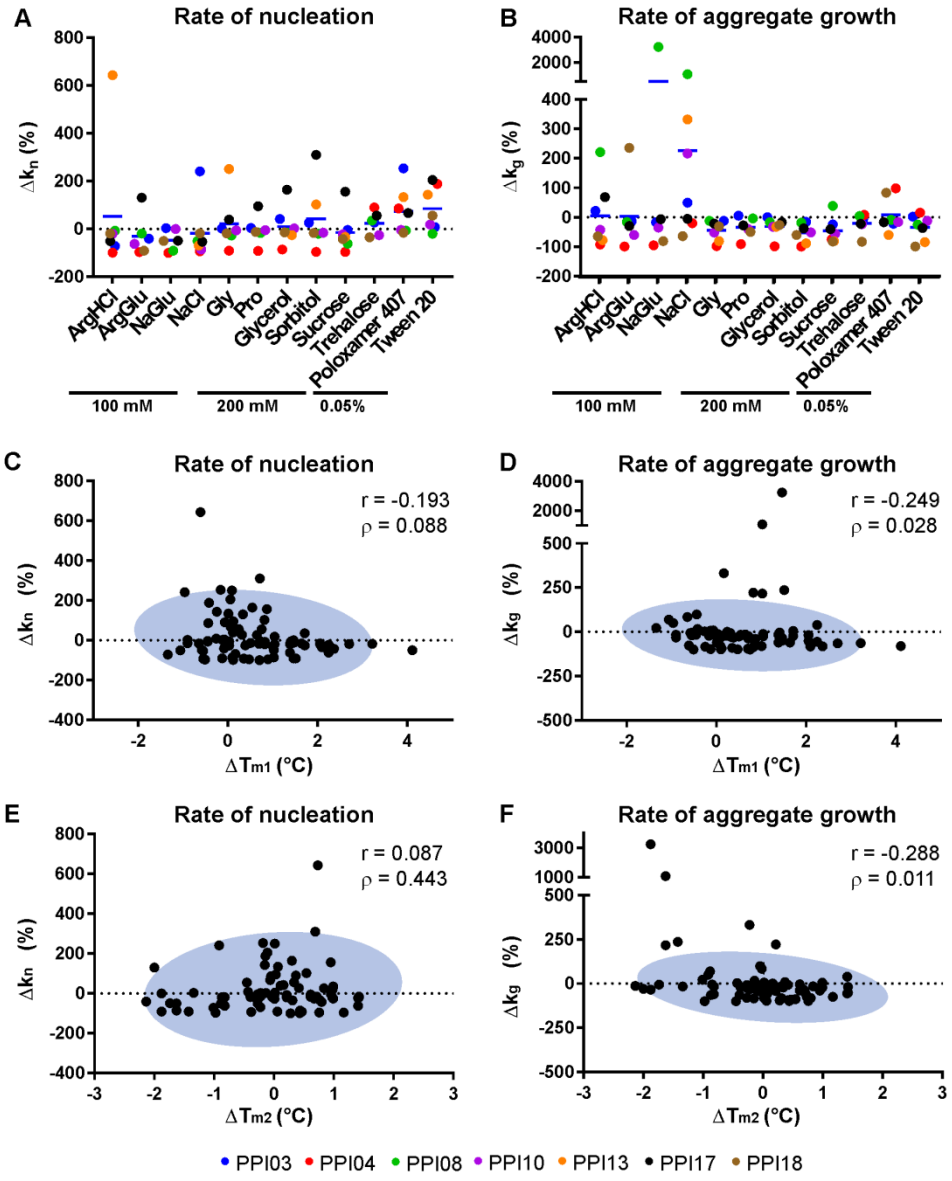

Figure S4.5: Effect of excipients on aggregation kinetics at elevated temperatures. A) Effect of excipients on nucleation rate (k_{nuc}) as compared to the control experiment in the absence of any excipients. B) Effect of excipients on the rate of aggregate growth (k_{agg}) as compared to the control experiment in the absence of excipients. C) Correlation between the effect of excipients on k_{nuc} and first melting transition (T_{m1}) D) Correlation between the effect of excipients on k_{agg} and first melting transition. E) Correlation between the effect of excipients on k_{nuc} and second melting transition (T_{m2}) F) Correlation between the effect of excipients on k_{agg} and T_{m2}

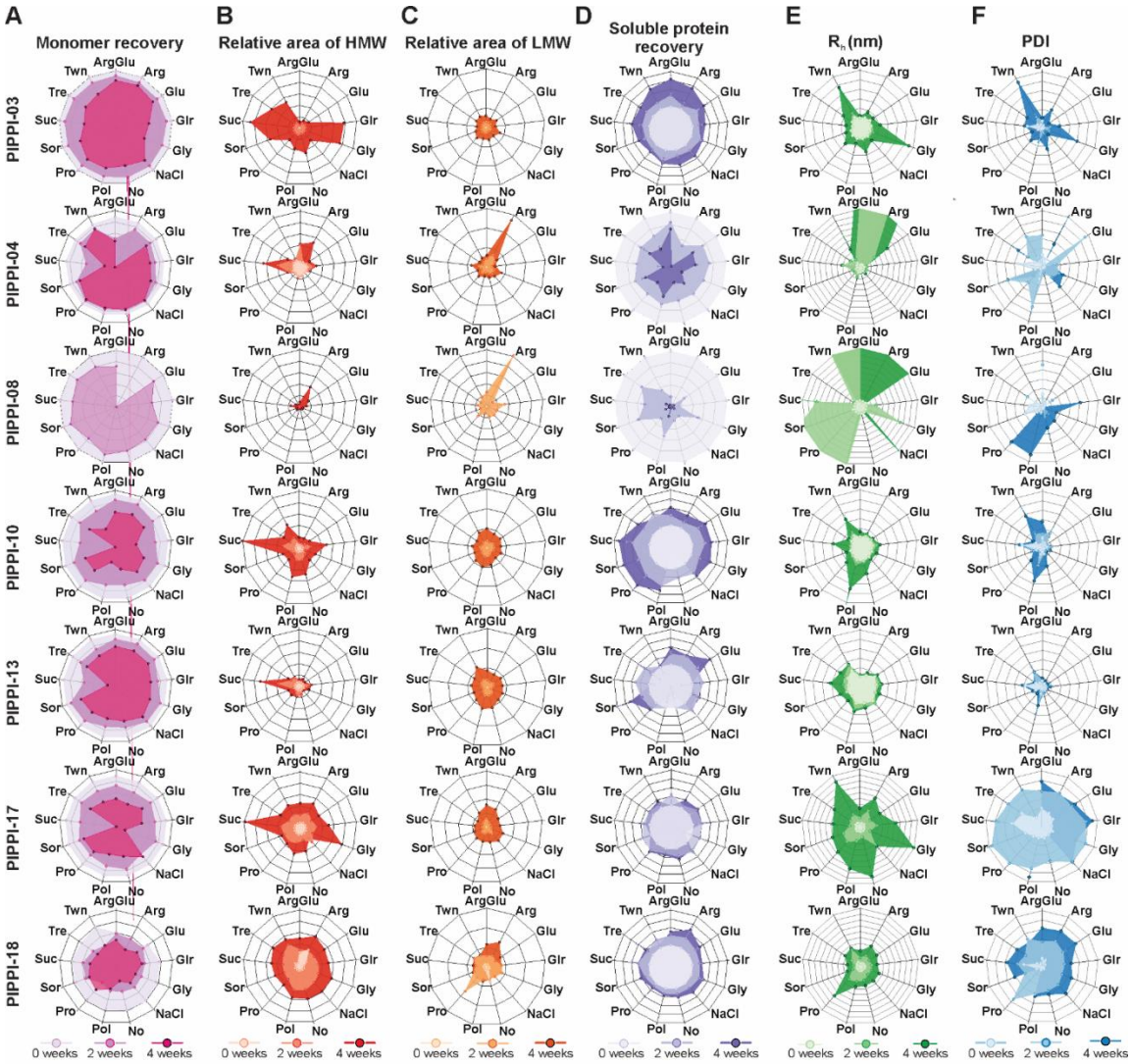

Figure S4.6: Radar chart analysis of the accelerated stability study. Overall stability of proteins in accelerated stability was analysed by A) Monomer recovery during storage time B) fraction of higher molecular species (HMW), C) fraction of fragmentation (LMW) D) total soluble protein recovery as measured by SEC and in terms of E) apparent hydrodynamic radius and F) polydispersity index (PDI) determined from DLS data. Each axis in the individual radar chart represents one of the experimental conditions

4.7.2 Supplementary Tables

Table S 4.1: NMR characterisation of protein-excipient interactions.

Protein	Excipient	S	W	H	C
PPI03	ArgHCl	96.21	0.57	0.76	1.00
PPI03	ArgGlu	0.00	0.47	0.70	0.97
PPI03	NaGlu	176.14	0.68	0.81	1.16
PPI03	Trehalose	14.23	0.92	1.03	1.02
PPI03	Gly	0.00	0.30	0.53	1.04
PPI03	Sorbitol	0.00	1.00	1.19	1.05
PPI03	Pro	7.55	0.00	2.09	0.98
PPI03	Sucrose	241.11			1.16
PPI03	Glycerol	0.00	1.66	2.46	1.12
PPI03	Poloxamer 407	0.00	1.00	1.00	1.00
PPI03	Tween 20	363.68	0.18	0.26	0.83
PPI04	ArgHCl	393.82	0.54	0.52	0.97
PPI04	ArgGlu	421.03	0.48	0.67	1.00
PPI04	NaGlu	0.00	0.36	0.45	1.15
PPI04	Trehalose	224.58	0.67	1.05	1.19
PPI04	Gly	0.00	0.14	0.54	1.03
PPI04	Sorbitol	0.00		1.17	1.05
PPI04	Pro	324.58	1.12	1.64	1.01
PPI04	Sucrose	0.00		1.90	1.15
PPI04	Glycerol	0.00	0.76	2.65	1.07
PPI04	Poloxamer 407	0.00	0.76	0.96	0.96
PPI04	Tween 20	187.35	0.18	0.17	1.06
PPI08	ArgHCl	462.59	0.52	0.58	1.09
PPI08	ArgGlu	582.29	0.45	0.67	1.00
PPI08	NaGlu	521.66	0.54	0.64	1.28
PPI08	Trehalose	311.57	0.49	0.75	0.97
PPI08	Gly	0.00	0.13	0.23	0.94
PPI08	Sorbitol	0.00	0.74	1.17	1.05
PPI08	Pro	133.51	0.65	1.69	1.00
PPI08	Sucrose	0.00	0.25	0.70	1.06
PPI08	Glycerol	0.00	0.68	1.83	1.19
PPI08	Poloxamer 407	1460.6	0.75	0.65	1.69
PPI08	Tween 20	196.80	0.06	0.15	1.23
PPI10	ArgHCl	13.02	0.35	0.63	0.96
PPI10	ArgGlu	40.55	0.31	0.68	0.95
PPI10	NaGlu	165.20	0.36	0.71	1.06
PPI10	Trehalose	25.90	0.23	0.61	1.12
PPI10	Gly	0.00	0.09	0.28	1.03
PPI10	Sorbitol	563.31	0.74	0.89	1.03
PPI10	Pro	0.00	0.80	1.73	1.07

PPI10	Sucrose	0.00	0.42	2.88	1.74
PPI10	Glycerol	584.26	0.80	1.94	1.20
PPI10	Poloxamer 407	150.39	0.66	0.62	1.27
PPI10	Tween 20	153.18	0.09	0.16	0.79
PPI 13	ArgHCl	577.47	0.36	0.76	0.88
PPI 13	ArgGlu	61.29	0.30	0.69	1.00
PPI 13	NaGlu	0.00	0.36	0.68	1.05
PPI 13	Trehalose	302.73	0.35	0.71	1.59
PPI 13	Gly	0.00	0.08	0.27	1.00
PPI 13	Sorbitol	0.00	0.91	0.63	1.05
PPI 13	Pro	0.00	0.60	1.67	1.09
PPI13	Sucrose	0.00	0.30	0.75	
PPI 13	Glycerol	37.84	0.85	1.95	1.15
PPI 13	Poloxamer 407	254.89	0.59	0.66	2.15
PPI 13	Tween 20	58.05	0.06	0.17	0.91
PPI 17	ArgHCl	567.95	0.31	0.66	0.92
PPI 17	ArgGlu	203.04	0.27	0.66	0.99
PPI 17	NaGlu	428.38	0.35	0.71	1.11
PPI 17	Trehalose	0.00	0.35	0.73	1.00
PPI 17	Gly	0.00	0.08	0.31	1.03
PPI 17	Sorbitol	0.00	0.92	0.59	1.06
PPI 17	Pro	0.00	0.65	1.64	1.02
PPI 17	Sucrose	0.00		1.03	1.51
PPI 17	Glycerol	279.96	0.86	2.01	1.26
PPI 17	Poloxamer 407	63.86	0.68	0.62	1.26
PPI 17	Tween 20	9.05	0.02	0.17	1.04
PPI 18	ArgHCl	90.29	0.28	0.64	0.98
PPI 18	ArgGlu	78.68	0.29	0.68	1.02
PPI 18	NaGlu	0.00	0.33	0.64	0.95
PPI 18	Trehalose	41.40	0.33	0.71	0.89
PPI 18	Gly	0.00	0.06	0.23	1.03
PPI 18	Sorbitol	0.00	1.18	0.90	1.05
PPI 18	Pro	939.02	0.62	1.76	1.39
PPI 18	Sucrose	0.00		1.64	1.33
PPI 18	Glycerol	0.00	0.82	2.06	1.10
PPI 18	Poloxamer 407	278.31	0.63	0.63	2.35
PPI 18	Tween 20	557.08	0.01	0.18	1.56

Table S4.2: Evaluation of thermal and colloidal stability of proteins in the presence of excipients

	PPI03					PPI04				
Excipient	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**
ArgHCl	77.07	70.52	-6.55	78.33	1.27	66.33	66.75	0.43	75.57	9.24
ArgGlu	79.32	71.64	-7.68	79.59	0.27	63.90	65.35	1.45	76.20	12.30
NaGlu	78.08	71.31	-6.77	79.60	1.52	62.82	66.60	3.78	76.78	13.96
NaCl	76.98	71.74	-5.24	78.71	1.73	63.36	65.94	2.58	75.53	12.17
Glycine	77.68	72.52	-5.16	79.55	1.87	60.99	66.32	5.33	76.08	15.09
Proline	77.79	72.67	-5.12	79.54	1.75	63.85	66.38	2.53	75.95	12.10
No additive	76.48	72.65	-3.83	79.67	3.19	64.41	66.79	2.38	76.43	12.02
Glycerol	76.86	72.81	-4.05	79.92	3.05	62.72	66.76	4.04	76.18	13.46
Sorbitol	77.96	73.49	-4.47	80.32	2.36	64.17	67.08	2.91	76.55	12.39
Sucrose	77.91	73.61	-4.30	80.39	2.48	65.31	67.49	2.18	76.98	11.67
Trehalose	76.63	73.07	-3.56	79.99	3.36	59.61	66.69	7.08	76.01	16.40
Poloxamer 407	76.05	72.47	-3.58	79.51	3.46	60.06	66.27	6.22	75.64	15.58
Tween 20	74.72	72.22	-2.50	79.41	4.69	59.15	66.18	7.03	75.66	16.51
	PPI08					**PPI10**				
Excipient	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**
ArgHCl	64.33	73.03	8.70	83.70	19.37	82.18	71.09	-11.10	83.52	1.33
ArgGlu	63.15	71.99	8.84	83.14	19.99	81.07	70.05	-11.02	82.46	1.39
NaGlu	62.21	70.94	8.72	83.78	21.57	81.22	70.59	-10.63	83.05	1.83
NaCl	62.62	71.19	8.57	83.34	20.72	80.60	71.04	-9.56	82.52	1.92
Glycine	62.52	72.82	10.29	82.15	19.63	80.19	71.87	-8.32	83.23	3.04
Proline	63.36	72.83	9.46	82.35	18.98	81.05	71.88	-9.17	83.27	2.23
No additive	63.31	73.43	10.12	83.73	20.42	80.37	71.91	-8.45	83.38	3.02
Glycerol	63.23	72.59	9.36	81.42	18.19	80.23	72.18	-8.05	83.88	3.65
Sorbitol	64.25	74.24	9.99	84.37	20.12	80.57	72.59	-7.97	83.77	3.20
Sucrose	64.57	74.22	9.66	84.57	20.00	81.24	72.80	-8.44	84.07	2.84
Trehalose	63.06	73.81	10.74	84.03	20.97	80.13	72.18	-7.95	83.34	3.20
Poloxamer 407	62.53	72.54	10.02	81.87	19.34	80.09	71.77	-8.33	83.28	3.19
Tween 20	62.47	73.79	11.32	83.36	20.88	75.04	71.52	-3.52	83.11	8.06

	PPI13					PPI17				
Excipient	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**
ArgHCl	82.20	64.99	-17.20	83.38	1.18	74.21	73.32	-0.89	79.56	5.35
ArgGlu	80.67	64.00	-16.67	82.25	1.58	73.51	72.21	-1.30	78.16	4.65
NaGlu	80.57	65.31	-15.27	82.63	2.05	73.99	72.47	-1.53	78.64	4.65
NaCl	79.80	64.78	-15.02	82.42	2.62	73.39	73.36	-0.03	78.67	5.28
Glycine	80.06	64.61	-15.45	82.64	2.58	72.95	74.21	1.26	79.22	6.27
Proline	81.00	64.70	-16.30	82.66	1.66	75.20	74.32	-0.88	79.34	4.14
No additive	80.99	64.89	-16.11	83.29	2.29	72.86	74.24	1.38	79.34	6.48
Glycerol	80.36	65.00	-15.36	83.31	2.95	73.99	74.51	0.52	79.77	5.77
Sorbitol	81.14	65.43	-15.71	83.19	2.05	74.45	74.90	0.45	79.94	5.49
Sucrose	81.32	65.62	-15.69	83.41	2.10	75.50	75.16	-0.34	80.09	4.59
Trehalose	80.16	64.98	-15.18	82.81	2.65	73.60	74.55	0.95	79.47	5.87
Poloxamer 407	80.28	64.62	-15.66	82.71	2.43	74.27	74.15	-0.13	79.28	5.01
Tween 20	71.68	64.36	-7.32	82.49	10.81	69.10	74.10	5.00	79.28	10.18
	PPI18									
Excipient	**T_{agg}**	**T_{m1}**	**T_{m1}-T_{agg}**	**T_{m2}**	**T_{m2}-T_{agg}**	**T_{m3}**	**T_{m3}-T_{agg}**			
ArgHCl	56.50	53.81	-2.69	68.74	12.24	83.24	26.74			
ArgGlu	65.28	53.26	-12.02	69.94	4.65	82.16	16.88			
NaGlu	66.94	54.41	-12.53	71.34	4.40	82.75	15.80			
NaCl	65.06	54.29	-10.77	70.45	5.39	82.56	17.50			
Glycine	54.43	54.68	0.26	67.23	12.80	82.63	28.20			
Proline	56.36	54.65	-1.71	67.93	11.56	82.66	26.30			
No additive	55.42	54.66	-0.76	68.71	13.28	82.99	27.57			
Glycerol	56.28	54.42	-1.86	67.73	11.45	83.03	26.75			
Sorbitol	56.23	55.40	-0.83	68.87	12.65	83.22	26.99			
Sucrose	66.60	55.48	-11.12	69.60	3.00	83.41	16.81			
Trehalose	57.94	54.99	-2.95	69.33	11.39	82.98	25.04			
Poloxamer 407	55.15	54.66	-0.49	66.58	11.44	82.44	27.29			
Tween 20	55.57	54.23	-1.34	67.99	12.42	82.56	26.99			

BLANK PAGE

5 Use of protein- and ligand-observed NMR to comprehensively study protein-excipient interactions

This chapter was written for peer review as:

Matja Zalar[1], Alexander P. Golovanov[1*],

[1]Manchester Institute of Biotechnology and Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

* Correspondence to: A.Golovanov@manchester.ac.uk

Author contributions:

M.Z. prepared the samples, performed the experiments, analysed the data and wrote the first draft of the manuscript. A.P.G. provided conceptual guidance and made significant contributions to the final draft of the manuscript.

5.1 Abstract

Formulation of protein solutions, including the addition of small molecule excipients, such as amino acids, sugars, sugar polyols and detergents, is vital for stability and development of biopharmaceutical products. However, this typically requires extensive screenings as the mechanisms and interactions which underly protein stabilisation with excipients are not clear. As protein-excipient interactions are weak and transient they are challenging to characterise and are traditionally assessed using non-direct methods such as vapour pressure osmometry and densimetry. One of the methods that have only recently been applied to characterise these weak interactions is NMR. In this study, we have used both ligand- and protein-observed NMR methods to characterise interactions between a model protein SQT-1C$^{Q46C,\ N59C}$ and eleven common excipients. We show that ligand-observed methods have limited applicability at excipient concentrations typically used in biopharmaceutical formulation. However, ^{15}N and ^{13}C HSQC proved to be a sensitive tool in identifying transient excipient interaction sites on the protein surface, allow for distinguishing between preferential exclusion and interaction mechanism of action and even provide estimates of dissociation constants.

5.2 Introduction

In biotechnology and biopharmaceutical industries, stabilisation of proteins against chemical and physical instabilities is a critical quality attribute. During production and storage, proteins may undergo various stresses, leading to aggregation, fragmentation, liquid-liquid phase separation, and other forms of degradation[388]. Protein aggregation is one of the biggest challenges in protein formulation, particularly at high concentrations. Aggregation may be minimised through protein engineering approaches or optimisation of formulation conditions, including buffers, pH, ionic strength, and addition of small molecule excipients[389].

Excipients, such as amino acids, sugars, and surfactants, are added to protein formulations to solubilise the active ingredient, stabilise it and reduce solution viscosity. This typically requires relatively high excipient concentrations, in the 10s-100s mM range and in large excess compared to the active ingredient[143]. While there is a clear understanding that the addition of excipients may improve the stability of the active ingredient, their mechanisms of action are largely unknown. In general, excipients are thought to either

undergo preferential interaction with the protein surface, or preferential exclusion from the surface[135-137, 147, 360]. However, given the weak and transient nature of these processes[209], lack of distinct binding sites and defined stoichiometry, studying protein-excipient interactions is challenging.

Protein-excipient interaction (exclusion) parameters are typically measured using densimetry[137, 390], vapour pressure osmometry[148, 391], microscale thermophoresis[357], or through thermal shift assays[151], where it is assumed that the change in a measurable parameter, e.g., the extent of temperature shift, is proportional to the affinity of the ligand for a given protein[392]. Such methods are indirect and do not provide much detail into protein-excipient interactions and often rely on complicated models with a lot of assumptions. Standard biophysical techniques used to study protein-ligand interactions, such as isothermal calorimetry (ITC), surface plasmon resonance (SPR), and fluorescence, are best suited to detect strong specific interactions in the nano- to micromolar range[366-367, 393], and require a knowledge of the number of binding sites. Therefore, such techniques are not easily applicable to characterising protein-excipient interactions. Conversely, nuclear magnetic resonance (NMR) spectroscopy is widely used to study weaker protein-ligand interactions in the low millimolar range, for example, in fragment-based drug discovery (FBDD)[394]. Additionally, NMR can report of protein structure and the effect of ligand binding on this structure or changes in the local environment at the transient binding sites. Therefore, it may be suitable for studying both protein-excipient interactions and their mechanism of action, i.e., of protein stabilisation.

NMR may either examine ligand signals (ligand-observed) or protein signals (protein-observed) to report on protein-ligand interactions. So far, efforts to characterise these interactions were predominantly based on ligand-observed experiments in solution[215-216, 395] and solid state[396]. Previously we attempted to quantify the weak protein-excipient interactions for non-labelled proteins using a weighed empirical interaction parameter based on ligand-observed 1D NMR methods, and showed that it poorly correlates with other measures of protein stability[397]. Furthermore, ligand-observed experiments do not provide the in-depth structural or site-specific information that is necessary to understand the mechanism of action of individual excipients and provide guidance which excipient might be best for certain proteins. Therefore, protein-observed experiments, which may report on both interactions and the influence of ligands on protein structure or local environment, may offer a greater understanding of the mechanisms of protein-excipient interactions.

Previously, ^{15}N HSQC NMR fingerprinting was used to assess conformation of interferon-alpha in various biopharmaceutical formulations[398] while the protein-observe strategy has been applied to explore the binding of Hofmeister series ions to model protein and peptide surfaces[117, 399-401].

In this study, we have explored how commonly-used NMR experiments can be applied to probe protein-excipient interactions in biopharmaceutical formulations. We used a model protein, SQT-1C$^{Q46C, N59C}$ [402] and systematically applied both ligand and protein observed approaches to compare the two and see what their advantages and drawbacks are and whether they correlate. We show that while ligand-observed methods can detect protein-excipient interactions in our system at lower concentrations, they have limited applicability at concentrations of excipients typically used in biopharmaceutical formulations. We also show that both ^{15}N and ^{13}C HSQC provide a sensitive tool to identify transient interaction sites on the protein surface. ^{15}N additionally allowed us to distinguish between preferential interaction and exclusion mechanism of individual excipients, while binding constants could be estimated using ^{13}C HSQC spectra.

5.3 Results

5.3.1 Choice of the protein model and excipients

In this study, we used ^{15}N and ^{13}C labelled SQT-1C$^{Q46C, N59C}$, an engineered scaffold protein derived from stefin A[56] and stabilised against domain swap oligomerisation[402] as a model to probe potential protein-excipient interactions. Engineered scaffold proteins can be designed to bind to arbitrary targets, similar to antibodies but typically have a much smaller frame (often ~15 kDa), less complex structure and can be expressed in simpler recombinant expression systems[266] which makes them good model systems to explore protein-excipient interactions. Backbone assignment of SQT-1C$^{Q46C, N59C}$ has been previously reported[331, 402] and is 80% complete. The unassigned peaks are not visible in the ^{15}N HSQC spectra and involve five proline residues, N-terminal residues (M1-R4), and parts of the loop regions (L49-R52 and F77-L84).

Here, we have tested a set of eleven biopharmaceuticaly relevant excipients including arginine hydrochloride (ArgHCl), arginine glutamate (ArgGlu), sodium glutamate (NaGlu), glycine (Gly), proline (Pro), glycerol, sorbitol, trehalose, sucrose, poloxamer 407 and tween 20 against SQT-1C$^{Q46C, N59C}$. Using ^{15}N and ^{13}C isotopically labelled protein allowed us to

compare systematically the results of ligand-observed and protein-observed experiments on the same samples.

5.3.2 Ligand-observed screening depends on the protein to excipient ratio

Initially, protein-excipient interactions were assessed using ligand-observed experiments as they are the simplest and fastest type of NMR experiments to detect protein-ligand binding, at least for tighter-binding ligands. Additionally, they do not require any labelling or specific sample preparation. The results of such screening can, however, depend on the protein to excipient ratio which is typically very high in biopharmaceutical formulations, with active protein ingredient used at <1 mM concentration, and excipients at 10s-100s mM. We have, therefore, applied a previously used combination of four ligand-observed experiments[397], namely 1D ^{1}H, transverse T_2 relaxation filter (CPMG filter), saturation transfer difference (STD) and WaterLOGSY spectra[397] to detect excipient interactions with SQT-1C$^{Q46C,\ N59C}$ and explore how these methods perform at different protein to excipient (P:L) ratios. Results of the hit identification screen at different protein to ligand ratios are summarised in Figure 5.1.

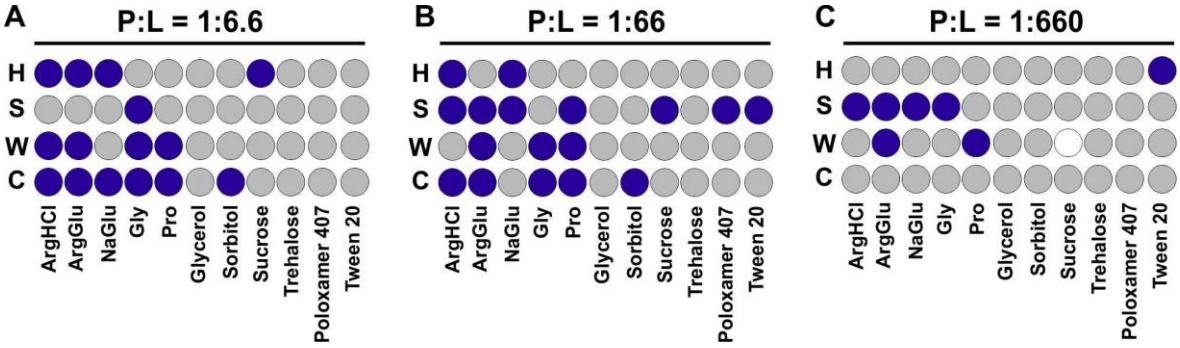

Figure 5.1: Screening of excipient interactions with SQT-1C$^{Q46C,\ N59C}$ by ligand-observed NMR. Qualitative evaluation of protein excipient interactions as detected by ^{1}H (H), STD (S), WaterLOGSY (W) and CPMG (C) NMR methods and different protein to ligand ratios. Ratios used were A) 1:6.6, B) 1:66 and C) 1:660. Positive hits for excipient binding identified by the individual method are denoted in blue, non-hits in grey while data points excluded from analysis due to experimental error are indicated with empty circles.

Our results show that ligand-observed screening depends heavily on the protein to ligand ratio used in the experiment which is in line with previous reports on these methods[256,

[403-404]. We observed that line broadening, WaterLOGSY, and CPMG filter perform better at a lower concentration of excipients (P:L=1:6.6), which is to be expected as the binding effects are not masked by the bulk of the free excipients. On the other hand, STD performed better at higher concentrations of excipients (P:L=1:66), where other methods already failed to detect the interactions. At excipient concentrations (P:L=1:660) that are typically used in biopharmaceutical formulations, the ligand-observed methods performed very poorly, with only a small fraction of interactions identified typically by a single method. It is still important to mention that protein-excipient interactions can be detected to some extent using ligand-observed methods even when excipient is in more than 50-fold excess of the protein which is typically the case in biopharmaceutical formulations. Such ligand-observed methods can, therefore, provide a useful tool for distinguishing excipients that interact with proteins but have limited applications at excipient concentrations typically used in biopharmaceutical formulations. However, we have shown previously that there is a very poor correlation between the excipient interaction and the overall protein stability in the presence of this excipient[397]. Moreover, ligand observed methods do not provide any information about the number and the location of the interaction sites on the protein. We have therefore explored the application of 2D protein-observed methods, using the same protein model, SQT-1C^{Q46C,N59C}, to detect weak protein-excipient interactions and potentially deduce the physical mode of action of individual excipients.

5.3.3 Perturbation mapping of excipient interaction detected by 2D ^{1}H-^{15}N HSQC experiments

Given that ligand detected experiments proved insensitive at excipients concentrations typically used in biopharmaceutical formulations, we next explored the use of two-dimensional protein-observed experiments to detect weak protein-excipient interactions. Additionally, these experiments may provide information about the number and location of interaction sites on the protein surface, and, potentially, estimates of the binding constants (K_d).

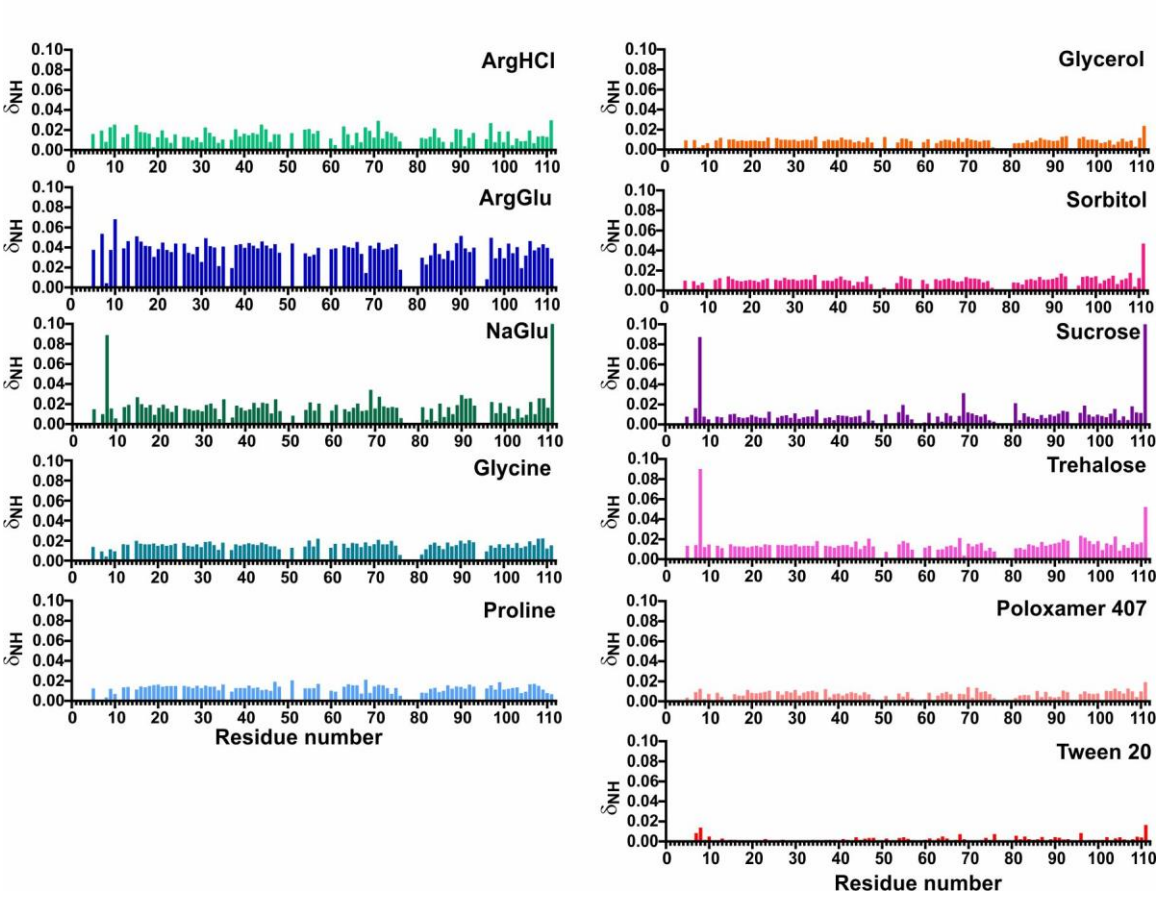

Figure 5.2: Maximum chemical shift perturbation plot. CSP values were calculated as change of chemical shifts in the absence of excipients and at the highest excipient concentrations. (200 mM, except for sucrose at 100 mM).

Initially, we have monitored chemical shift perturbations (CSP) in 2D ^{15}N HSQC across a wide range of excipient concentrations. ^{15}N HSQC spectra report on the structure of the protein backbone, and changes to the local environment of amide groups. The maximum observed CSPs (Figure 5.2) reveal that the protein-excipient interactions occur across the entire protein surface without obvious specific binding sites, which can be perhaps expected. The absolute CSP values are relatively small, reflecting the weak and transient nature of protein-excipient interactions. However, it may be reasonable to suggest that the degree of changes in the local environment, represented by CSP values, may reflect the relative "strength" of the transient interactions: this appears to vary significantly between different excipient classes and also within excipient classes. For example, ArgGlu leads to much greater CSPs than ArgHCl and NaGlu on their own.

Following the dependence of CSP values on increasing excipient concentrations (Figure 5.3, S1- S11) reveals three distinct patterns. Firstly, fast chemical exchange can be

present, where peaks move linearly, and may or may not reach saturation (Figure 5.3A,D). Secondly, step-like behaviour where significant CSP is observed only after a certain concentration of excipient is added to the solution (Figure 5.3B,E). This behaviour imply some cooperativity in local excipient binding. Thirdly, for some sites no significant CSP is observed (Figure 5.3C,F). We attempted to fit the CSP titration isotherms of peaks that moved linearly to a simple 1:1 binding model to estimate the binding constants (K_d) of protein-excipient interactions. However, the meaningful estimation of K_d values was not possible as the CSP did not saturate for any of the excipients. This implies that the nominal K_d values are >100 mM.

The step-like change of CSP at a certain concentration is intriguing. It is also accompanied by the appearance of additional peaks in the 2D ^{15}N HSQC spectrum (Figures S3, S6, S8 and S9). This behaviour was observed for trehalose, sucrose and sorbitol which are known to be preferentially excluded from the protein surface[135-136], and for glutamate. As a sudden step-wise change in protein structure is unlikely (and such change should propagate to several sites, affecting multiple signals), it suggests that such sudden changes are caused by a cooperative change in the local environment, most probably due to a rearrangement of the water molecules indicating preferential exclusion mechanism of some of the excipients. Our results show that 2D ^{15}N HSQC provides a sensitive tool to probe protein-excipient interactions, identify transient interaction sites on the protein surface, and detect changes in the local environment, such as rearrangement of solvent molecules.

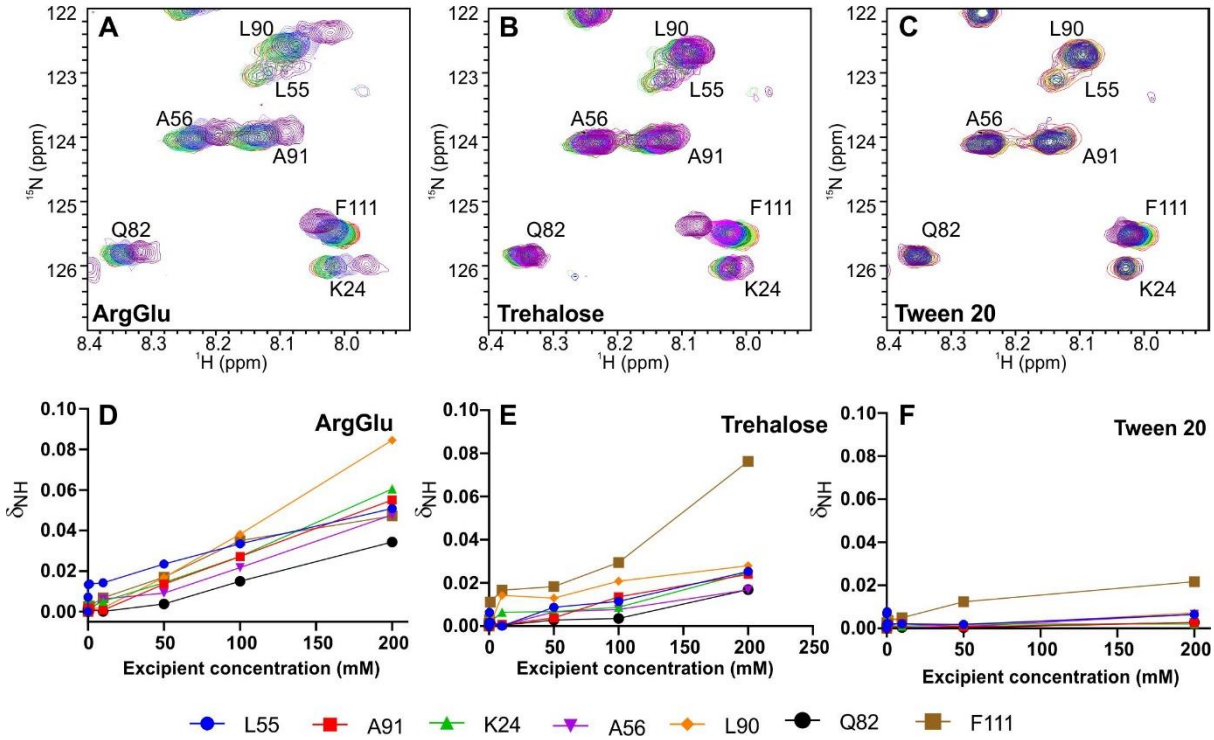

Figure 5.3: Expansion of a region in the ^{1}H-^{15}N HSQC spectrum of SQT-1C$^{Q46C,\ N59C}$ and chemical shift perturbations in the presence of increasing concentrations of A,D) ArgGlu, B,E) Trehalose and C,F) Tween 20.Assignments are show next to the peaks. Concentrations used: 0 (red), 0.075 (orange) 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (pink) and 200 mM (purple). Concentration of SQT-1C$^{Q46C,\ N59C}$ was 150 µM.

5.3.4 Analysis of protein-excipient interaction detected by 2D ^{1}H-^{13}C HSQC experiments

Excipients typically do not bind into a specific binding pocket but are expected to interact with the protein surface. Therefore, the sidechains may be in more contact with excipients than the backbone amide groups, that are usually buried and involved in hydrogen bonds. We have therefore explored whether the signals from the sidechains, represented by the CH correlations in 2D ^{13}C HSQC, are responsive to the excipient interactions.

We have monitored chemical shift perturbations (CSP) in 2D ^{13}C HSQC across a wide range of excipient concentrations (Figure S5.12-S5.22). Assignment of the CH correlations for SQT-1C$^{Q46C,\ N59C}$ was unfortunately not available. Instead, we have arbitrarily numbered the signals and monitored CSP of C_{α}-H, $C_{\beta/\gamma}$-H and CH_3 correlations without reference to the protein sequence.

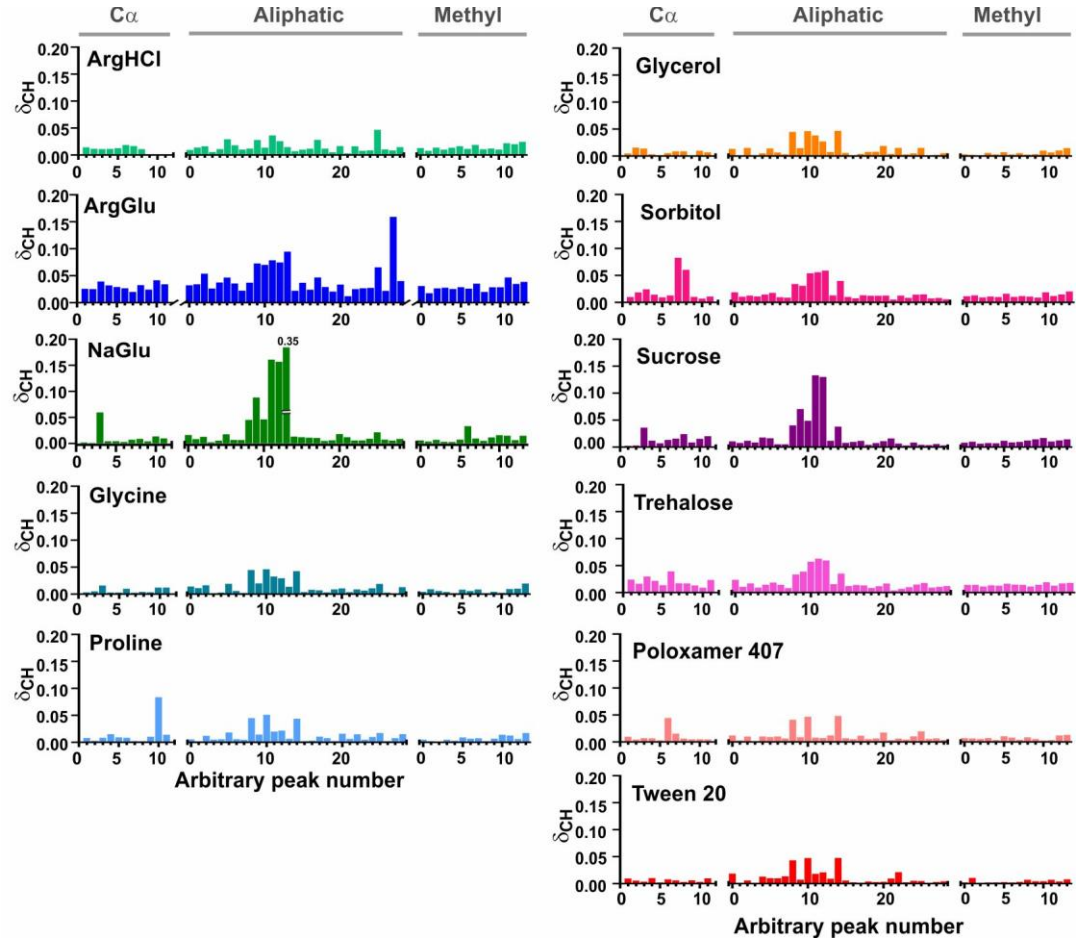

Figure 5.4: Maximum chemical shift perturbation of arbitrary-numbered spectral peaks, detected using ^{13}C HSQC spectra. CSP values were calculated as change of chemical shifts in the absence of excipients and at the highest ±excipient concentrations. (200 mM, except for sucrose at 100 mM).

The maximum observed CSPs (Figure 5.4) reveal that the effects in the ^{13}C HSQC seem to be more localised than in the ^{15}N HSQC, as fewer signals were perturbed. However, these results might be biased due to the limited amount of peaks used in the analysis. Similar to N-H CSP, the majority of absolute C-H CSP values are relatively small, reflecting the weak and transient nature of protein-excipient interactions. More localised observed CSPs suggest that the effects observed in ^{13}C HSQC would make it easier to determine the interaction sites on the protein if the assignments were available. However, the observed CSP patterns were similar to those obtained for the ^{1}H-^{15}N HSQC experiments, in that residues display all three chemical shift patterns (Figure S5.12-Figure S5.22). In contrast to

^{1}H-^{15}N HSQC experiments, saturation of CSP was observed for selected peaks and hence the K_d values could be estimated. For example, the K_d for NaGlu- SQT-1C$^{Q46C,\ N59C}$ interactions under the studied conditions is ~80±20mM (Figure 5.5). Our results indicate that excipients preferentially interact with side chains, which is in line with the common consensus. Additionally, we show that 2D ^{13}C HSQC provides a valuable complement to ^{15}N HSQC experiments to probe protein-excipient interactions, and identify interaction sites on the protein surface, and can be used to estimate K_d values.

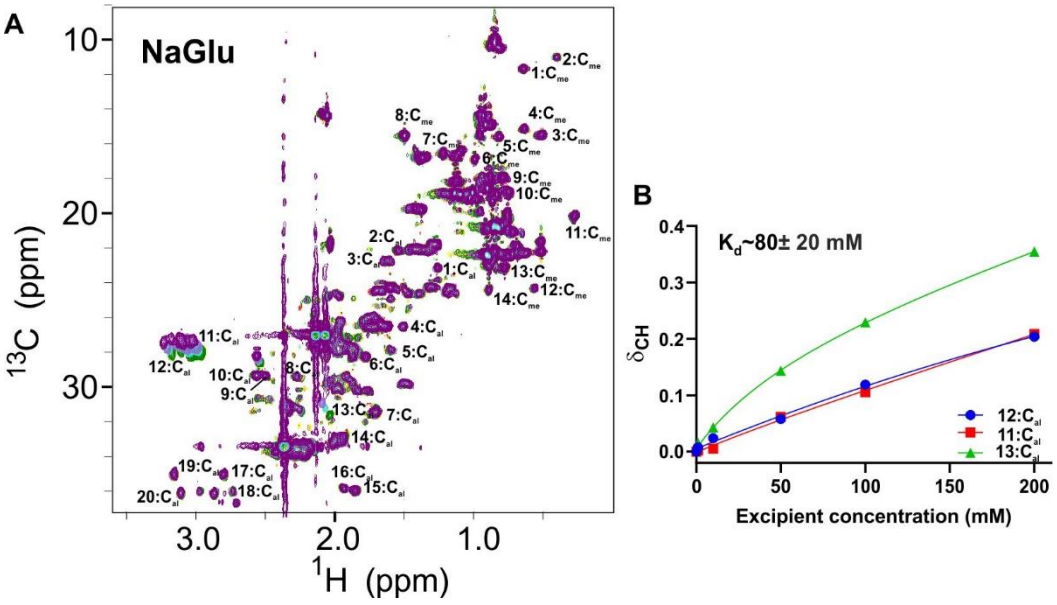

Figure 5.5: A) Expansion of the methyl-aliphatic region in the ^{1}H-^{13}C HSQC spectrum of SQT-1C$^{Q46C,\ N59C}$ in the presence of increasing concentrations of NaGlu Labels with arbitrary numbering are shown next to the peaks. Concentrations used: 0 (red), 0.075 (orange) 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (pink) and 200 mM (purple). Concentration of SQT-1C$^{Q46C,\ N59C}$ was 150 μM. **B) Chemical shift perturbations in the presence of increasing concentrations of NaGlu for selected peaks.** Solid lines represent the best fit to the data based on the 1:1 binding model.

5.4 Discussion

It has been established previously that weak protein-excipient interactions may influence certain protein stability indicators, and there has been a significant effort made towards trying to prevent protein aggregation by identifying an excipient that would bind to aggregation-prone regions thus masking them, and perhaps competitively inhibit protein-protein interactions[151, 357, 405]. Typical methods involved in such screening rely on indirect

measurement of the observable and systematically fail to detect weak protein-excipient interactions.

In this study, we have employed both ligand- and protein- observed NMR approaches to explore the potential of NMR to study weak and transient protein-excipient interactions that cannot be measured directly by any other method. Although side-chain assignment of SQT-1C$^{Q46C, N59C}$ was not available, the shifts of CH signals from the sidechains could still be analysed, for the degree of site-specific changes and shape of the concentration dependencies, for different excipients. Sequence-specific signal assignments were available for the amide signals, allowing sequence-specific analysis of perturbations caused by transient excipient binding. We show that ligand and protein-observed methods identify the same excipients as binders, especially when multiple ligand-observed experiments are combined to eliminate false positives and false negatives. Caution should, however, be taken as the results heavily depend on protein to ligand ratio in solution, making these experiments less reliable at higher concentration of excipients, as the effects are masked by the bulk solution.

^{15}N HSQC proved to be a sensitive tool to probe weak protein-excipient interactions, however it was not possible to estimate the K_d of these interactions due to their weak and transient nature, and possibly due to high salt concentration conditions, which screened some of the electrostatic interactions. We show, however, that such an experiment could be used to deduce whether an excipient acts by a preferential interaction or preferential exclusion mechanism, and determine the concentration of excipients needed to achieve this.

In contrast with ^{15}N HSQC perturbation data, CSP in ^{13}C HSQC seems to be more localised, highlighting the residues that preferentially interact with the excipients. Additionally, saturation of selected peaks was observed, allowing for estimating the K_d values for protein- excipient interactions. However, interference of the water and excipient signals, and signal overlap make ^{13}C HSQC more challenging to analyse. Often, full side-chain assignment is also not possible. Overall, the ^{13}C HSQC chemical shift perturbations offers a nice alternative to a more classical approach of detecting protein-excipient interactions with ^{15}N HSQC. Additionally, it can be envisioned that methyl ^{13}C HSQC at natural abundance could be used in screening protein-excipient interactions in biopharmaceutical formulations, especially where isotopic labelling is not an option or

where the protein is too big to be analysed by ^{15}N HSQC (for example monoclonal antibodies).

In our study, we have observed fairly small CSP values, which is to be expected as excipients interact weakly and transiently with the protein surface and there is no defined pocket for their binding. Additionally, we carried out our experiments at high ionic strengths, similar to the ionic strength typically used in biopharmaceutical formulations, masking the electrostatic interactions. In contrast, the majority of other studies where protein-excipient interactions have been evaluated so far were carried out at low ionic conditions [116-117, 400]. They have subsequently observed much bigger effects in chemical shift perturbation data[117]. It can, therefore, be envisioned that our results may vary significantly if they were to be repeated at low ionic strength.

In conclusion, using isotopically-labelled model proteins enables to test NMR approaches to measuring weak protein-excipient interactions, explore the protein-specific and site-specific mechanisms of excipient action and may be used to train computational models aimed at predicting site-specific interactions for arbitrary protein structure [359, 406] or design of novel excipients [356, 407].

5.5 Materials and Methods

5.5.1 Protein expression, purification and sample preparation

^{13}C and ^{15}N labelled SQT-1C$^{Q46C,\ N59C}$ was expressed in M9 minimal media supplemented with $^{15}NH_4Cl$ and $^{13}C_6$ glucose and purified as described previously [402]. After purification, samples were extensively dialysed into 50 mM sodium phosphate buffer, pH 7.2 with the addition of 150 mM NaCl. The stock solution of SQT-1C^{Q46C,N59C} contained 0.15 mM ^{13}C and ^{15}N labelled SQT-1C^{Q46C,N59C}, 10% v/v ^2H_2O in 50 mM sodium phosphate buffer, pH 7.2, 150 mM NaCl, and 0.0002 % 3-trimethylsilyl-2,2,3,3-($^2H_4$) propionate (TSP, Sigma Aldrich). 1M stock solutions of ArgHCl, ArgGlu, NaGlu, Pro, Gly, glycerol, sorbitol, trehalose, and sucrose were prepared in highly purified water, and pH was adjusted to pH 7.2. For poloxamer 407 and tween 20 1% w/v stock solutions were prepared in highly purified water, and pH was adjusted to pH 7.2. Aliquots of excipients were prepared from stock solutions, freeze-dried and successively reconstituted with protein solution to achieve accurate concentrations of excipients without sample dilution.

5.5.2 NMR experiments

All NMR spectra were acquired at 25 °C on 800 MHz Bruker Avance III spectrometer equipped with 5mm triple resonance TCI cryoprobe and temperature control unit. The spectra were acquired and processed using Bruker Topspin 3.5 and analysed using NMRFAM-SPARKY[314], Dynamics Center 2.2.4 (Bruker), GraphPad Prism 7.04 (GraphPad Inc.) and MATLAB. Proton chemical shifts were referenced to TSP signal at 0.0 ppm. ^{15}N and ^{13}C chemical shifts were calculated relative to TSP by using the gyromagnetic ratios of ^{15}N, ^{13}C and ^{1}H nuclei.

The excipient titration experiments were carried out in 5 mm NMR tubes (Norell). Final concentrations of excipients were 0.075, 0.15, 0.30, 1, 2, 5, 10, 25, 50 100 and 200 mM for ArgHCl, ArgGlu, NaGlu, Pro, Gly, glycerol, sorbitol, trehalose and sucrose. Final concentration of poloxamer 407 and tween 20 were 0.001, 0.0025, 0.005, 001, 0.005, 0.01, 0.05, 0.1 and 0.5% w/v). Pulse widths were calibrated at each addition of excipients, and for saturation transfer difference experiments, saturation power was recalculated to ensure the equal saturation of all samples.^{1}H- ^{15}N chemical shifts of SQT-1C$^{Q46C,\ N59C}$ were monitored by the acquisition of ^{1}H-^{15} HSQC spectra using the hsqcfpf3gpphwg pulse sequence from the standard Bruker library. ^{1}H- ^{13}C chemical shifts were monitored by ^{1}H-^{13}C HSQC experiment with sensitivity enhancement, gradient coherence selection, and multiplicity editing which enabled us to easily distinguish between CH_2 and CH/CH_3 groups. Chemical shift changes were calculated either as $\Delta\delta_{NH} = \sqrt{0.5[\delta_H^2 + (0.1\delta_N)^2]}$ where $\Delta\delta_H$ and $\Delta\delta_N$ represent the chemical shift changes in proton and nitrogen dimensions or as $\Delta\delta_{CH} = \sqrt{0.5[\delta_H^2 + (0.3\delta_C)^2]}$, where $\Delta\delta_H$ and $\Delta\delta_C$ represent the chemical shift changes in proton and carbon dimensions. All chemical shift changes were fitted either to the linear or equation describing 1:1 binding, using MATLAB R2018a (MathWorks Inc). The full script for the fitting of the data can be found in section 7.2.2.

Additionally, a set of 1D ^{1}H, STD, WaterLOGSY and ^{1}H T_2-filter experiments were acquired. ^{1}H spectra were acquired using a standard Bruker pulse sequence with excitation sculpting with gradients for water suppression (zgesgp). For STD NMR studies a standard Bruker stddiffesgp.3 pulse sequence was used with the interleaved acquisition of on- and off-resonance spectra. On- and off-resonance saturation frequencies were 0.716 ppm and 40 ppm, respectively. Saturation time of 3 s was used, and 50 ms spin lock filter was applied to eliminate protein signals. STD spectra were obtained by subtracting on-resonance from off-

resonance spectrum. Appropriate control experiments without the addition of proteins were performed to confirm no direct irradiation of excipients. STD amplification factors were calculated using the following equation:

$$AF^{STD} = \frac{I_{off} - I_{on}}{I_{off}} * \frac{[L]}{[P]}$$

where AF^{STD} is the STD amplification factor, I_{off} the intensity of excipient signal in the off-resonance spectrum, I_{on} the intensity of the signal in the on resonance spectrum, and *[L]* and *[P]* represent total concentrations of excipients and SQT-1C$^{Q46C, N59C}$, respectively.

For WaterLOGSY experiments, ephogsygpno.2 pulse sequence from standard Bruker library was used with mixing time of 3s. To determine the reference spectra, appropriate control experiments of each excipient in the absence of protein were acquired. Changes in transverse relaxation rates of protons were monitored using two-point Carr-Purcell-Meiboom-Gill (CPMG) experiments, with the number of echoes 1 and 128. The fixed echo time to allow elimination of diffusion and J-modulation effect was 3 ms. Temperature compensation was used to ensure equal heating of the sample over the course of the experiment. Ratio between signal intensities in both spectra was calculated and compared to reference intensities of excipients alone, without the presence of protein.

5.6 Supplementary Information

5.6.1 Supplementary Figures

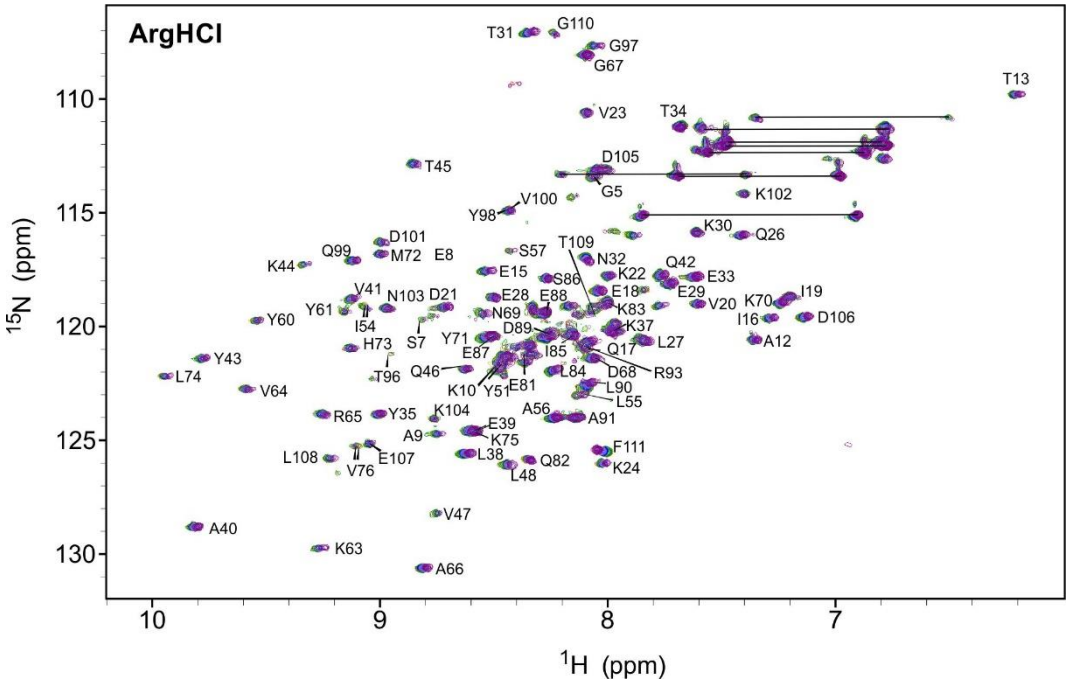

Figure S5.1 Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of ArgHCl: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

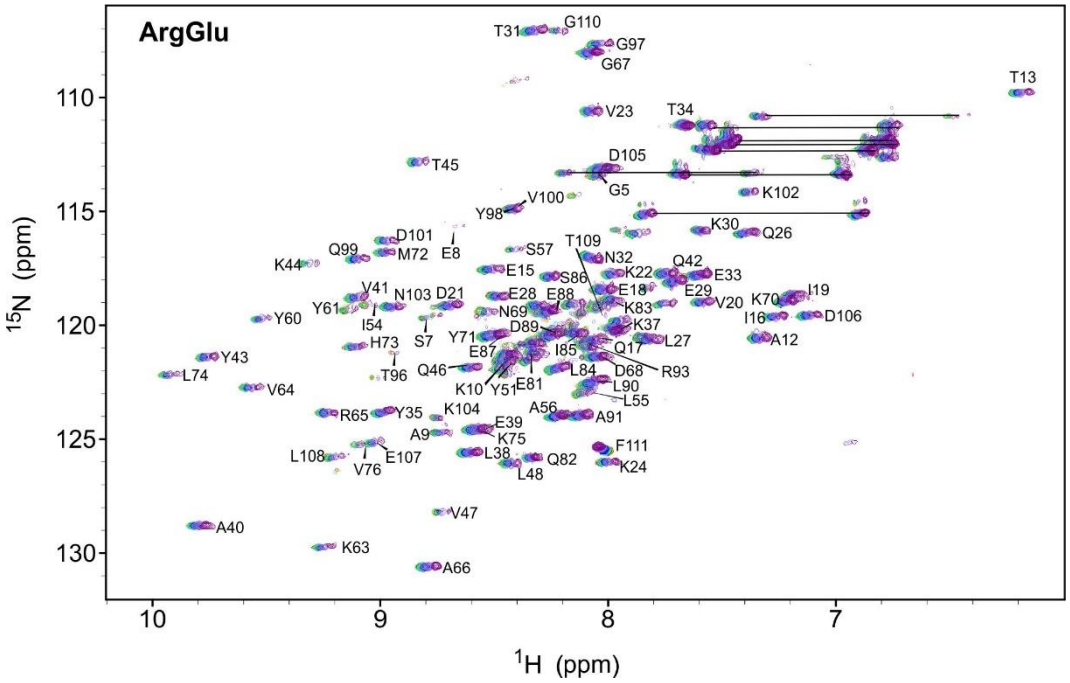

Figure S5.2: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of ArgGlu: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (), 100 (light purple) and 200 mM (purple).

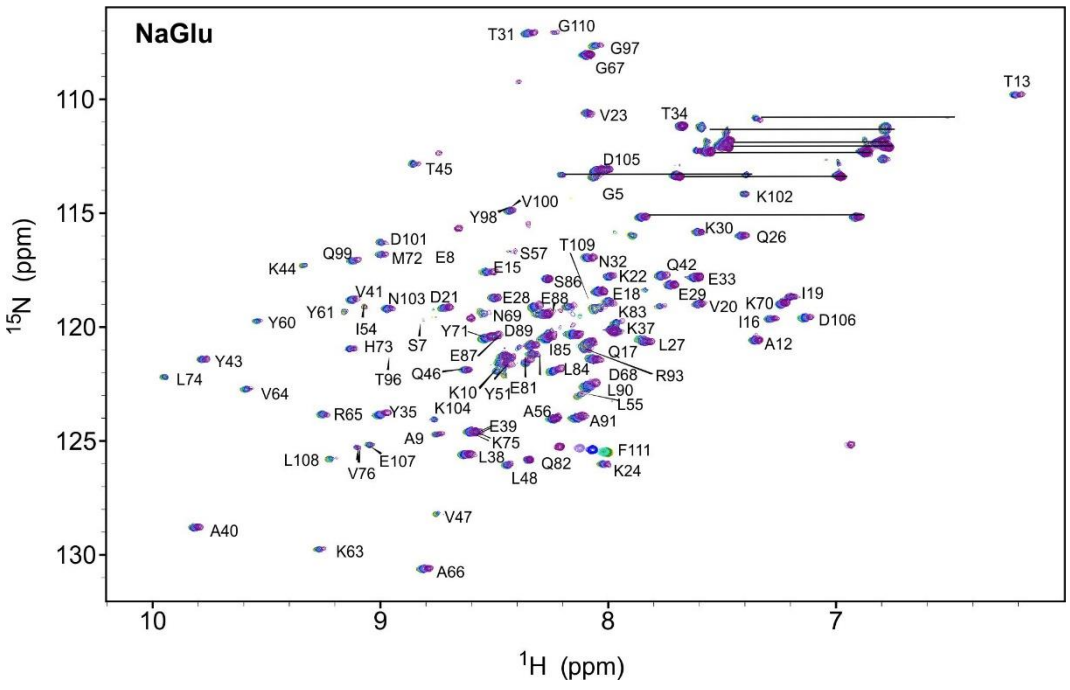

Figure S5.3: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of NaGlu: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

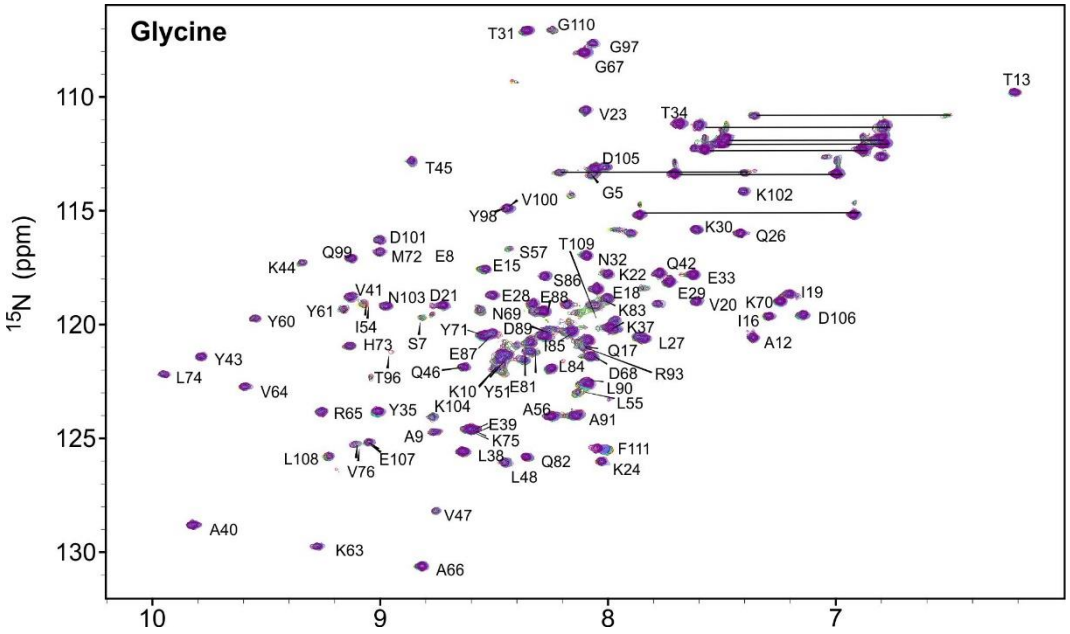

Figure S5.4: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of glycine: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

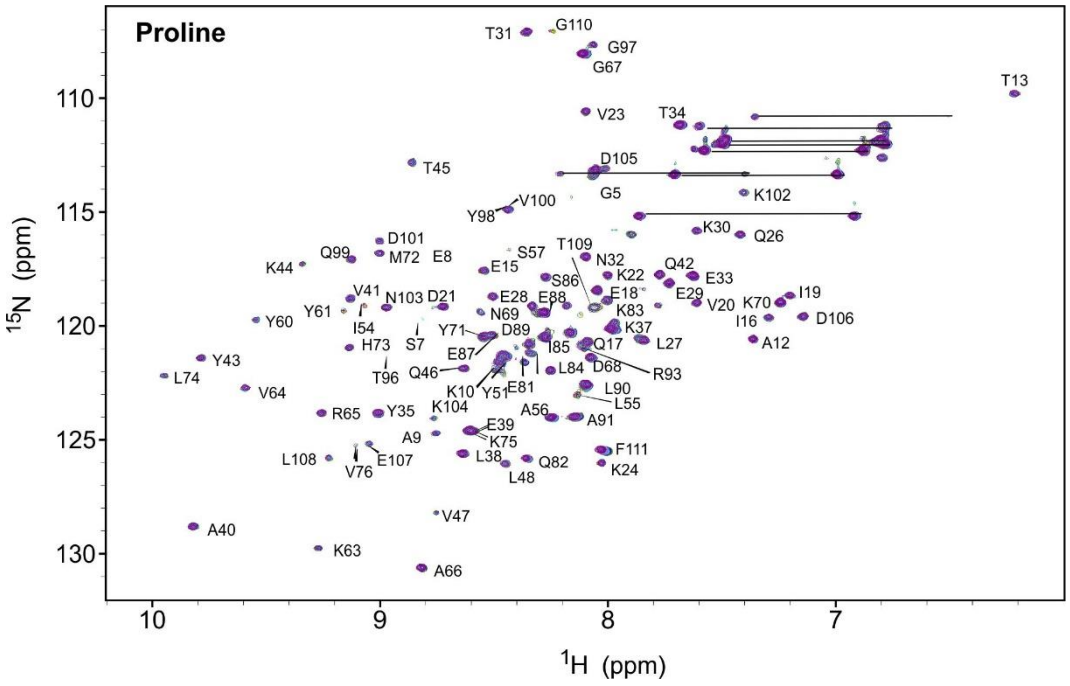

Figure S5.5: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of proline: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

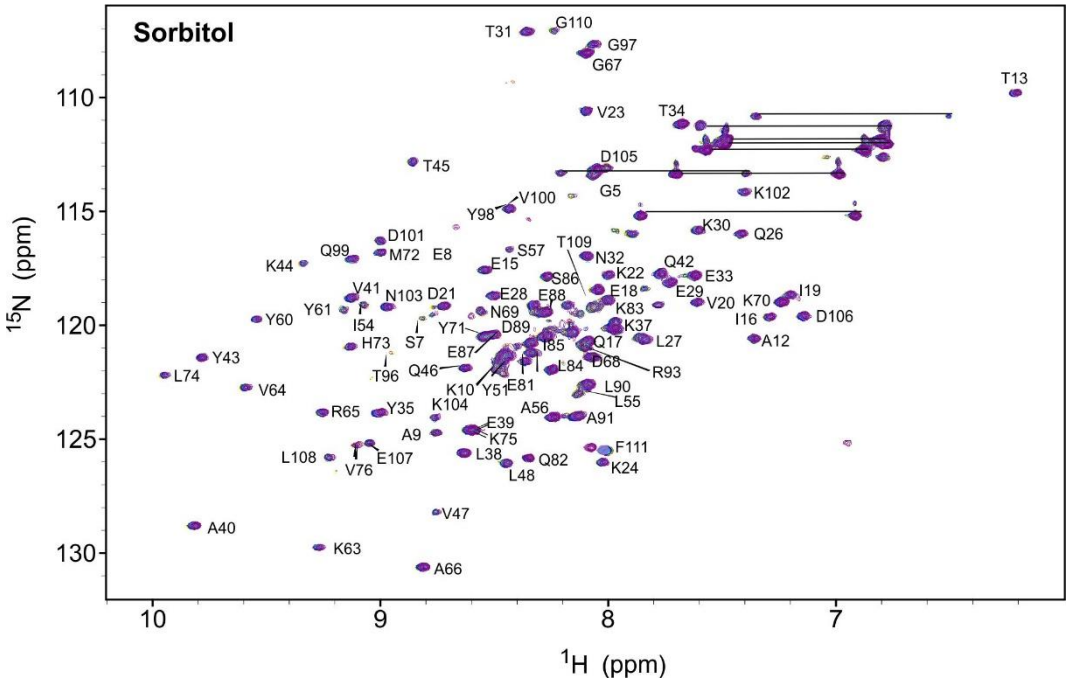

Figure S5.6: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of sorbitol: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

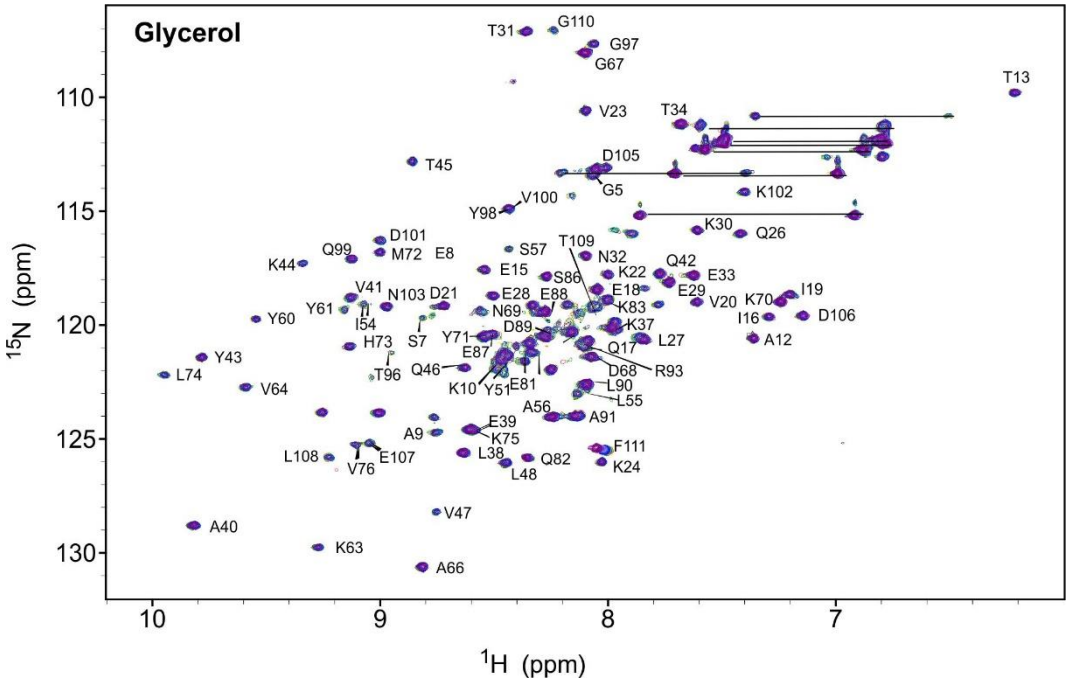

Figure S5.7: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of glycerol: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

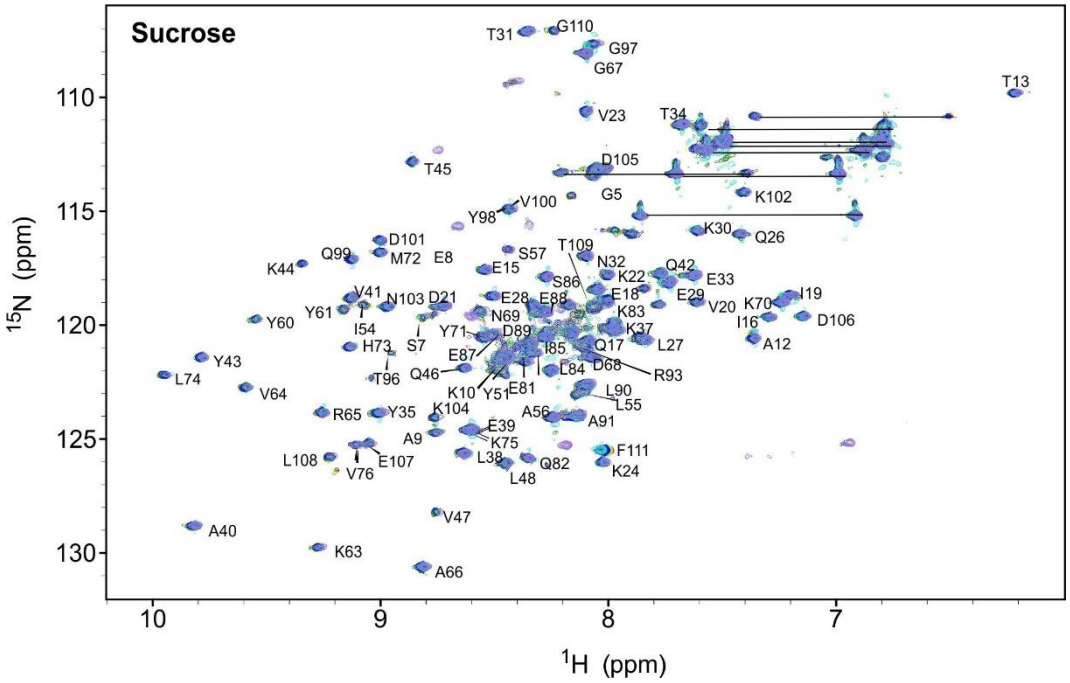

Figure S5.8: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of sucrose: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), and 100 (light purple).

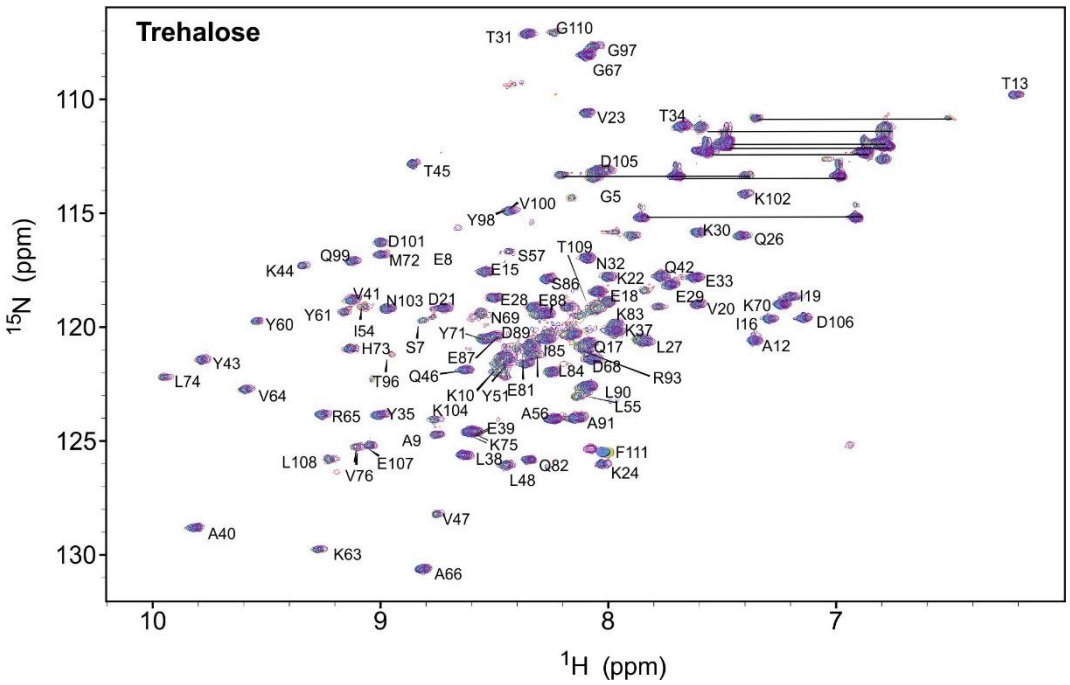

Figure S5.9 Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of trehalose: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

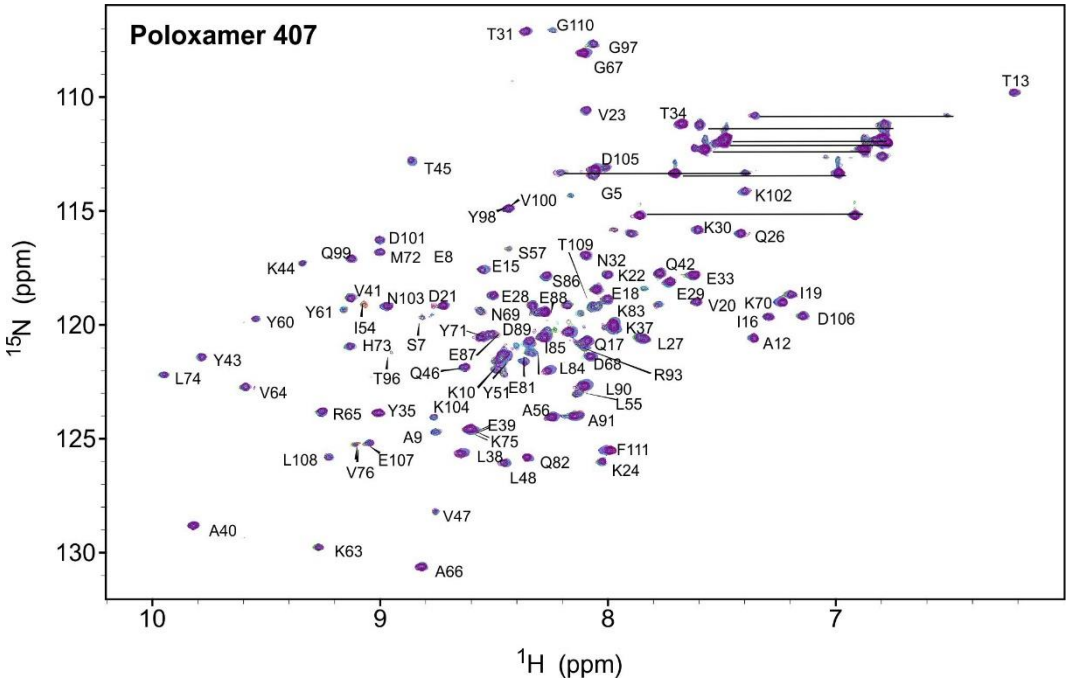

Figure S5.10: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of poloxamer 407: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

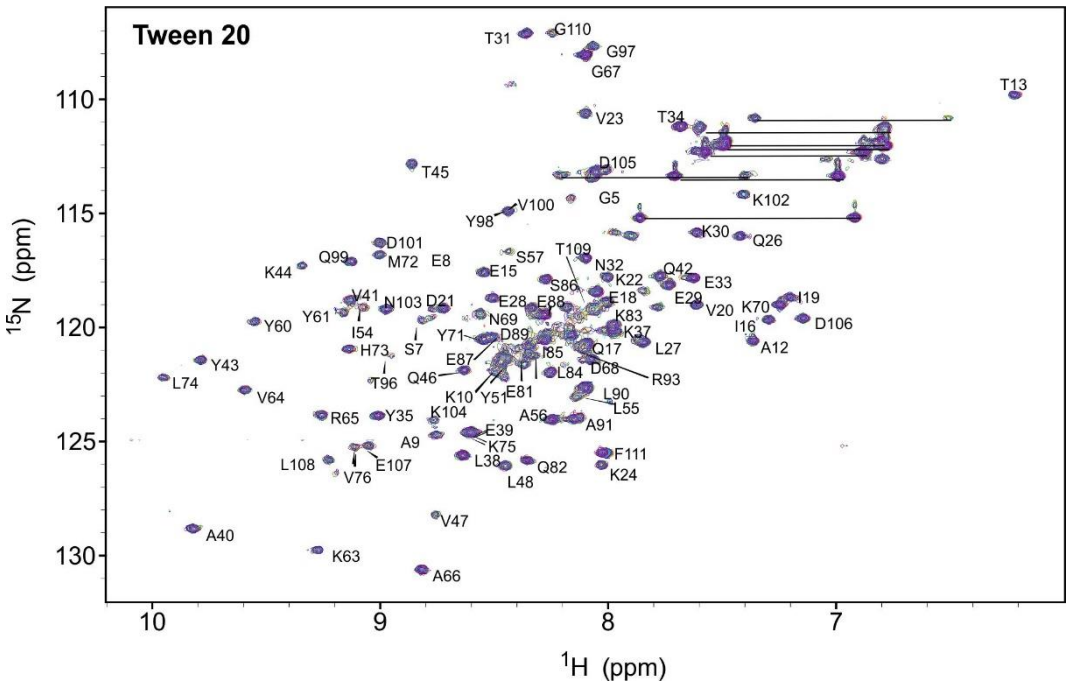

Figure S5.11: Overlay of ^{1}H-^{15}N HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of tween 20: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (turquoise), 50 (blue), 100 (light purple) and 200 mM (purple).

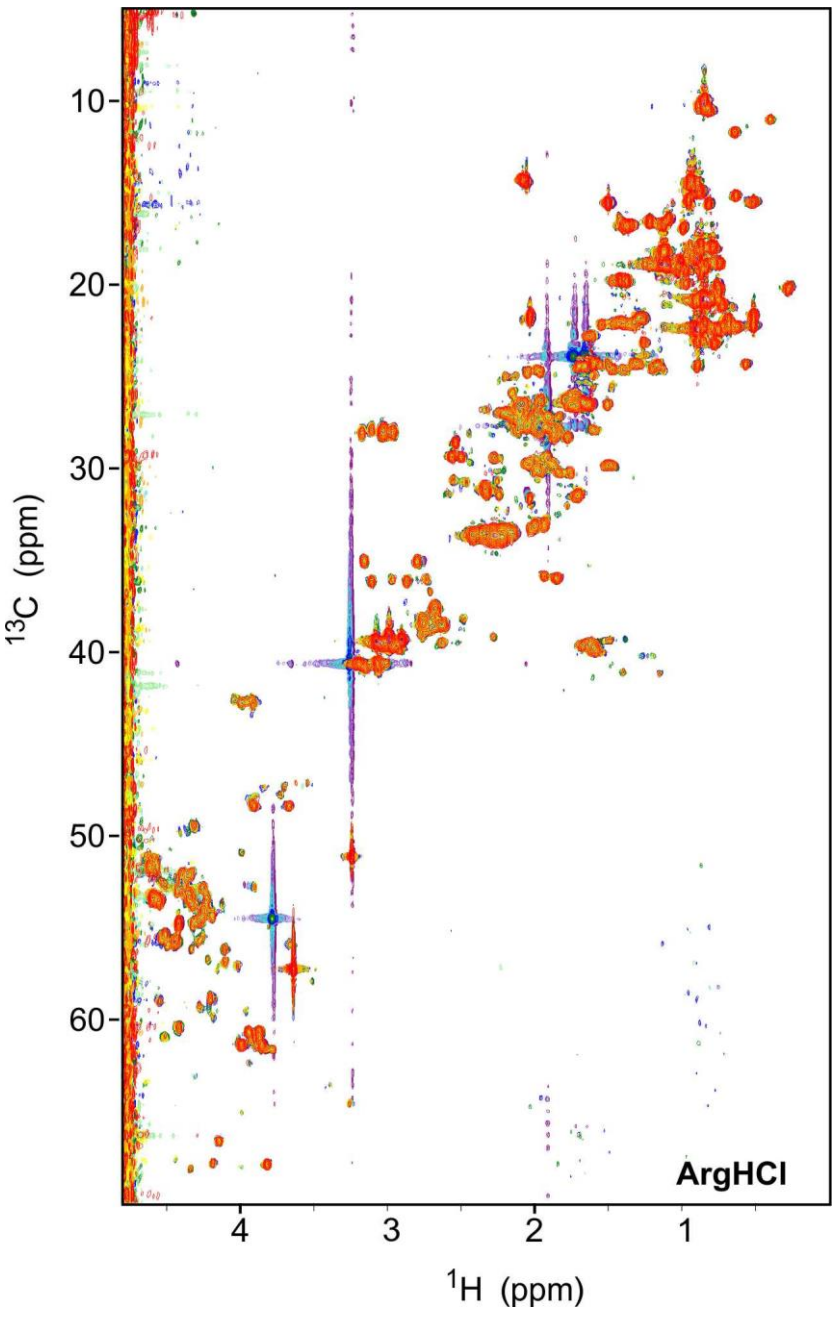

Figure S5.12: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of ArgHCl: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

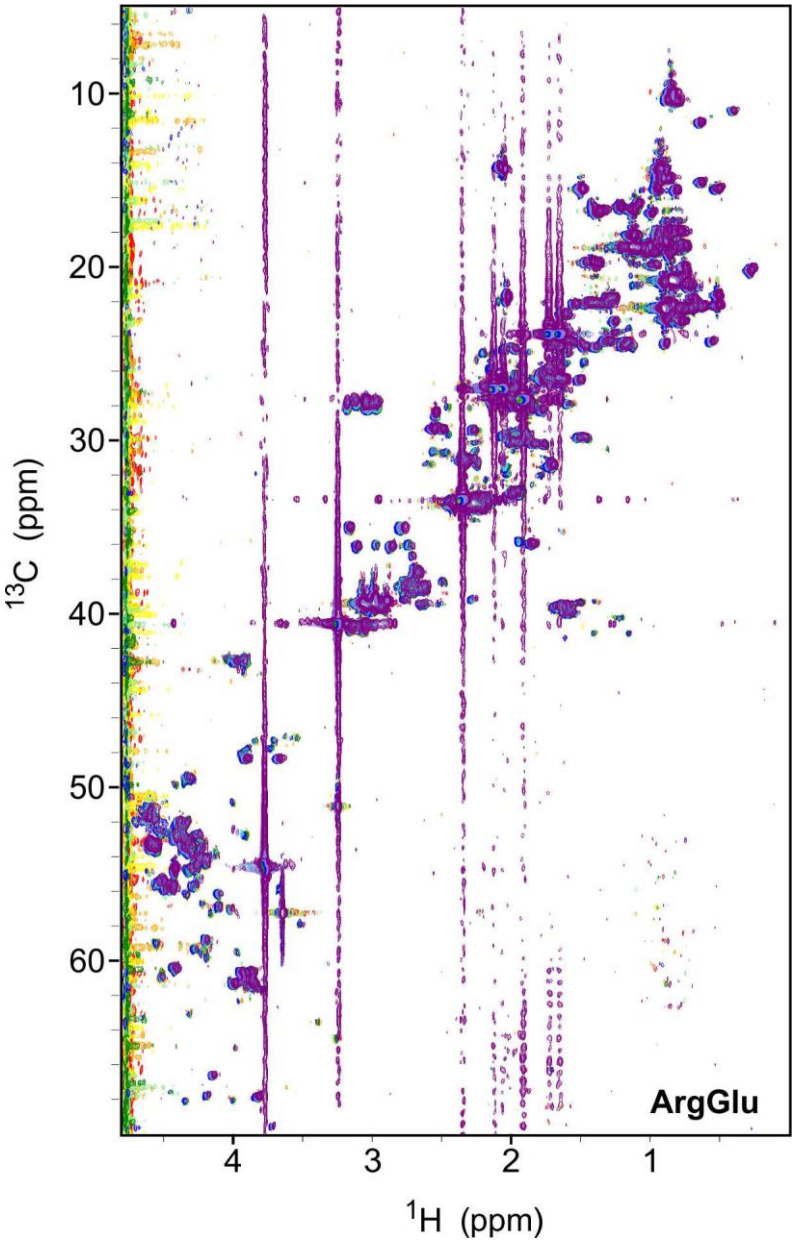

Figure S5.13: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of ArgGlu: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

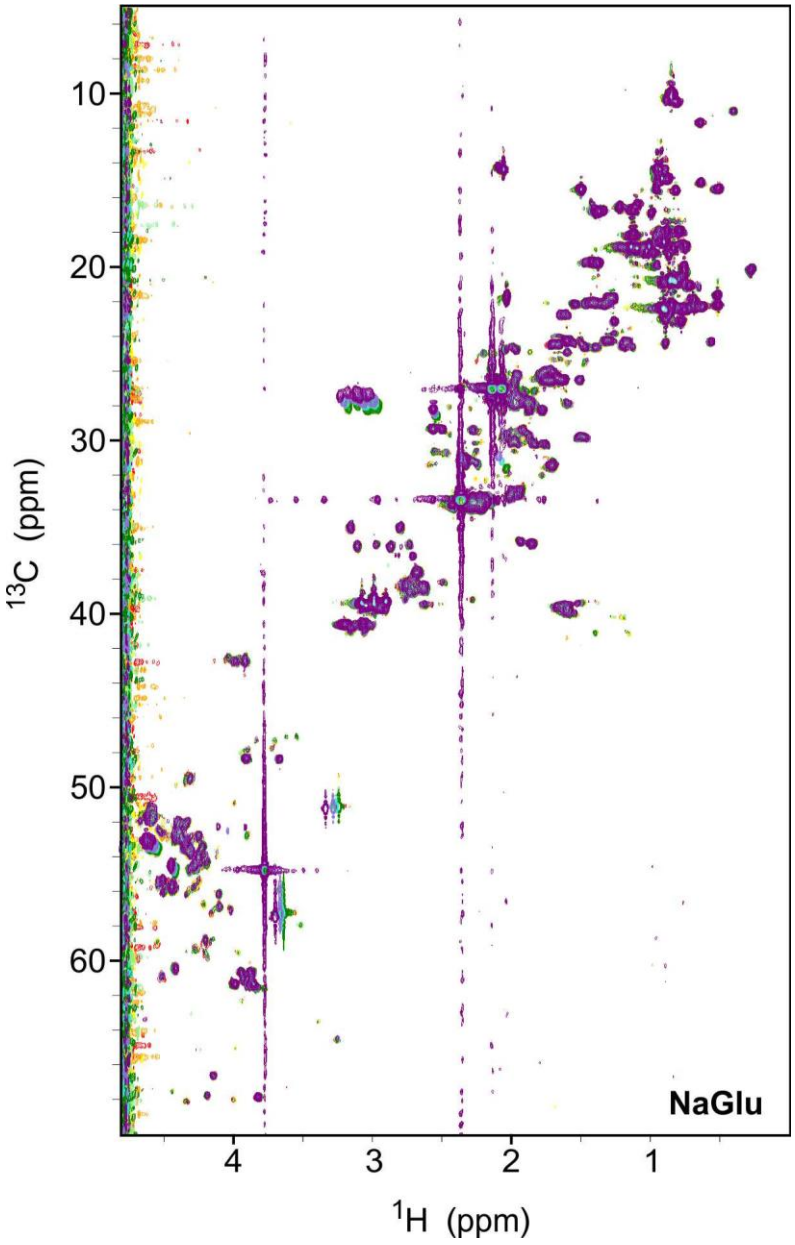

Figure S5.14: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of NaGlu: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

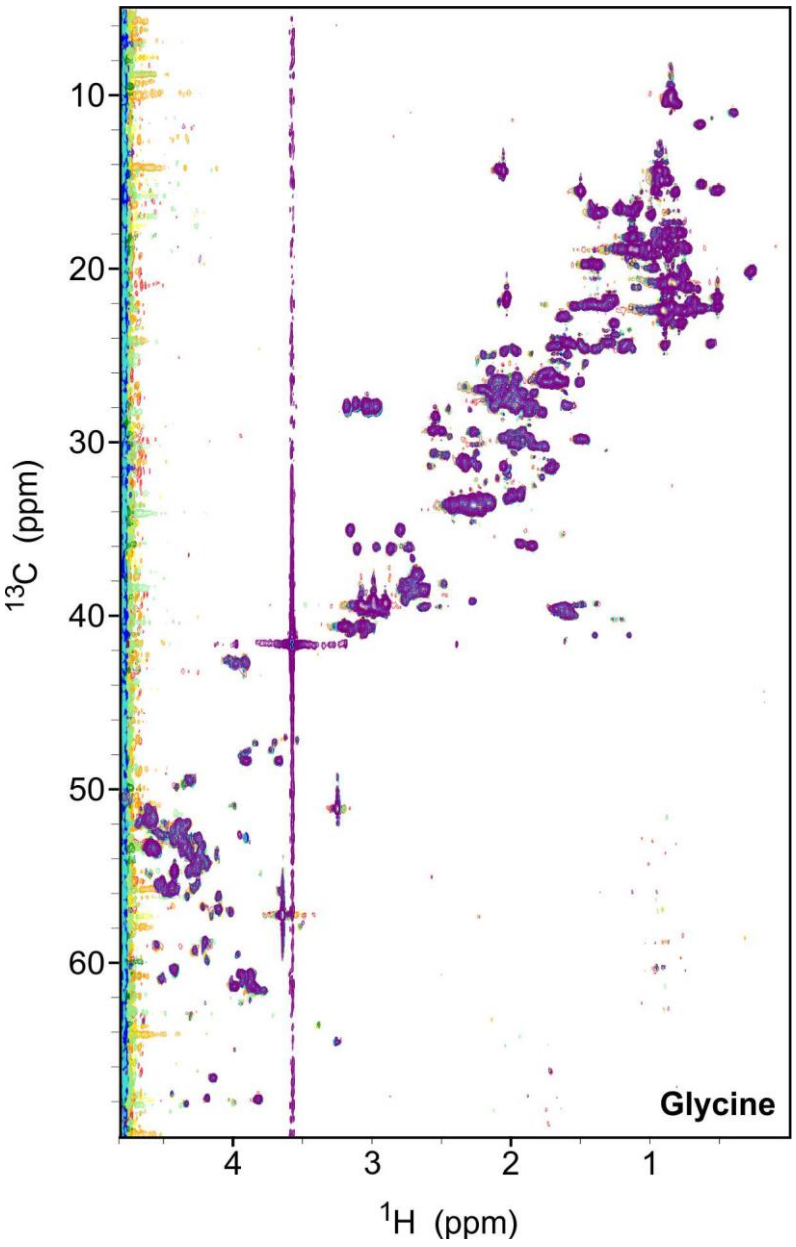

Figure S5.15: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of Gly: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

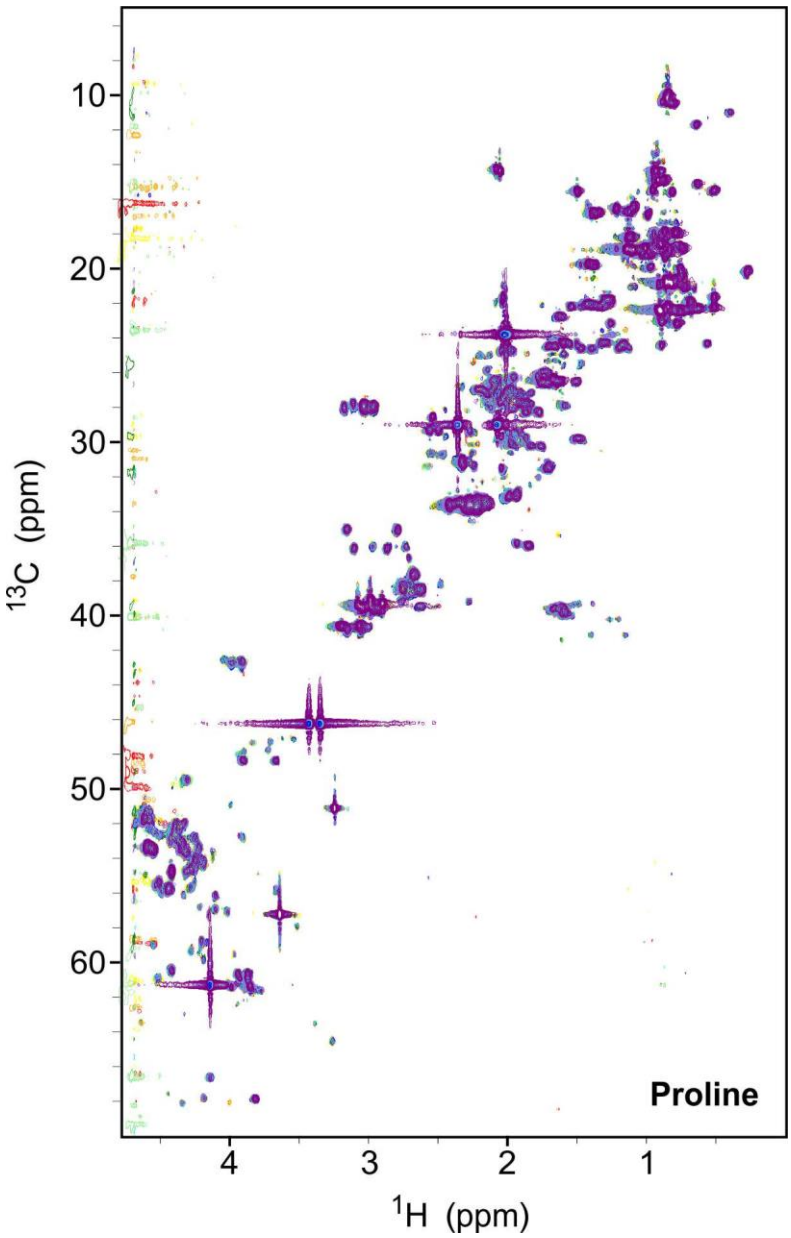

Figure S5.16: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of Pro: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

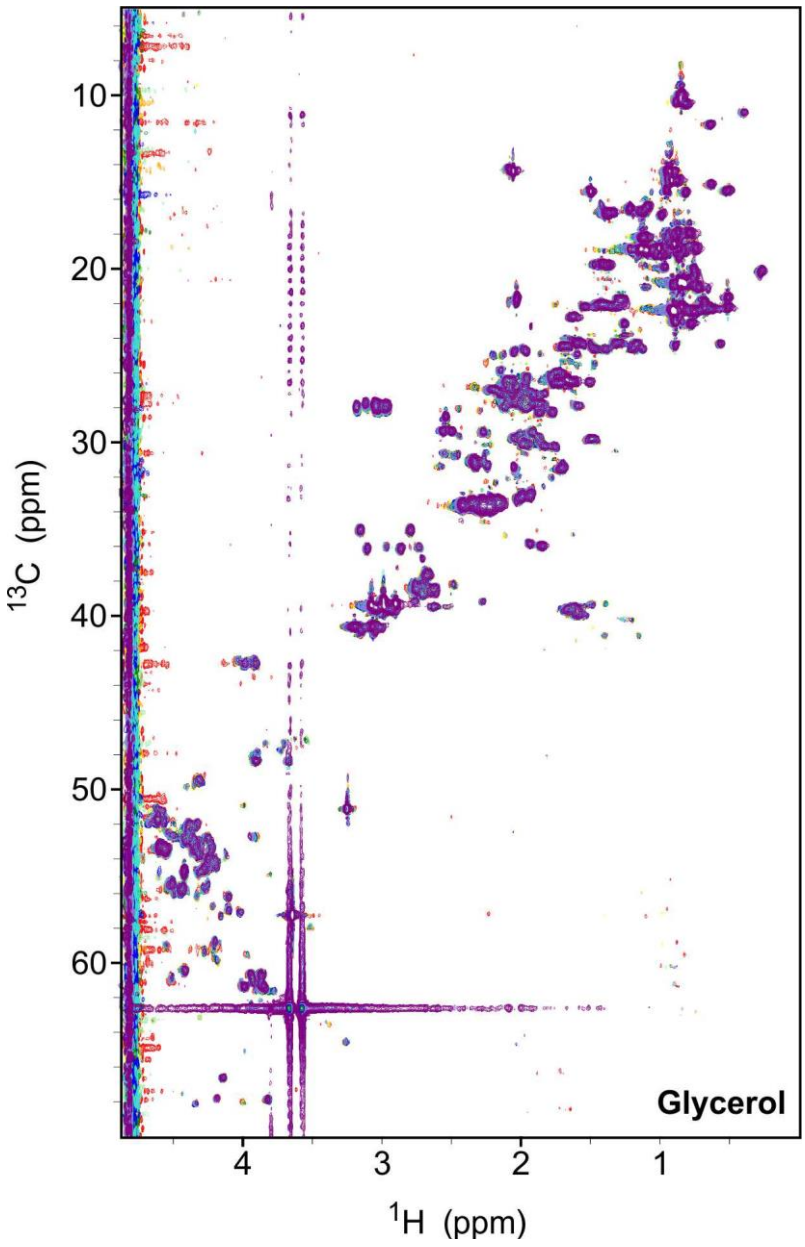

Figure S5.17: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of glycerol: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

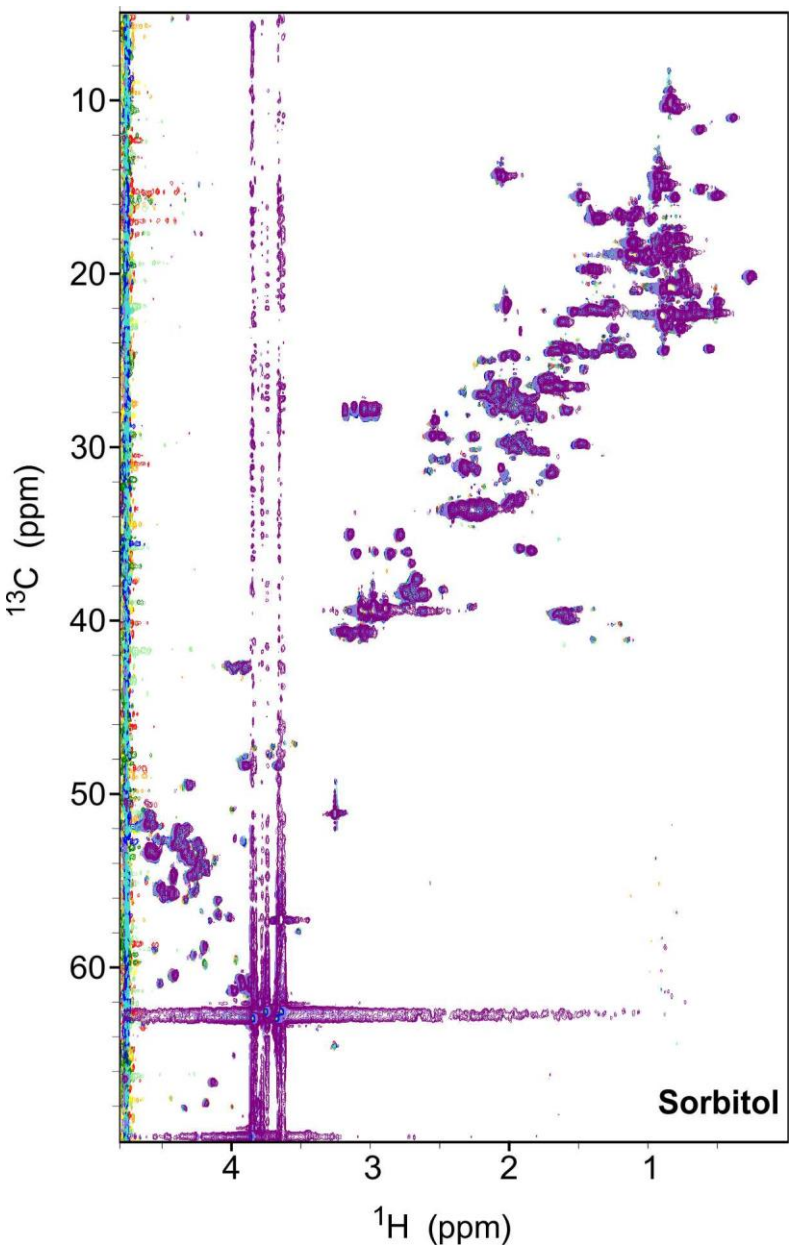

Figure S5.18: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of sorbitol: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

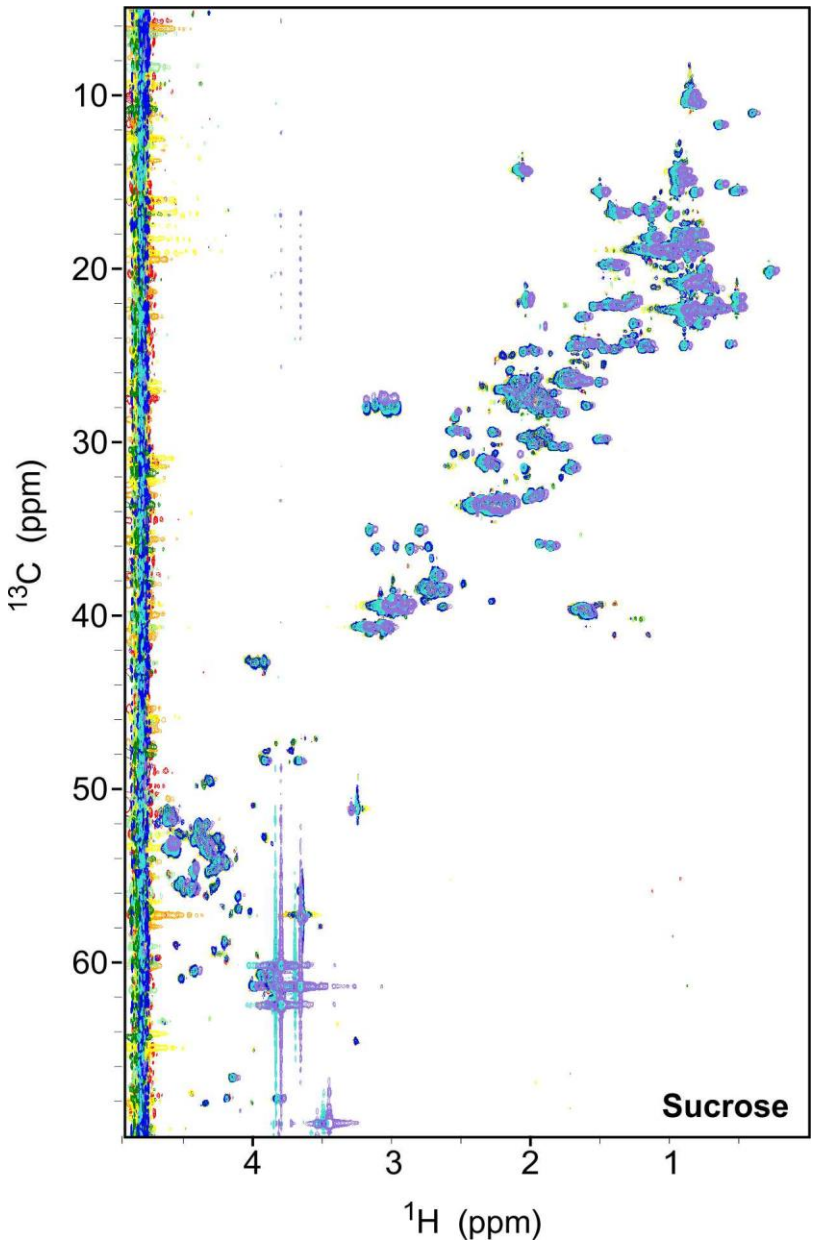

Figure S5.19: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of sucrose: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise) and 100 (light purple).

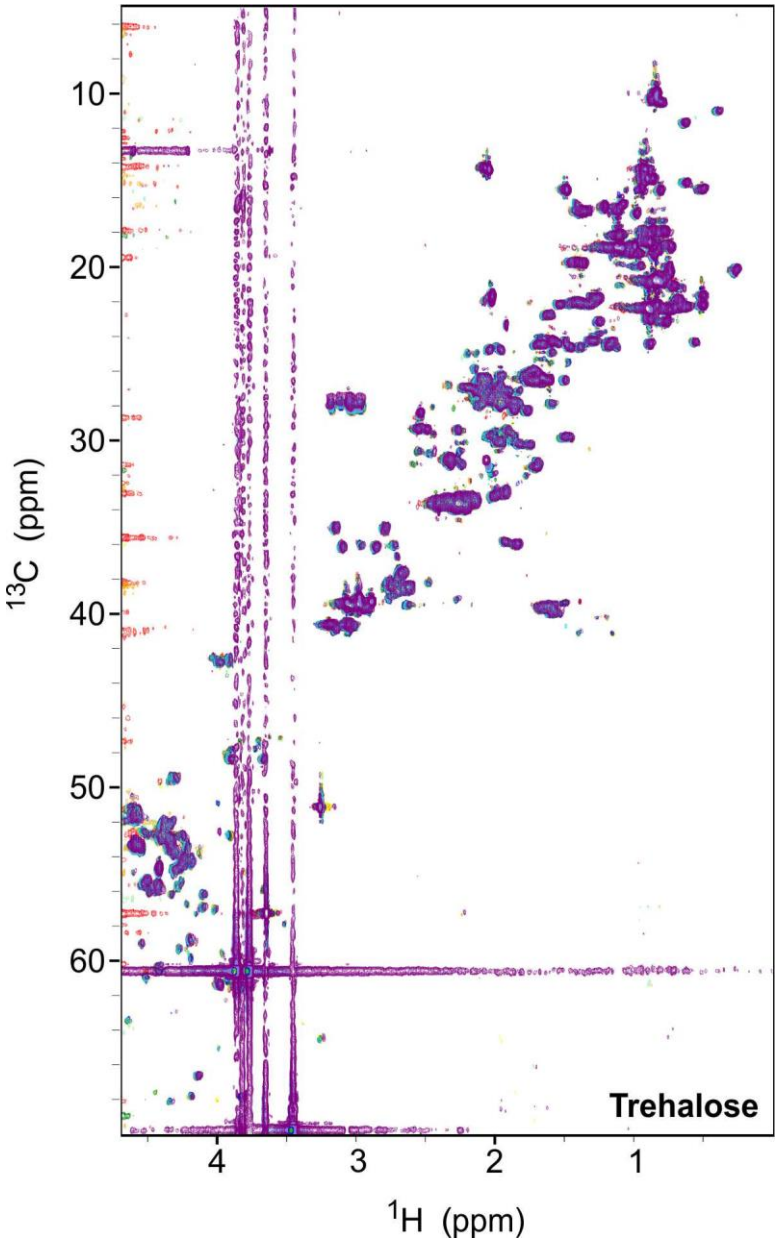

Figure S5.20: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of trehalose: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

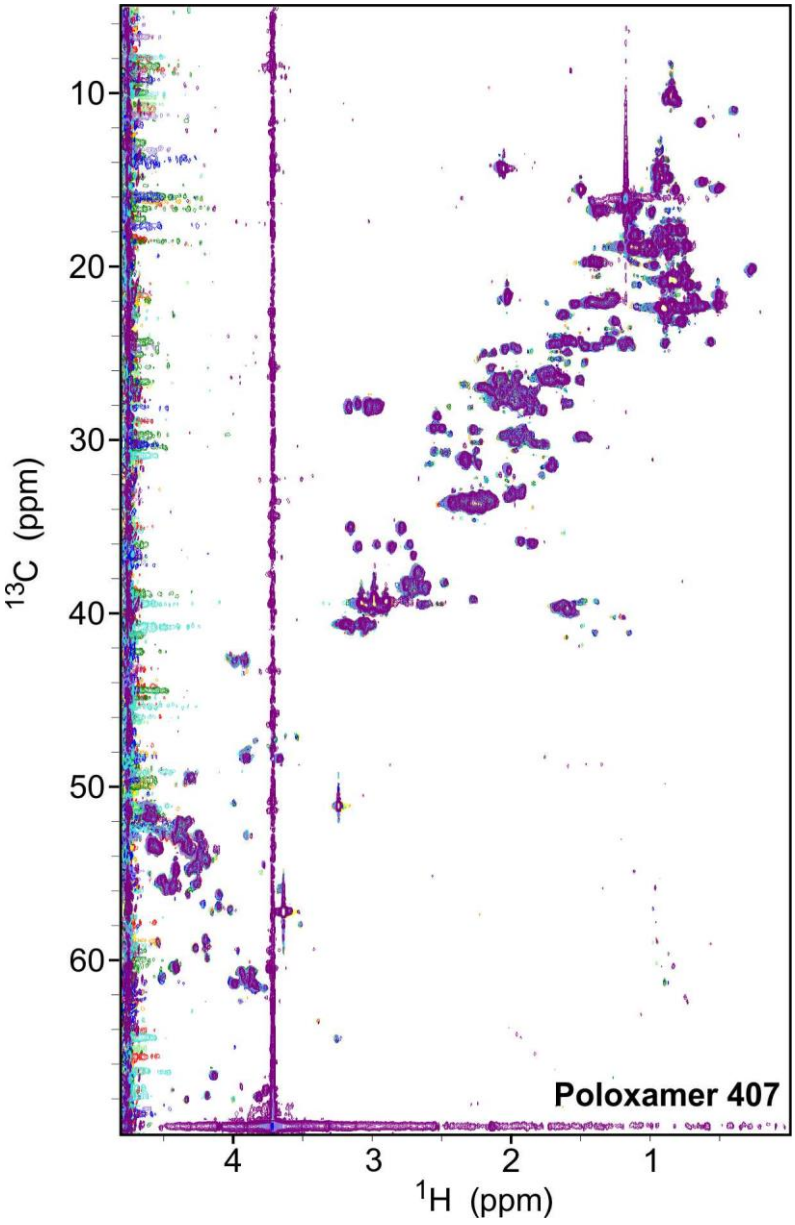

Figure S5.21: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of poloxamer 407: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

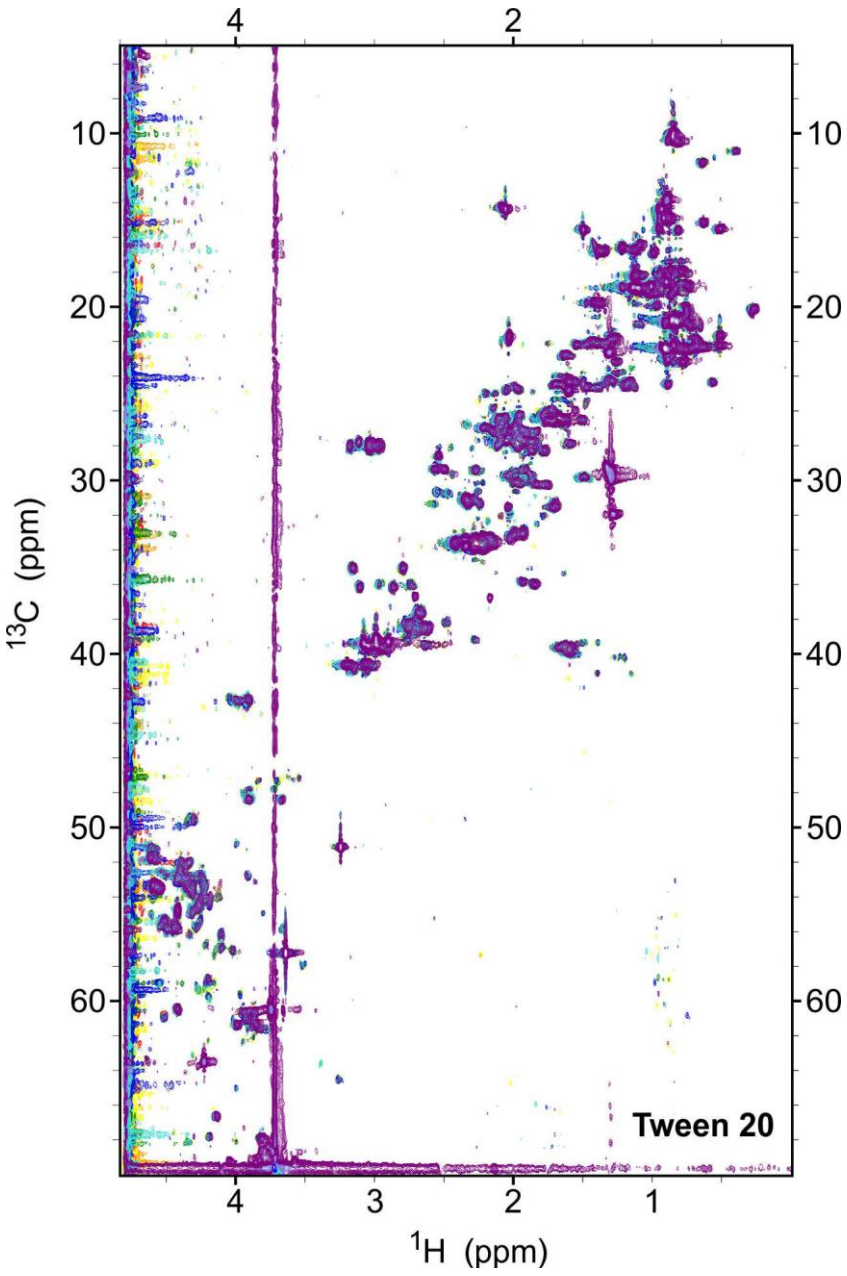

Figure S5.22: Overlay of ^{1}H-^{13}C HSQC spectra of SQT-1C^{Q46C,N59C} in the presence of increasing concentrations of tween 20: 0 (red), 0.075 (orange), 0.15 (yellow), 0.3 (light green), 1 (green), 10 (blue), 50 (turquoise), 100 (light purple) and 200 mM (purple).

BLANK PAGE

6 General conclusions and future directions

The main aim of this thesis was to characterise protein-protein and protein-excipient interactions in biopharmaceutical formulations using NMR spectroscopy and evaluate how these affect various aspects of protein stability using a range of complementary techniques. In our work we have investigated two groups of proteins; biopharmaceutical antibodies supplied by the partners of PIPPI consortium and a peptide aptamer, SQT-1C, as an alternative to mAbs.

The work presented in Chapter 2 and Chapter 3 has focused on the oligomerisation pathways of the model peptide aptamer, called SQT-1C, and approaches to prevent it. Peptide aptamers are typically built by inserting short peptides containing the desired binding epitope into a loop of a stable protein scaffold, and in principle mimic the CDR regions of an antibody. However, inserting peptides into the protein scaffold may unintentionally decrease thermal and colloidal stability of the scaffold protein[51]. In our case, we have inserted the AU1 and c-Myc peptides into the SQT scaffold, creating the SQT-1C and used it as a model system to study in detail the effect of peptide insertions on its conformational and thermal stability. In Chapter 2 we showed that while the secondary structure and the overall fold of SQT was not perturbed by the peptide insertions, it readily formed dimers and tetramers in solution. We furthermore established that inserted loops were directly involved in this process and determined using computational approaches that the oligomer formation most likely proceeds through domain swap mechanism. Formation of oligomers through loop regions that are involved in binding to the target of the protein may, however, decrease or fully diminish the binding activity of the peptide aptamer, which is detrimental to their use. This results also pointed out that instead of focusing solely on the stability of the scaffold protein on its own, routine assessment using a combination of analytical methods on the final peptide aptamers would reveal early problematic behaviour of designs, helping to troubleshoot protein instability and loss of structure and confirm the presence or absence of well-behaved stable protein fold.

In Chapter 3 we continued the work on SQT-1C and characterised in detail the pathway and kinetics of SQT-1C dimer and tetramer formation. We established that tetramer formation occurs through the association of domain-swapped dimers, where domain swap occurs through loop1. We also showed that the conserved Pro80 located at the base of the

loop 2 is one of the factors in tetramer formation, but not the limiting one. By introducing a double mutation, introducing disulphide bond at the base of loop 1 we have designed a conformationally stable mutant SQT-1C^{Q46C,N59C} that does not oligomerise and has improved thermal stability. Introduction of disulphide bonds into protein to improve their thermal and conformational stability is already an established strategy[121, 338], while the so-called 'peptide stapling' is emerging technology to constrain peptides into a defined conformation[308, 408-409]. We propose that adding disulphide bonds at the base of ligand binding loop(s) may increase scaffold stability and maximise its specific target-binding activity and could as such be applied to various other scaffold proteins that are subject to destabilisation upon peptide loop insertions. In our work we focused on a single aptamer where c-Myc and AU1 peptides were inserted into the loops of SQT-1C^{Q46C,N59C} scaffold while the N-terminal insertion site was not utilised. Stability and activity of SQT-1C^{Q46C,N59C} scaffold could therefore be further tested by insertion of peptides of various lengths into all three insertion sites.

In Chapter 4 and 5 we have addressed the interactions between biopharmaceutical proteins and excipients added to their formulation. Small molecule excipients, such as amino acids, sugars, sugar polyols, detergents, etc are typically added to protein formulations to ensure stability and solubility of the active ingredient whilst preventing aggregation and other pathways of degradation over longer periods of time. Due to the weak and transient nature of protein-excipient interactions, these are challenging to evaluate, which is the main reason for the lack of general knowledge on how excipients work/interact with a particular protein. In general, excipients act through preferential interaction or preferential exclusion mechanism, which were postulated in the 1980s[135-137]. These effects are usually measured indirectly and there are limited studies where NMR would be applied to study protein-excipient interactions in biopharmaceutical formulations.

In Chapter 4 we have used ligand-observed NMR spectroscopy to characterise interactions between seven biopharmaceuticaly relevant proteins and a set of eleven commonly used excipients. We applied a screening procedure, where a combination of ^{1}H, CPMG, WaterLOGSY and STD experiments was used to identify excipients that interact with proteins. We introduced an empirical protein-excipient interaction parameter (I^{N}) that reflects all four NMR observables at once and can be thus used as a proxy for interaction "strength" and as an empirical cut-off to distinguish between excipients which interact with proteins and those that do not. We have also measured the effect of excipients on thermal and colloidal protein stability, on aggregation kinetics and protein storage stability at

elevated temperatures. Our results showed that effects of excipients are very much protein specific and that there are no obvious correlations between the strength of protein-excipient interaction and different protein stability parameters. Therefore, the excipient binding as detected by NMR in general cannot be used itself for predicting whether an excipient will be stabilising or destabilising in terms of melting temperature, neither can it be used to predict the change in colloidal stability. Furthermore, we have also concluded that different parameters used as a measure of stabilising effects in protein formulation, e.g. general monomer loss, polydispersity index, rate of aggregation, apparent hydrodynamic radius, all reflect on different attributes of protein stability and provide orthogonal views, not correlated with each other in any obvious way. This further implies that multiple methods should be used to assess different aspects of protein stability and that case studies based on a single model protein may oversimplify the complexity of the formulation space and may produce misleading results that cannot be easily applied to another protein system. It would be interesting to expand this work to a larger number of proteins including other classes of biopharmaceuticals. Additionally, the meaning of empirical protein-excipient interaction parameter (I^N) should be further assessed on a model protein with their known binders to validate our assumption that it reflects the strength of protein-ligand (excipient) interaction.

In Chapter 5 we have explored the application of both ligand- and protein-observed NMR methods to characterise protein-excipient interactions in biopharmaceutical formulations. In this proof-of-concept study we have measured interactions between a model protein SQT-1C$^{Q46C,\ N59C}$ and the same eleven excipients we used in Chapter 4. For ligand-observed experiments we have again used a combination of ^{1}H, CPMG, WaterLOGSY and STD experiments. We show that results of the ligand-observed screen depend heavily on protein to excipient ratio in solution, making these experiments less reliable at higher concentration of excipients, as the effects are masked by the bulk solution. As ligand-observed experiments proved somewhat insensitive at excipients concentrations typically used in biopharmaceutical formulations, we next explored the use of two-dimensional protein-observed ^{13}C HSQC and ^{15}N HSQC experiments to detect the transient protein-excipient interactions. We show that both experiments can be used to probe transient protein-excipient interactions on the protein surface. Our results based on ^{15}N CSP data indicate that there is no specific, well defined binding site for excipients on the protein surface, which could perhaps be anticipated. While ^{13}C CSP data seemed more site-specific, it was not possible to deduce potential interaction sites as C-H assignments were not available. We

could however, use the concentration dependent ^{13}C CSP data to estimate the K_d values for excipients that interact with SQT-1C$^{Q46C, N59C}$. On the other hand, no saturation of CSP was observed in ^{15}N HSQC titration experiment. We show that ^{15}N HSQC titration experiment can distinguish between a preferential interaction mechanism, where chemical shifts are in fast exchange, and preferential exclusion mechanism, where a sudden jump of CSP across the majority of cross-peaks was observed above a certain concentration of excipient. This opens up a possibility to study directly how excipients interact with individual proteins and to deduce their mechanism of action and the concentration needed to achieve a specific stabilisation effect. It can be envisioned that this type of data could be used to guide prediction models on possible protein-excipient interaction sites and guide the selection and the concentration of excipients used in biopharmaceutical formulations. For example, it would be interesting to apply this methodology to study how different point mutations that influence the overall charge of the protein affect protein-excipient interactions.

BLANK PAGE

7 Appendix

7.1 List of other publications

- M. Martinez Morales, M. Zalar, S. Sonzini, A.P. Golovanov, C. F. van der Walle, and J.P. Derrick, **Interaction of a Macrocycle with an Aggregation-Prone Region of a Monoclonal Antibody,** Molecular Pharmaceutics 2019 16 (7), 3100-3108.

Contribution: Design of the NMR experimental setup, and help with data acquisition. Analysis of the data, interpretation of the NMR results, figure design and write up of the NMR sections of the paper. Substantial contribution to paper writing.

- S. Indrakumar, M. Zalar, C. Pohl, A. Nørgaard, W. Streicher, P. Harris, A. P. Golovanov, and G.H.J. Peters, **Conformational Stability Study of a Therapeutic Peptide Plectasin Using Molecular Dynamics Simulations in Combination with NMR,** The Journal of Physical Chemistry B, 2019 123(23), 4867-4877.

Contribution: Design of the NMR experimental setup, NMR data acquisition, processing and analysis. Interpretation of the results, figure design and write up of the NMR sections of the paper. Substantial contribution to paper writing.

- S. Indrakumar, M. Zalar, N. Tschammer, A. Nørgaard, W. Streicher, P. Harris, A. P. Golovanov, G.H.J. Peters, **Synergistic applications of molecular dynamics, microscale thermophoresis and NMR to probe excipient interactions with therapeutic peptide Plectasin**, 2020, *in review.*

Contributions: Design of the NMR experimental setup and helping with NMR data acquisition and processing. Performed analysis of the NMR data and interpreted the results, designed figures and wrote the NMR sections of the paper.

- H.L. Svilenov, A.V. Kulakova, M. Zalar, A. P. Golovanov, P. Harris and G. Winter, **Orthogonal Techniques to Study the Effect of pH, Sucrose and Arginine Salts on Monoclonal Antibody Physical Stability and Aggregation During Long-term Storage**, Journal of Pharmaceutical Sciences, 2020, 109(1):584-94.

Contributions: Design of the NMR experimental setup, NMR data acquisition, processing and analysis. Interpretation of the data, figure design and write up of the NMR sections of the paper. Substantial contribution to paper writing.

- C. Pohl, M. Zalar, I. El Bialy, S. Indrakumar, G.H.J. Peters, A. P. Golovanov, W. Streicher, A. Nørgaard and P. Harris, **The effect of point mutations on the protein properties in solution: development of a screening protocol of an anti-microbial peptide,** 2020, Molecular Pharmaceutics, *accepted.*.

Contributions: Design of the NMR experimental setup, NMR data acquisition, processing and analysis. Interpretation of the data, figure design and write up of the NMR sections of the paper. Substantial contribution to paper writing

- A.V. Kulakova, L. Gentiluomo, H.L. Svilenov, D. Augustijn, I. El Bialy, ML. Greco, S. Indrakumar, S Mahapatra, M. Martinez Morales, C. Pohl, A. Roche, M. Zalar A Tosstorff, R. Curtis, J. P. Derrick, A.Nørgaard, Tarik A. Khan, A.P. Golovanov, G.H.J. Peters, A. Pluen, Å. Rinnan, W. Streicher, C. F. van der Walle, S. Uddin, G Winter, D Roessner, Wolfgang Frieß and P Harris, **Advancing excipient selection for protein formulation through comprehensive biophysical characterisation**, 2020, *in preparation.*

Contributions: Design of the NMR experimental setup, NMR data acquisition, processing and analysis. Interpretation of the data, figure design and write up of the NMR sections of the paper.

7.2 MATLAB scripts

7.2.1 Script for fitting aggregation data using Finke-Watzky 2-step mechanism

```
%% SCRIPT FOR FITTING AGGREGATION DATA USING FINKE-WATZKY 2-STEP MECHANISM
%Matja Zalar and Jack Bramham, February 2019
clear
infile = "PPI10 Isothermal 75C 3,17,13 rep2.xlsx";
outfile = "PPI8 3 13 17 FW_3 output.xlsx";
disp_fits = 1; %0 = no, 1 = yes
out_dataandfits = 0; % 0 = output only fitted parameters, 1 = output fitted parameters
and curves (can take a long time!)
user_MaxNumChanges = 10;

inmatrix = xlsread(infile,5);
[~,titles,~] = xlsread(infile,1,'G:G');
titles(1)=[]; %remove first cell
% separate time and scattering data
sz = size(inmatrix);
n = sz(2)/3;
tn = sz(1);
times = inmatrix(:,1:3:end);
scat = inmatrix(:,3:3:end);

%baseline scattering data
scat_base = scat(1,:);
corr_scat = scat - scat_base;

%identify transitions
%4 transitions per curve
[trans_0_1,S1, S2] = ischange(corr_scat,'linear','MaxNumChanges',user_MaxNumChanges);
%trans_0_1 = 1 means transition point
%index of transition points
trans = cell(1,n);
for transi=1:n
    snt = sum(trans_0_1(:,transi)==1); %identify number of transitions from previous
step (=1), upto 10
    trans = find(trans_0_1(:,transi)==1,snt);
end
%convert transition index to time
trans_t = cell(1,n);
for ni=1:n
    sztrans = size(trans);
    trans_t_hold = [];
    for trans_ti=1:sztrans(1)
        trans_t_hold(trans_ti) = times(trans(trans_ti),ni);
    end
   trans_t1 = trans_t_hold;
end
```

```
%user evaluation of data
try
    load user_ins
catch
    user_in = zeros(1,n);
    for ni=1:n
        fign = figure();
        hold on
        plot(times(:,ni),corr_scat(:,ni))
        sztrans = size(trans{ni});
        for li=1:sztrans
            line([trans_t{ni}(li) trans_t{ni}(li) ],ylim);
            text(trans_t{ni}(li),0,num2str(li));
        end
        title(titles{ni})
        xlabel('Time')
        ylabel('Scattering')
        t_n = input(strcat('Line indicating end of first transition: 0-
',num2str(user_MaxNumChanges)));
        while t_n < 0 || t_n > user_MaxNumChanges
            t_n = input(strcat('Line indicating end of first transition: 0-
',num2str(user_MaxNumChanges)));
        end
        user_in(ni) = t_n;
        close(fign)
    end
    save('user_ins.mat','user_in')
end


%fitting
%initial parameters
x0 = [0.000001, 0.00001, 70];
lb = [0, 0, 0]; % lower bounds
ub = [inf, inf, 500]; % upper bounds

out_param = zeros(n, 7);
out_ci = zeros(n, 3);
out_fit = zeros(n, 1);
out_time = cell(1,n);
out_data = cell(1,n);
out_curve = cell(1,n);
%fitting for each curve
for ni=1:n
    if user_in(ni)>=1 %use only first part of curve
        x1 = times(1:trans{ni}(user_in(ni)),ni);
        y1 = corr_scat(1:trans{ni}(user_in(ni)),ni);
    else
        x1 = times(:,ni);
        y1 = corr_scat(:,ni);
    end
    out_time{ni} = x1;
    out_data{ni} = y1;

    [param, F2, fitresid, jacobian] = fitting(x0, lb,ub, x1, y1);
    out_param(ni,1:3) = param;
    paramfit = sig(param(1), param(2), param(3), x1);
    out_curve{ni} = paramfit;

    ci = nlparci(param,fitresid,'jacobian',jacobian,'alpha',0.05);
    cie = param - ci(:,1)';
    out_ci(ni,:) = cie;
```

```
    %calculate extra parameters
    %tlag = param(4) - (2*param(5));
    % rate = 1/param(5);
    %out_param(ni,6) = tlag;
    % out_param(ni,7) = rate;

    cost_func = 'NRMSE';
    fit = goodnessOfFit(paramfit,y1,cost_func);
    out_fit(ni,1) = fit;

    if disp_fits==1
        figure()
        hold on
        plot(x1,y1)
        plot(x1,paramfit)
        title(titles{ni})
        xlabel('Time')
        ylabel('Scattering')
    else
    end
end



% export fit parameters to xls file
xls_param_header = ["Code", "k1", "k1 CI","k2", "k2 CI","a0", "a0 CI", "nrmse fit"];
xlswrite(outfile,xls_param_header,'Fitted Values','A1');
xlswrite(outfile,titles,'Fitted Values','A2');
xlswrite(outfile,out_param(:,1),'Fitted Values','B2');
xlswrite(outfile,out_ci(:,1),'Fitted Values','C2');
xlswrite(outfile,out_param(:,2),'Fitted Values','D2');
xlswrite(outfile,out_ci(:,2),'Fitted Values','E2');
xlswrite(outfile,out_param(:,3),'Fitted Values','F2');
xlswrite(outfile,out_ci(:,3),'Fitted Values','G2');
xlswrite(outfile,out_fit,'Fitted Values','H2');

if out_dataandfits == 1
%export times, exp, and fitted curves for each sample
for ni=1:n
    xls_data_header = ["Time 's'","Corr exp scattering","Fitted curve"];
    xlswrite(outfile,xls_data_header,titles{ni},'A1');
    xlswrite(outfile,out_time{ni}(:,1),titles{ni},'A2');
    xlswrite(outfile,out_data{ni}(:,1),titles{ni},'B2');
    xlswrite(outfile,out_curve{ni}(:,1),titles{ni},'C2');
end
else
end

%% functions

function F=sig(k1,k2,a0,x)
out = zeros(length(x),1);
for xi=1:length(x)
    out(xi) = a0-((k1/k2+a0)/(1+k1/(k2*a0)*exp((k1+k2*a0)*x(xi))));
end
F = out;
end

function F=chi(k1,k2,a0,x,data)
%call function for fit
fit = sig(k1,k2,a0,x);
F = fit - data;
end

function [F, F2, F3,F4] =fitting(x0,lb,ub, x, data)
fun = @(x0)(chi(x0(1),x0(2),x0(3),x,data));
options = optimoptions('lsqnonlin','FiniteDifferenceType', 'central','Algorithm',
'trust-region-reflective', 'OptimalityTolerance', 1e-12,'TolFun',1e-12,'TolX', 1e-12,
'MaxFunEvals', 2000);
[param, resnorm,residual,~,~,~,jacobian] = lsqnonlin(fun, x0, lb, ub,options);
F=param;
F2 = resnorm;
F3 = residual;
F4 = jacobian;
end
```

7.2.2 Script for fitting chemical shift perturbation data

```
%SCRIPT FOR FITTING CSP DATA
% Matja Zalar and Jack Bramham
% August 2019
clear
infile = "infile.xlsx";
outfile = "outfile.xlsx";
imagefile = "imagename_";
sheetn = 1; %sheet number in excel file

inmatrix = xlsread(infile,sheetn);
[~,titles,~] = xlsread(infile,sheetn);
titles(1)=[]; %remove first cell
sz = size(inmatrix);
n = sz(2)-1;
conc = inmatrix(2:end,1);
max_conc = max(conc);
CSP = inmatrix(2:end,2:end);
P = 0.015; %insert protein concentration in M

%fitting
%initial parameters [max chem shift, protein concentration, Kd]
x0 = [0.05, 0.00001];
lb = [0, 0]; % lower bounds
ub = [inf, inf]; % upper bounds
x_fit = [0:0.00001:max_conc];

out_param = zeros(n, 2);
out_ci = zeros(n, 2);
out_curve = cell(1,n);
out_fit = zeros(n, 1);
%fitting for each curve
for ni=1:n
    y1 = CSP(:,ni);
    [param, F2, fitresid, jacobian] = fitting(x0,lb,ub, conc,P, y1);
    out_param(ni,1:2) = param;
    paramfit = sig(param(1), P, param(2), conc);
    fit_curve = sig(param(1), P, param(2), x_fit);
    out_curve{ni} = fit_curve;

    ci = nlparci(param,fitresid,'jacobian',jacobian,'alpha',0.32);
    cie = param - ci(:,1)';
    out_ci(ni,:) = cie;

    cost_func = 'NRMSE';
    fit = goodnessOfFit(paramfit,y1,cost_func);
    out_fit(ni,1) = fit;

    figure('visible','off')
    hold on
    scatter(conc,y1)
    plot(x_fit,fit_curve)
    title(titles{ni})
    xlabel('Concentration (mM)')
    ylabel('CSP (ppm)')
    ylim([0 0.1])
    imagefile_n = strcat(imagefile,titles{ni},'.emf');
    saveas(gcf,imagefile_n)
    close gcf
end
```

```
% export fit parameters to xls file
xls_param_header = ["Name", "D", "D CI","Kd", "Kd CI", "nrmse fit"];
xlswrite(outfile,xls_param_header,'Fitted Values','A1');
xlswrite(outfile,titles,'Fitted Values','A2');
xlswrite(outfile,out_param(:,1),'Fitted Values','B2');
xlswrite(outfile,out_ci(:,1),'Fitted Values','C2');
xlswrite(outfile,out_param(:,2),'Fitted Values','D2');
xlswrite(outfile,out_ci(:,2),'Fitted Values','E2');
xlswrite(outfile,out_fit,'Fitted Values','F2');

%% functions

function F=sig(D,P,Kd,x)
out = zeros(length(x),1);
for xi=1:length(x)
    out(xi) = D/(2*P)*((P+x(xi)+Kd)-sqrt((P+x(xi)+Kd)*(P+x(xi)+Kd)-4*P*x(xi)));
end
F = out;
end

function F=chi(D,P,Kd,x,data)
%call function for fit
fit = sig(D,P,Kd,x);
F = fit - data;
end

function [F, F2, F3,F4] =fitting(x0,lb,ub, x,P, data)
fun = @(x0)(chi(x0(1),P,x0(2),x,data));
options = optimoptions('lsqnonlin','FiniteDifferenceType', 'central','Algorithm',
'trust-region-reflective', 'OptimalityTolerance', 1e-12,'TolFun',1e-12,'TolX', 1e-12,
'MaxFunEvals', 2000, 'Display', 'iter');
[param, resnorm,residual,~,~,~,jacobian] = lsqnonlin(fun, x0, lb, ub,options);
F=param;
F2 = resnorm;
F3 = residual;
F4 = jacobian;
end
```

BLANK PAGE

8 References

1. Birch, J. R.; Onakunle, Y., Biopharmaceutical proteins. *Therapeutic proteins: Methods and protocols* **2005**, 1-16.
2. Walsh, G., Biopharmaceutical benchmarks 2018. *Nat. Biotechnol* **2018,** *36*, 1136-1145.
3. Walsh, G., Biopharmaceuticals and biotechnology medicines: an issue of nomenclature. *European Journal of Pharmaceutical Sciences* **2002,** *15* (2), 135-138.
4. Reddy, B. B. K.; Karunakar, A., Biopharmaceutics classification system: a regulatory approach. *Dissolution Technol* **2011,** *18* (1), 31-37.
5. Sekhon, B. S., Biopharmaceuticals: an overview. *Thai J. Pharm. Sci* **2010,** *34* (34), 1-19.
6. Rader, R. A., (Re) defining biopharmaceutical. *Nature biotechnology* **2008,** *26* (7), 743-751.
7. Wang, W., Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International journal of pharmaceutics* **1999,** *185* (2), 129-188.
8. Mitragotri, S.; Burke, P. A.; Langer, R., Overcoming the challenges in administering biopharmaceuticals: formulation and delivery strategies. *Nature reviews Drug discovery* **2014,** *13* (9), 655-672.
9. Schellekens, H., Immunogenicity of therapeutic proteins: clinical implications and future prospects. *Clinical therapeutics* **2002,** *24* (11), 1720-1740.
10. Schellekens, H.; Jiskoot, W., Immunogenicity of therapeutic proteins. In *Pharmaceutical biotechnology*, Springer: 2013; pp 133-141.
11. Ma, H.; O'Kennedy, R., The structure of natural and recombinant antibodies. In *Peptide Antibodies*, Springer: 2015; pp 7-11.
12. Schroeder Jr, H. W.; Cavacini, L., Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology* **2010,** *125* (2), S41-S52.
13. Tao, M.-H.; Morrison, S. L., Studies of aglycosylated chimeric mouse-human IgG. Role of carbohydrate in the structure and effector functions mediated by the human IgG constant region. *The Journal of Immunology* **1989,** *143* (8), 2595-2601.
14. Wright, A.; Morrison, S. L., Effect of glycosylation on antibody function: implications for genetic engineering. *Trends in biotechnology* **1997,** *15* (1), 26-32.
15. Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S., Antibody structure, instability, and formulation. *Journal of pharmaceutical sciences* **2007,** *96* (1), 1-26.
16. Polonelli, L.; Pontón, J.; Elguezabal, N.; Moragues, M. D.; Casoli, C.; Pilotti, E.; Ronzi, P.; Dobroff, A. S.; Rodrigues, E. G.; Juliano, M. A., Antibody complementarity-determining regions (CDRs) can display differential antimicrobial, antiviral and antitumor activities. *PLoS One* **2008,** *3* (6), e2371.
17. Walsh, G., Biopharmaceutical benchmarks 2014. *Nature biotechnology* **2014,** *32* (10), 992-1000.
18. Liu, J. K., The history of monoclonal antibody development–progress, remaining challenges and future innovations. *Annals of Medicine and Surgery* **2014,** *3* (4), 113-116.
19. LLP, K. S. I. *Monoclonal Antibodies Market - Forecasts from 2018 to 2023*; ID: 4618236; 2018.
20. Cui, Y.; Cui, P.; Chen, B.; Li, S.; Guan, H., Monoclonal antibodies: formulations of marketed products and recent advances in novel delivery system. *Drug development and industrial pharmacy* **2017,** *43* (4), 519-530.
21. Jiskoot, W.; Randolph, T. W.; Volkin, D. B.; Middaugh, C. R.; Schoneich, C.; Winter, G.; Friess, W.; Crommelin, D. J.; Carpenter, J. F., Protein instability and immunogenicity: roadblocks to clinical application of injectable protein delivery systems for sustained release. *J Pharm Sci* **2012,** *101* (3), 946-54.
22. Hermeling, S.; Crommelin, D. J.; Schellekens, H.; Jiskoot, W., Structure-immunogenicity relationships of therapeutic proteins. *Pharm Res* **2004,** *21* (6), 897-903.
23. Ratanji, K. D.; Derrick, J. P.; Dearman, R. J.; Kimber, I., Immunogenicity of therapeutic proteins: Influence of aggregation. *J Immunotoxicol* **2014,** *11* (2), 99-109.
24. Better, M.; Chang, C. P.; Robinson, R. R.; Horwitz, A. H., Escherichia coli secretion of an active chimeric antibody fragment. *Science* **1988,** *240* (4855), 1041-1043.
25. Nelson, A. L.; Reichert, J. M., Development trends for therapeutic antibody fragments. *Nature biotechnology* **2009,** *27* (4), 331.
26. Skerra, A.; Pluckthun, A., Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science* **1988,** *240* (4855), 1038-1041.
27. Holliger, P.; Hudson, P. J., Engineered antibody fragments and the rise of single domains. *Nature biotechnology* **2005,** *23* (9), 1126.
28. Siontorou, C. G., Nanobodies as novel agents for disease diagnosis and therapy. *International journal of nanomedicine* **2013,** *8*, 4215.

29. Holliger, P.; Prospero, T.; Winter, G., " Diabodies": small bivalent and bispecific antibody fragments. *Proceedings of the National Academy of Sciences* **1993,** *90* (14), 6444-6448.
30. Sedykh, S. E.; Prinz, V. V.; Buneva, V. N.; Nevinsky, G. A., Bispecific antibodies: design, therapy, perspectives. *Drug design, development and therapy* **2018,** *12*, 195.
31. Wu, X.; Demarest, S. J., Building blocks for bispecific and trispecific antibodies. *Methods* **2019,** *154*, 3-9.
32. Gronwall, C.; Stahl, S., Engineered affinity proteins-Generation and applications. *J Biotechnol* **2009,** *140* (3-4), 254-269.
33. Binz, H. K.; Amstutz, P.; Pluckthun, A., Engineering novel binding proteins from nonimmunoglobulin domains. *Nature Biotechnology* **2005,** *23* (10), 1257-1268.
34. Löfblom, J.; Frejd, F. Y.; Ståhl, S., Non-immunoglobulin based protein scaffolds. *Current opinion in biotechnology* **2011,** *22* (6), 843-848.
35. Nygren, P.-Å.; Skerra, A., Binding proteins from alternative scaffolds. *Journal of immunological methods* **2004,** *290* (1-2), 3-28.
36. Woodman, R.; Yeh, J. T.-H.; Laurenson, S.; Ferrigno, P. K., Design and validation of a neutral protein scaffold for the presentation of peptide aptamers. *Journal of molecular biology* **2005,** *352* (5), 1118-1133.
37. Colas, P.; Cohen, B.; Jessen, T.; Grishina, I., Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **1996,** *380* (6574), 548.
38. Hoppe-Seyler, F.; Crnkovic-Mertens, I.; Tomai, E.; Butz, K., Peptide aptamers: specific inhibitors of protein function. *Current molecular medicine* **2004,** *4* (5), 529-538.
39. Lopez-Ochoa, L.; Ramirez-Prado, J.; Hanley-Bowdoin, L., Peptide aptamers that bind to a geminivirus replication protein interfere with viral replication in plant cells. *Journal of virology* **2006,** *80* (12), 5841-5853.
40. Bedford, R.; Tiede, C.; Hughes, R.; Curd, A.; McPherson, M.; Peckham, M.; Tomlinson, D. C., Alternative reagents to antibodies in imaging applications. *Biophysical reviews* **2017,** *9* (4), 299-308.
41. Davis, J. J.; Tkac, J.; Humphreys, R.; Buxton, A. T.; Lee, T. A.; Ko Ferrigno, P., Peptide aptamers in label-free protein detection: 2. Chemical optimization and detection of distinct protein isoforms. *Analytical chemistry* **2009,** *81* (9), 3314-3320.
42. Ferrigno, P. K., Non-antibody protein-based biosensors. *Essays Biochem* **2016,** *60* (1), 19-25.
43. Justino, C. I.; Freitas, A. C.; Pereira, R.; Duarte, A. C.; Santos, T. A. R., Recent developments in recognition elements for chemical sensors and biosensors. *TrAC Trends in Analytical Chemistry* **2015,** *68*, 2-17.
44. Johnson, S.; Evans, D.; Laurenson, S.; Paul, D.; Davies, A. G.; Ko Ferrigno, P.; Wälti, C., Surface-immobilized peptide aptamers as probe molecules for protein detection. *Analytical chemistry* **2008,** *80* (4), 978-983.
45. Bardou, C.; Borie, C.; Bickle, M.; Rudkin, B. B.; Colas, P., Peptide aptamers for small molecule drug discovery. In *Nucleic Acid and Peptide Aptamers*, Springer: 2009; pp 373-388.
46. Zhou, J.; Rossi, J., Aptamers as targeted therapeutics: current potential and challenges. *Nature reviews Drug discovery* **2017,** *16* (3), 181.
47. Linhult, M.; Gülich, S.; Gräslund, T.; Simon, A.; Karlsson, M.; Sjöberg, A.; Nord, K.; Hober, S., Improving the tolerance of a protein a analogue to repeated alkaline exposures using a bypass mutagenesis approach. *Proteins: Structure, Function, and Bioinformatics* **2004,** *55* (2), 407-416.
48. Koide, A.; Bailey, C. W.; Huang, X.; Koide, S., The fibronectin type III domain as a scaffold for novel binding proteins1. *Journal of molecular biology* **1998,** *284* (4), 1141-1151.
49. Batori, V.; Koide, A.; Koide, S., Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein engineering* **2002,** *15* (12), 1015-1020.
50. Tiede, C.; Tang, A. A.; Deacon, S. E.; Mandal, U.; Nettleship, J. E.; Owen, R. L.; George, S. E.; Harrison, D. J.; Owens, R. J.; Tomlinson, D. C., Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Engineering, Design and Selection* **2014,** *27* (5), 145-155.
51. Hoffmann, T.; Stadler, L. K. J.; Busby, M.; Song, Q.; Buxton, A. T.; Wagner, S. D.; Davis, J. J.; Ko Ferrigno, P., Structure-function studies of an engineered scaffold protein derived from stefin A. I: Development of the SQM variant. *Protein Engineering, Design & Selection* **2010,** *23* (5), 403-413.
52. Gilbreth, R. N.; Koide, S., Structural insights for engineering binding proteins based on non-antibody scaffolds. *Current opinion in structural biology* **2012,** *22* (4), 413-420.
53. Klevenz, B.; Butz, K.; Hoppe-Seyler, F., Peptide aptamers: exchange of the thioredoxin-A scaffold by alternative platform proteins and its influence on target protein binding. *Cellular and molecular life sciences* **2002,** *59* (11), 1993-1998.
54. Nygren, P. Å., Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *Febs Journal* **2008,** *275* (11), 2668-2676.

55. Binz, H. K.; Stumpp, M. T.; Forrer, P.; Amstutz, P.; Plückthun, A., Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of molecular biology* **2003,** *332* (2), 489-503.
56. Stadler, L. K.; Hoffmann, T.; Tomlinson, D. C.; Song, Q.; Lee, T.; Busby, M.; Nyathi, Y.; Gendra, E.; Tiede, C.; Flanagan, K.; Cockell, S. J.; Wipat, A.; Harwood, C.; Wagner, S. D.; Knowles, M. A.; Davis, J. J.; Keegan, N.; Ferrigno, P. K., Structure-function studies of an engineered scaffold protein derived from Stefin A. II: Development and applications of the SQT variant. *Protein engineering, design & selection : PEDS* **2011,** *24* (9), 751-63.
57. Martin, J. R.; Craven, J. C.; Jerala, R.; Kroon-Žitko, L.; Žerovnik, E.; Turk, V.; Waltho, J. P., The three-dimensional solution structure of human stefin A. *Journal of molecular biology* **1995,** *246* (2), 331-343.
58. Jenko, S.; Dolenc, I.; Gunčar, G.; Doberšek, A.; Podobnik, M.; Turk, D., Crystal structure of Stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo-and exopeptidases. *Journal of molecular biology* **2003,** *326* (3), 875-885.
59. Bode, W.; Engh, R.; Musil, D.; Thiele, U.; Huber, R.; Karshikov, A.; Brzin, J.; Kos, J.; Turk, V., The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *The EMBO journal* **1988,** *7* (8), 2593.
60. Edwards, T. A.; Miles, J. A.; Hobor, F.; Taylor, J. L.; Tiede, C.; Rowell, P. L.; Trinh, C. H.; Jackson, B. R.; Nadat, F. A.; Kyle, H. F., Selective Affimers Recognize BCL-2 Family Proteins Through Non-Canonical Structural Motifs. *bioRxiv* **2019**, 651364.
61. Jenkins, T. P.; Fryer, T.; Dehli, R. I.; Jürgensen, J. A.; Fuglsang-Madsen, A.; Føns, S.; Laustsen, A. H., Toxin neutralization using alternative binding proteins. *Toxins* **2019,** *11* (1), 53.
62. van Wijk, S. J.; Fulda, S.; Dikic, I.; Heilemann, M., Visualizing ubiquitination in mammalian cells. *EMBO reports* **2019,** *20* (2).
63. Hesketh, E. L.; Tiede, C.; Adamson, H.; Adams, T. L.; Byrne, M. J.; Meshcheriakova, Y.; Kruse, I.; McPherson, M. J.; Lomonossoff, G. P.; Tomlinson, D. C., Affimer reagents as tools in diagnosing plant virus diseases. *Scientific reports* **2019,** *9* (1), 7524.
64. Adamson, H.; Nicholl, A.; Tiede, C.; Tang, A. A.; Davidson, A.; Curd, H.; Wignall, A.; Ford, R.; Nuttall, J.; McPherson, M. J., Affimers as anti-idiotypic affinity reagents for pharmacokinetic analysis of biotherapeutics. *BioTechniques* **2019,** (0).
65. Tiede, C.; Bedford, R.; Heseltine, S. J.; Smith, G.; Wijetunga, I.; Ross, R.; AlQallaf, D.; Roberts, A. P.; Balls, A.; Curd, A., Affimer proteins are versatile and renewable affinity reagents. *Elife* **2017,** *6*, e24903.
66. Pace, C. N., Measuring and increasing protein stability. *Trends in biotechnology* **1990,** *8*, 93-98.
67. Bischof, J. C.; He, X., Thermal stability of proteins. *Annals of the New York Academy of Sciences* **2005,** *1066*, 12-33.
68. 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.
69. Manning, M. C.; Chou, D. K.; Murphy, B. M.; Payne, R. W.; Katayama, D. S., Stability of protein pharmaceuticals: an update. *Pharm Res* **2010,** *27* (4), 544-575.
70. Liu, H.; Gaza-Bulseco, G.; Faldu, D.; Chumsae, C.; Sun, J., Heterogeneity of monoclonal antibodies. *Journal of pharmaceutical sciences* **2008,** *97* (7), 2426-2447.
71. Houchin, M.; Topp, E., Chemical degradation of peptides and proteins in PLGA: a review of reactions and mechanisms. *Journal of pharmaceutical sciences* **2008,** *97* (7), 2395-2404.
72. Snape, T. J.; Astles, A. M.; Davies, J., Understanding the chemical basis of drug stability and degradation. *Pharmaceutical Journal* **2010,** *285* (7622), 416-417.
73. Walters, R. H.; Bhatnagar, B.; Tchessalov, S.; Izutsu, K.-I.; Tsumoto, K.; Ohtake, S., Next generation drying technologies for pharmaceutical applications. *Journal of pharmaceutical sciences* **2014,** *103* (9), 2673-2695.
74. Shire, S. J.; Shahrokh, Z.; Liu, J., Challenges in the development of high protein concentration formulations. *Journal of pharmaceutical sciences* **2004,** *93* (6), 1390-1402.
75. Arzenšek, D.; Kuzman, D.; Podgornik, R., Colloidal interactions between monoclonal antibodies in aqueous solutions. *Journal of colloid and interface science* **2012,** *384* (1), 207-216.
76. Roberts, C. J., Therapeutic protein aggregation: mechanisms, design, and control. *Trends Biotechnol* **2014,** *32* (7), 372-80.
77. Mahler, H. C.; Friess, W.; Grauschopf, U.; Kiese, S., Protein aggregation: pathways, induction factors and analysis. *Journal of pharmaceutical sciences* **2009,** *98* (9), 2909-2934.
78. Wang, W.; Nema, S.; Teagarden, D., Protein aggregation—Pathways and influencing factors. *International journal of pharmaceutics* **2010,** *390* (2), 89-99.
79. Fink, A. L., Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and design* **1998,** *3* (1), R9-R23.

80. Roberts, C. J., Non-native protein aggregation kinetics. *Biotechnology and bioengineering* **2007,** *98* (5), 927-938.
81. Wang, W.; Roberts, C. J., Protein aggregation - Mechanisms, detection, and control. *Int J Pharm* **2018,** *550* (1-2), 251-268.
82. Wu, H.; Kroe-Barrett, R.; Singh, S.; Robinson, A. S.; Roberts, C. J., Competing aggregation pathways for monoclonal antibodies. *FEBS letters* **2014,** *588* (6), 936-41.
83. Rousseau, F.; Schymkowitz, J. W.; Itzhaki, L. S., The unfolding story of three-dimensional domain swapping. *Structure* **2003,** *11* (3), 243-251.
84. Tsitsanou, K. E.; Drakou, C. E.; Thireou, T.; Gruber, A. V.; Kythreoti, G.; Azem, A.; Fessas, D.; Eliopoulos, E.; Iatrou, K.; Zographos, S. E., Crystal and solution studies of the "Plus-C" odorant-binding protein 48 from Anopheles gambiae control of binding specificity through three-dimensional domain swapping. *Journal of Biological Chemistry* **2013,** *288* (46), 33427-33438.
85. Bennett, M. J.; Sawaya, M. R.; Eisenberg, D., Deposition diseases and 3D domain swapping. *Structure* **2006,** *14* (5), 811-824.
86. Palsdottir, A.; Snorradottir, A.; Thorsteinsson, L., Hereditary cystatin C amyloid angiopathy: genetic, clinical, and pathological aspects. *Brain pathology* **2006,** *16* (1), 55-59.
87. Orlikowska, M.; Jankowska, E.; Kolodziejczyk, R.; Jaskolski, M.; Szymanska, A., Hinge-loop mutation can be used to control 3D domain swapping and amyloidogenesis of human cystatin C. *J Struct Biol* **2011,** *173* (2), 406-13.
88. Rodziewicz-Motowidło, S.; Iwaszkiewicz, J.; Sosnowska, R.; Czaplewska, P.; Sobolewski, E.; Szymańska, A.; Stachowiak, K.; Liwo, A., The role of the Val57 amino-acid residue in the hinge loop of the human cystatin C. Conformational studies of the beta2-L1-beta3 segments of wild-type human cystatin C and its mutants. *Biopolymers* **2009,** *91* (5), 373-383.
89. O'Neill, J. W.; Kim, D. E.; Johnsen, K.; Baker, D.; Zhang, K. Y., Single-site mutations induce 3D domain swapping in the B1 domain of protein L from Peptostreptococcus magnus. *Structure* **2001,** *9* (11), 1017-1027.
90. Nandwani, N.; Surana, P.; Negi, H.; Mascarenhas, N. M.; Udgaonkar, J. B.; Das, R.; Gosavi, S., A five-residue motif for the design of domain swapping in proteins. *Nature communications* **2019,** *10* (1), 452.
91. Ha, J.-H.; Karchin, J. M.; Walker-Kopp, N.; Castañeda, C. A.; Loh, S. N., Engineered domain swapping as an on/off switch for protein function. *Chemistry & biology* **2015,** *22* (10), 1384-1393.
92. Liu, J.; Nguyen, M. D.; Andya, J. D.; Shire, S. J., Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. *J Pharm Sci* **2005,** *94* (9), 1928-40.
93. Esfandiary, R.; Parupudi, A.; Casas-Finet, J.; Gadre, D.; Sathish, H., Mechanism of reversible self-association of a monoclonal antibody: role of electrostatic and hydrophobic interactions. *Journal of pharmaceutical sciences* **2015,** *104* (2), 577-586.
94. Arora, J.; Hu, Y.; Esfandiary, R.; Sathish, H. A.; Bishop, S. M.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. In *Charge-mediated Fab-Fc interactions in an IgG1 antibody induce reversible self-association, cluster formation, and elevated viscosity*, MAbs, Taylor & Francis: 2016; pp 1561-1574.
95. Stradner, A.; Sedgwick, H.; Cardinaux, F.; Poon, W. C.; Egelhaaf, S. U.; Schurtenberger, P., Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature* **2004,** *432* (7016), 492.
96. Cardinaux, F.; Zaccarelli, E.; Stradner, A.; Bucciarelli, S.; Farago, B.; Egelhaaf, S. U.; Sciortino, F.; Schurtenberger, P., Cluster-driven dynamical arrest in concentrated lysozyme solutions. *The Journal of Physical Chemistry B* **2011,** *115* (22), 7227-7237.
97. Pathak, J. A.; Sologuren, R. R.; Narwal, R., Do clustering monoclonal antibody solutions really have a concentration dependence of viscosity? *Biophysical journal* **2013,** *104* (4), 913-923.
98. Fukuda, M.; Moriyama, C.; Yamazaki, T.; Imaeda, Y.; Koga, A., Quantitative correlation between viscosity of concentrated MAb solutions and particle size parameters obtained from small-angle x-ray scattering. *Pharm Res* **2015,** *32* (12), 3803-3812.
99. Austerberry, J. I.; Dajani, R.; Panova, S.; Roberts, D.; Golovanov, A. P.; Pluen, A.; van der Walle, C. F.; Uddin, S.; Warwicker, J.; Derrick, J. P., The effect of charge mutations on the stability and aggregation of a human single chain Fv fragment. *European Journal of Pharmaceutics and Biopharmaceutics* **2017,** *115*, 18-30.
100. Galm, L.; Amrhein, S.; Hubbuch, J., Predictive approach for protein aggregation: Correlation of protein surface characteristics and conformational flexibility to protein aggregation propensity. *Biotechnology and bioengineering* **2017,** *114* (6), 1170-1183.
101. Mehta, S. B.; Bee, J. S.; Randolph, T. W.; Carpenter, J. F., Partial unfolding of a monoclonal antibody: role of a single domain in driving protein aggregation. *Biochemistry* **2014,** *53* (20), 3367-3377.
102. Andrews, J. M.; Roberts, C. J., Non-native aggregation of α-chymotrypsinogen occurs through nucleation and growth with competing nucleus sizes and negative activation energies. *Biochemistry* **2007,** *46* (25), 7558-7571.

103. Watzky, M. A.; Finke, R. G., Transition metal nanocluster formation kinetic and mechanistic studies. A new mechanism when hydrogen is the reductant: slow, continuous nucleation and fast autocatalytic surface growth. *Journal of the American Chemical Society* **1997,** *119* (43), 10382-10400.
104. Bentea, L.; Watzky, M. A.; Finke, R. G., Sigmoidal nucleation and growth curves across nature fit by the Finke–Watzky model of slow continuous nucleation and autocatalytic growth: explicit formulas for the lag and growth times plus other key insights. *The Journal of Physical Chemistry C* **2017,** *121* (9), 5302-5312.
105. Iashchishyn, I. A.; Sulskis, D.; Nguyen Ngoc, M.; Smirnovas, V.; Morozova-Roche, L. A., Finke–Watzky Two-Step Nucleation–Autocatalysis Model of S100A9 Amyloid Formation: Protein Misfolding as "Nucleation" Event. *ACS Chemical Neuroscience* **2017,** *8* (10), 2152-2158.
106. Morris, A. M.; Watzky, M. A.; Agar, J. N.; Finke, R. G., Fitting neurological protein aggregation kinetic data via a 2-step, minimal/"Ockham's Razor" model: The Finke− Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry* **2008,** *47* (8), 2413-2427.
107. Meric, G.; Robinson, A. S.; Roberts, C. J., Driving forces for nonnative protein aggregation and approaches to predict aggregation-prone regions. *Annual review of chemical and biomolecular engineering* **2017,** *8*, 139-159.
108. Elvin, J. G.; Couston, R. G.; van der Walle, C. F., Therapeutic antibodies: market considerations, disease targets and bioprocessing. *International journal of pharmaceutics* **2013,** *440* (1), 83-98.
109. Buck, P. M.; Kumar, S.; Singh, S. K., On the role of aggregation prone regions in protein evolution, stability, and enzymatic catalysis: insights from diverse analyses. *PLoS computational biology* **2013,** *9* (10), e1003291.
110. Brubaker, W. D.; Freites, J. A.; Golchert, K. J.; Shapiro, R. A.; Morikis, V.; Tobias, D. J.; Martin, R. W., Separating instability from aggregation propensity in γS-crystallin variants. *Biophysical journal* **2011,** *100* (2), 498-506.
111. Kazlauskas, R., Engineering more stable proteins. *Chemical Society Reviews* **2018,** *47* (24), 9026-9045.
112. Buck, P. M.; Kumar, S.; Singh, S. K. In *Consequences of glycan truncation on Fc structural integrity*, MAbs, Taylor & Francis: 2013; pp 904-916.
113. Wang, X.; Kumar, S.; Buck, P. M.; Singh, S. K., Impact of deglycosylation and thermal stress on conformational stability of a full length murine IgG2a monoclonal antibody: observations from molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **2013,** *81* (3), 443-460.
114. Treuheit, M. J.; Kosky, A. A.; Brems, D. N., Inverse relationship of protein concentration and aggregation. *Pharm Res* **2002,** *19* (4), 511-516.
115. Gokarn, Y. R.; Fesinmeyer, R. M.; Saluja, A.; Razinkov, V.; Chase, S. F.; Laue, T. M.; Brems, D. N., Effective charge measurements reveal selective and preferential accumulation of anions, but not cations, at the protein surface in dilute salt solutions. *Protein Science* **2011,** *20* (3), 580-587.
116. Majumdar, R.; Manikwar, P.; Hickey, J. M.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D., Effects of salts from the Hofmeister series on the conformational stability, aggregation propensity, and local flexibility of an IgG1 monoclonal antibody. *Biochemistry* **2013,** *52* (19), 3376-3389.
117. Bye, J. W.; Baxter, N. J.; Hounslow, A. M.; Falconer, R. J.; Williamson, M. P., Molecular Mechanism for the Hofmeister Effect Derived from NMR and DSC Measurements on Barnase. *ACS Omega* **2016,** *1* (4), 669-679.
118. Agrawal, N. J.; Kumar, S.; Wang, X.; Helk, B.; Singh, S. K.; Trout, B. L., Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions. *Journal of pharmaceutical sciences* **2011,** *100* (12), 5081-5095.
119. Marshall, S. A.; Lazar, G. A.; Chirino, A. J.; Desjarlais, J. R., Rational design and engineering of therapeutic proteins. *Drug discovery today* **2003,** *8* (5), 212-221.
120. Culajay, J. F.; Blaber, S. I.; Khurana, A.; Blaber, M., Thermodynamic characterization of mutants of human fibroblast growth factor 1 with an increased physiological half-life. *Biochemistry* **2000,** *39* (24), 7153-7158.
121. Dombkowski, A. A.; Sultana, K. Z.; Craig, D. B., Protein disulfide engineering. *FEBS letters* **2014,** *588* (2), 206-12.
122. Courtois, F.; Schneider, C. P.; Agrawal, N. J.; Trout, B. L., Rational design of biobetters with enhanced stability. *Journal of pharmaceutical sciences* **2015,** *104* (8), 2433-2440.
123. Karow, A. R.; Bahrenburg, S.; Garidel, P., Buffer capacity of biologics—from buffer salts to buffering by antibodies. *Biotechnology progress* **2013,** *29* (2), 480-492.
124. Sek, D., Breaking old habits: Moving away from commonly used buffers in pharmaceuticals. *Eur Pharm Rev* **2012,** *7*, 37-41.

125. Kang, J.; Lin, X.; Penera, J., Rapid formulation development for monoclonal antibodies. *BioProcess Int* **2016,** *14* (4), 40.
126. Svilenov, H.; Markoja, U.; Winter, G., Isothermal chemical denaturation as a complementary tool to overcome limitations of thermal differential scanning fluorimetry in predicting physical stability of protein formulations. *European Journal of Pharmaceutics and Biopharmaceutics* **2018,** *125*, 106-113.
127. Katdare, A.; Chaubal, M., *Excipient development for pharmaceutical, biotechnology, and drug delivery systems*. CRC Press: 2006.
128. Maggio, E., Use of excipients to control aggregation in peptide and protein formulations. *Journal of Excipients and Food Chemicals* **2010,** *1* (2), 40-49.
129. Hofmeister, F., Zur Lehre von der Wirkung der Salze Nauyen-Schmiedebergs. *Arch Exp Pathol Pharmacol* **1888,** *24*, 247-260.
130. Kunz, W.; Henle, J.; Ninham, B., About the science of the effect of salts: Franz Hofmeister's historical papers. *Curr Opin Coll Interface Sci* **2004,** *9* (1-2), 19-37.
131. Von Hippel, P. H.; Wong, K.-Y., On the Conformational Stability of Globular Proteins the effects of various electrolytes and nonelectrolytes on the thermal ribonuclease transition. *Journal of Biological Chemistry* **1965,** *240* (10), 3909-3923.
132. Bye, J. W.; Falconer, R. J., Thermal stability of lysozyme as a function of ion concentration: a reappraisal of the relationship between the Hofmeister series and protein stability. *Protein Science* **2013,** *22* (11), 1563-1570.
133. Wang, S.; Zhang, N.; Hu, T.; Dai, W.; Feng, X.; Zhang, X.; Qian, F., Viscosity-lowering effect of amino acids and salts on highly concentrated solutions of two IgG1 monoclonal antibodies. *Molecular pharmaceutics* **2015,** *12* (12), 4478-4487.
134. Hong, T.; Iwashita, K.; Shiraki, K., Viscosity control of protein solution by small solutes: a review. *Current Protein and Peptide Science* **2018,** *19* (8), 746-758.
135. Arakawa, T.; Timasheff, S. N., Stabilization of protein structure by sugars. *Biochemistry* **1982,** *21* (25), 6536-6544.
136. Arakawa, T.; Timasheff, S., The stabilization of proteins by osmolytes. *Biophysical journal* **1985,** *47* (3), 411-414.
137. Lee, J. C.; Timasheff, S. N., The stabilization of proteins by sucrose. *Journal of Biological Chemistry* **1981,** *256* (14), 7193-7201.
138. Kim, Y. S.; Jones, L. S.; Dong, A.; Kendrick, B. S.; Chang, B. S.; Manning, M. C.; Randolph, T. W.; Carpenter, J. F., Effects of sucrose on conformational equilibria and fluctuations within the native-state ensemble of proteins. *Protein Science* **2003,** *12* (6), 1252-1261.
139. Abbas, S. A.; Sharma, V. K.; Patapoff, T. W.; Kalonia, D. S., Opposite effects of polyols on antibody aggregation: thermal versus mechanical stresses. *Pharm Res* **2012,** *29* (3), 683-694.
140. Khan, T. A.; Mahler, H.-C.; Kishore, R. S., Key interactions of surfactants in therapeutic protein formulations: a review. *European journal of pharmaceutics and biopharmaceutics* **2015,** *97*, 60-67.
141. Chen, B.; Bautista, R.; Yu, K.; Zapata, G. A.; Mulkerrin, M. G.; Chamow, S. M., Influence of histidine on the stability and physical properties of a fully human antibody in aqueous and solid forms. *Pharm Res* **2003,** *20* (12), 1952-1960.
142. Lam, X. M.; Yang, J. Y.; Cleland, J. L., Antioxidants for prevention of methionine oxidation in recombinant monoclonal antibody HER2. *Journal of pharmaceutical sciences* **1997,** *86* (11), 1250-1255.
143. Kamerzell, T. J.; Esfandiary, R.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B., Protein–excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. *Advanced drug delivery reviews* **2011,** *63* (13), 1118-1159.
144. Maeder, W.; Lieby, P.; Sebald, A.; Spycher, M.; Pedrussio, R.; Bolli, R., Local tolerance and stability up to 24 months of a new 20% proline-stabilized polyclonal immunoglobulin for subcutaneous administration. *Biologicals* **2011,** *39* (1), 43-49.
145. Shukla, D.; Trout, B. L., Interaction of arginine with proteins and the mechanism by which it inhibits aggregation. *The Journal of Physical Chemistry B* **2010,** *114* (42), 13426-13438.
146. Shukla, D.; Trout, B. L., Preferential interaction coefficients of proteins in aqueous arginine solutions and their molecular origins. *The Journal of Physical Chemistry B* **2010,** *115* (5), 1243-1253.
147. Arakawa, T.; Ejima, D.; Tsumoto, K.; Obeyama, N.; Tanaka, Y.; Kita, Y.; Timasheff, S. N., Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. *Biophysical chemistry* **2007,** *127* (1-2), 1-8.
148. Sudrik, C.; Cloutier, T.; Pham, P.; Samra, H. S.; Trout, B. L. In *Preferential interactions of trehalose, L-arginine. HCl and sodium chloride with therapeutically relevant IgG1 monoclonal antibodies*, MAbs, Taylor & Francis: 2017; pp 1155-1168.

149. Kheddo, P.; Tracka, M.; Armer, J.; Dearman, R. J.; Uddin, S.; van der Walle, C. F.; Golovanov, A. P., The effect of arginine glutamate on the stability of monoclonal antibodies in solution. *International journal of pharmaceutics* **2014,** *473* (1-2), 126-133.
150. Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L.-Y., A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society* **2004,** *126* (29), 8933-8939.
151. Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.; Fedorov, O. Y.; Finerty, P. J.; Wasney, G. A.; Yeung, R.; Arrowsmith, C.; Ball, L. J.; Berglund, H., Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences* **2006,** *103* (43), 15835-15840.
152. Bhambhani, A.; Thakkar, S.; Joshi, S. B.; Middaugh, C. R., A formulation method to improve the physical stability of macromolecular-based drug products. In *Therapeutic Protein Drug Products*, Elsevier: 2012; pp 13-45.
153. Mahler, H.-C.; Jiskoot, W., *Analysis of aggregates and particles in protein pharmaceuticals*. John Wiley & Sons: 2011.
154. Makhatadze, G. I., Measuring protein thermostability by differential scanning calorimetry. *Current protocols in protein science* **1998,** *12* (1), 7.9. 1-7.9. 14.
155. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2007,** *2* (9), 2212.
156. Haffke, M.; Rummel, G.; Boivineau, J.; Münch, A.; Jaakola, V.-P., nanoDSF: label-free thermal unfolding assay of G-protein-coupled receptors for compound screening and buffer composition optimization. *Nanotemper Application Note NT-PR-007* **2016,** (NT-PR-), 1-4.
157. Brader, M. L.; Estey, T.; Bai, S.; Alston, R. W.; Lucas, K. K.; Lantz, S.; Landsman, P.; Maloney, K. M., Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies. *Molecular pharmaceutics* **2015,** *12* (4), 1005-1017.
158. Greenfield, N. J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature protocols* **2006,** *1* (6), 2527.
159. Chaudhuri, R.; Cheng, Y.; Middaugh, C. R.; Volkin, D. B., High-throughput biophysical analysis of protein therapeutics to examine interrelationships between aggregate formation and conformational stability. *The AAPS journal* **2014,** *16* (1), 48-64.
160. Kenrick, S.; Some, D., The diffusion interaction parameter kd as an indicator of colloidal and thermal stability. *Wyatt Technology Corporation.–2014* **2018**.
161. Lehermayr, C.; Mahler, H.-C.; Mäder, K.; Fischer, S., Assessment of net charge and protein–protein interactions of different monoclonal antibodies. *Journal of pharmaceutical sciences* **2011,** *100* (7), 2551-2562.
162. Den Engelsman, J.; Garidel, P.; Smulders, R.; Koll, H.; Smith, B.; Bassarab, S.; Seidl, A.; Hainzl, O.; Jiskoot, W., Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. *Pharm Res* **2011,** *28* (4), 920-933.
163. Folta-Stogniew, E., Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. In *New and Emerging Proteomic Techniques*, Springer: 2006; pp 97-112.
164. Xu, R., Light scattering: A review of particle characterization applications. *Particuology* **2015,** *18*, 11-21.
165. Lebowitz, J.; Lewis, M. S.; Schuck, P., Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science* **2002,** *11* (9), 2067-2079.
166. Vendruscolo, M.; Knowles, T. P.; Dobson, C. M., Protein solubility and protein homeostasis: a generic view of protein misfolding disorders. *Cold Spring Harbor perspectives in biology* **2011,** *3* (12), a010454.
167. Roy, A.; Nair, S.; Sen, N.; Soni, N.; Madhusudhan, M., In silico methods for design of biological therapeutics. *Methods* **2017,** *131*, 33-65.
168. Cannon, D. A.; Shan, L.; Du, Q.; Shirinian, L.; Rickert, K. W.; Rosenthal, K. L.; Korade III, M.; van Vlerken-Ysla, L. E.; Buchanan, A.; Vaughan, T. J., Experimentally guided computational antibody affinity maturation with de novo docking, modelling and rational design. *PLoS computational biology* **2019,** *15* (5), e1006980.
169. Elcock, A. H.; Sept, D.; McCammon, J. A., Computer simulation of protein− protein interactions. ACS Publications: 2001.
170. Baaden, M.; Marrink, S. J., Coarse-grain modelling of protein–protein interactions. *Current opinion in structural biology* **2013,** *23* (6), 878-886.
171. Buck, P. M.; Kumar, S.; Wang, X.; Agrawal, N. J.; Trout, B. L.; Singh, S. K., Computational methods to predict therapeutic protein aggregation. In *Therapeutic Proteins*, Springer: 2012; pp 425-451.

172. Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L., A comparative study of the relationship between protein structure and β-aggregation in globular and intrinsically disordered proteins. *Journal of molecular biology* **2004,** *342* (1), 345-353.
173. Conchillo-Solé, O.; de Groot, N. S.; Avilés, F. X.; Vendrell, J.; Daura, X.; Ventura, S., AGGRESCAN: a server for the prediction and evaluation of" hot spots" of aggregation in polypeptides. *BMC bioinformatics* **2007,** *8* (1), 65.
174. Trovato, A.; Seno, F.; Tosatto, S. C., The PASTA server for protein aggregation prediction. *Protein Engineering, Design & Selection* **2007,** *20* (10), 521-523.
175. Tartaglia, G. G.; Vendruscolo, M., The Zyggregator method for predicting protein aggregation propensities. *Chemical Society Reviews* **2008,** *37* (7), 1395-1401.
176. Zambrano, R.; Jamroz, M.; Szczasiuk, A.; Pujols, J.; Kmiecik, S.; Ventura, S., AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic acids research* **2015,** *43* (W1), W306-W313.
177. Chennamsetty, N.; Helk, B.; Voynov, V.; Kayser, V.; Trout, B. L., Aggregation-prone motifs in human immunoglobulin G. *Journal of molecular biology* **2009,** *391* (2), 404-413.
178. Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L., Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences* **2009,** *106* (29), 11937-11942.
179. Sormanni, P.; Aprile, F. A.; Vendruscolo, M., The CamSol method of rational design of protein mutants with enhanced solubility. *Journal of molecular biology* **2015,** *427* (2), 478-490.
180. Sormanni, P.; Amery, L.; Ekizoglou, S.; Vendruscolo, M.; Popovic, B., Rapid and accurate in silico solubility screening of a monoclonal antibody library. *Scientific reports* **2017,** *7* (1), 1-9.
181. Hebditch, M.; Carballo-Amador, M. A.; Charonis, S.; Curtis, R.; Warwicker, J., Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **2017,** *33* (19), 3098-3100.
182. Hebditch, M.; Roche, A.; Curtis, R. A.; Warwicker, J., Models for antibody behavior in hydrophobic interaction chromatography and in self-association. *Journal of pharmaceutical sciences* **2019,** *108* (4), 1434-1441.
183. Hebditch, M.; Warwicker, J., Charge and hydrophobicity are key features in sequence-trained machine learning models for predicting the biophysical properties of clinical-stage antibodies. *PeerJ* **2019,** *7*, e8199.
184. Gentiluomo, L.; Roessner, D.; Augustijn, D.; Svilenov, H.; Kulakova, A.; Mahapatra, S.; Winter, G.; Streicher, W.; Rinnan, Å.; Peters, G. H., Application of interpretable artificial neural networks to early monoclonal antibodies development. *European Journal of Pharmaceutics and Biopharmaceutics* **2019,** *141*, 81-89.
185. Yang, Y.; Ye, Z.; Su, Y.; Zhao, Q.; Li, X.; Ouyang, D., Deep learning for in vitro prediction of pharmaceutical formulations. *Acta pharmaceutica sinica B* **2019,** *9* (1), 177-185.
186. Landin, M.; Rowe, R. C., Artificial neural networks technology to model, understand, and optimize drug formulations. In *Formulation Tools for Pharmaceutical Development*, Elsevier: 2013; pp 7-37.
187. Norman, R. A.; Ambrosetti, F.; Bonvin, A. M.; Colwell, L. J.; Kelm, S.; Kumar, S.; Krawczyk, K., Computational approaches to therapeutic antibody design: established methods and emerging trends. *Briefings in Bioinformatics* **2019**.
188. Wang, G.; Zhang, Z.-T.; Jiang, B.; Zhang, X.; Li, C.; Liu, M., Recent advances in protein NMR spectroscopy and their implications in protein therapeutics research. *Analytical and bioanalytical chemistry* **2014,** *406* (9-10), 2279-2288.
189. Mas, G.; Crublet, E.; Hamelin, O.; Gans, P.; Boisbouvier, J., Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *Journal of biomolecular NMR* **2013,** *57* (3), 251-262.
190. Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J., Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Current Opinion in Structural Biology* **2015,** *32*, 113-122.
191. Tugarinov, V.; Hwang, P. M.; Kay, L. E., Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annual review of biochemistry* **2004,** *73* (1), 107-146.
192. Riek, R.; Wider, G.; Pervushin, K.; Wüthrich, K., Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences* **1999,** *96* (9), 4918-4923.
193. Tugarinov, V.; Sprangers, R.; Kay, L. E., Line narrowing in methyl-TROSY using zero-quantum 1H-13C NMR spectroscopy. *Journal of the American Chemical Society* **2004,** *126* (15), 4921-4925.
194. Takeuchi, K.; Wagner, G., NMR studies of protein interactions. *Current opinion in structural biology* **2006,** *16* (1), 109-117.
195. Harner, M. J.; Frank, A. O.; Fesik, S. W., Fragment-based drug discovery using NMR spectroscopy. *Journal of biomolecular NMR* **2013,** *56* (2), 65-75.

196. Cala, O.; Guillière, F.; Krimm, I., NMR-based analysis of protein–ligand interactions. *Analytical and bioanalytical chemistry* **2014,** *406* (4), 943-956.
197. Bernadó, P.; Åkerud, T.; García de la Torre, J.; Akke, M.; Pons, M., Combined use of NMR relaxation measurements and hydrodynamic calculations to study protein association. Evidence for tetramers of low molecular weight protein tyrosine phosphatase in solution. *Journal of the American Chemical Society* **2003,** *125* (4), 916-923.
198. Fawzi, N. L.; Ying, J.; Ghirlando, R.; Torchia, D. A.; Clore, G. M., Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* **2011,** *480* (7376), 268.
199. Taraban, M. B.; Truong, H. C.; Feng, Y.; Jouravleva, E. V.; Anisimov, M. A.; Yu, Y. B., Water proton NMR for in situ detection of insulin aggregates. *Journal of pharmaceutical sciences* **2015,** *104* (12), 4132-4141.
200. Feng, Y.; Taraban, M. B.; Yu, Y. B., Water proton NMR—A sensitive probe for solute association. *Chemical Communications* **2015,** *51* (31), 6804-6807.
201. Taraban, M. B.; DePaz, R. A.; Lobo, B.; Yu, Y. B., Water proton NMR: a tool for protein aggregation characterization. *Analytical chemistry* **2017,** *89* (10), 5494-5502.
202. Aubin, Y.; Gingras, G.; Sauvé, S., Assessment of the three-dimensional structure of recombinant protein therapeutics by NMR fingerprinting: demonstration on recombinant human granulocyte macrophage-colony stimulation factor. *Analytical chemistry* **2008,** *80* (7), 2623-2627.
203. Ghasriani, H.; Hodgson, D. J.; Brinson, R. G.; McEwen, I.; Buhse, L. F.; Kozlowski, S.; Marino, J. P.; Aubin, Y.; Keire, D. A., Precision and robustness of 2D-NMR for structure assessment of filgrastim biosimilars. *Nature biotechnology* **2016,** *34* (2), 139.
204. Poppe, L.; Jordan, J. B.; Lawson, K.; Jerums, M.; Apostol, I.; Schnier, P. D., Profiling formulated monoclonal antibodies by 1H NMR spectroscopy. *Analytical chemistry* **2013,** *85* (20), 9623-9629.
205. Arbogast, L. W.; Brinson, R. G.; Marino, J. P., Mapping monoclonal antibody structure by 2D ^{13}C NMR at natural abundance. *Analytical chemistry* **2015,** *87* (7), 3556-3561.
206. Arbogast, L. W.; Brinson, R. G.; Formolo, T.; Hoopes, J. T.; Marino, J. P., 2D ^{1}HN, ^{15}N correlated NMR methods at natural abundance for obtaining structural maps and statistical comparability of monoclonal antibodies. *Pharm Res* **2016,** *33* (2), 462-475.
207. Brinson, R. G.; Marino, J. P.; Delaglio, F.; Arbogast, L. W.; Evans, R. M.; Kearsley, A.; Gingras, G.; Ghasriani, H.; Aubin, Y.; Pierens, G. K. In *Enabling adoption of 2D-NMR for the higher order structure assessment of monoclonal antibody therapeutics*, mAbs, Taylor & Francis: 2019; pp 94-105.
208. Japelj, B.; Ilc, G.; Marušič, J.; Senčar, J.; Kuzman, D.; Plavec, J., Biosimilar structural comparability assessment by NMR: from small proteins to monoclonal antibodies. *Scientific reports* **2016,** *6*, 32201.
209. Kheddo, P.; Cliff, M. J.; Uddin, S.; van der Walle, C. F.; Golovanov, A. P. In *Characterizing monoclonal antibody formulations in arginine glutamate solutions using ^{1}H NMR spectroscopy*, Mabs, Taylor & Francis: 2016; pp 1245-1258.
210. Kheddo, P.; Bramham, J. E.; Dearman, R. J.; Uddin, S.; van der Walle, C. F.; Golovanov, A. P., Investigating Liquid–Liquid Phase Separation of a Monoclonal Antibody Using Solution-State NMR Spectroscopy: Effect of ArgGlu and ArgHCl. *Molecular pharmaceutics* **2017,** *14* (8), 2852-2860.
211. Falk, B. T.; Liang, Y.; Bailly, M.; Raoufi, F.; Kekec, A.; Pissarnitski, D.; Feng, D.; Yan, L.; Lin, S.; Fayadat-Dilman, L., NMR Assessment of Therapeutic Peptides and Proteins: Correlations That Reveal Interactions and Motions. *ChemBioChem* **2020,** *21* (3), 315-319.
212. Falk, B. T.; Liang, Y.; McCoy, M. A., Diffusion profiling of therapeutic proteins by using solution NMR spectroscopy. *ChemBioChem* **2019,** *20* (7), 896-899.
213. Skidmore, K.; Hewitt, D.; Kao, Y. H., Quantitation and characterization of process impurities and extractables in protein-containing solutions using proton NMR as a general tool. *Biotechnology progress* **2012,** *28* (6), 1526-1533.
214. Magarian, N.; Lee, K.; Nagpal, K.; Skidmore, K.; Mahajan, E., Clearance of extractables and leachables from single-use technologies via ultrafiltration/diafiltration operations. *Biotechnology progress* **2016,** *32* (3), 718-724.
215. Martinez Morales, M.; Zalar, M.; Sonzini, S.; Golovanov, A. P.; van der Walle, C. F.; Derrick, J. P., Interaction of a macrocycle with an aggregation-prone region of a monoclonal antibody. *Molecular pharmaceutics* **2019**.
216. Torosantucci, R.; Furtmann, B.; Elshorst, B.; Pfeiffer-Marek, S.; Hartleb, T.; Andres, N.; Bussemer, T., Protein-excipient interactions evaluated via nuclear magnetic resonance studies in polysorbate-based multidose protein formulations: influence on antimicrobial efficacy and potential study approach. *Journal of pharmaceutical sciences* **2018,** *107* (10), 2531-2537.

217. Edwards, J. M.; Bramham, J. E.; Podmore, A.; Bishop, S. M.; van der Walle, C. F.; Golovanov, A. P., ^{19}F Dark-State Exchange Saturation Transfer NMR Reveals Reversible Formation of Protein-Specific Large Clusters in High-Concentration Protein Mixtures. *Analytical chemistry* **2019,** *91* (7), 4702-4708.
218. Edwards, J. M.; Derrick, J. P.; van der Walle, C. F.; Golovanov, A. P., ^{19}F NMR as a Tool for Monitoring Individual Differentially Labeled Proteins in Complex Mixtures. *Molecular pharmaceutics* **2018,** *15* (7), 2785-2796.
219. Edwards, J. M.; Harris, P.; Bukrinski, J. T.; Golovanov, A. P., Use of ^{19}F Differential Labelling for the Simultaneous Detection and Monitoring of Three Individual Proteins in a Serum Environment. *ChemPlusChem* **2019,** *84* (5), 443-446.
220. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K., Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **1997,** *94* (23), 12366-12371.
221. Zuiderweg, E. R., Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **2002,** *41* (1), 1-7.
222. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* **2013,** *73*, 1-16.
223. Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: a protein− protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* **2003,** *125* (7), 1731-1737.
224. Pellecchia, M.; Sem, D. S.; Wüthrich, K., NMR in drug discovery. *Nature Reviews Drug Discovery* **2002,** *1* (3), 211-219.
225. Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W., NMR-based screening of proteins containing 13C-labeled methyl groups. *Journal of the American Chemical Society* **2000,** *122* (33), 7898-7904.
226. Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W., The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *Journal of the American Chemical Society* **2000,** *122* (6), 1241-1242.
227. Bain, A. D., Chemical Exchange Effects in NMR. **1999**.
228. Becker, W.; Bhattiprolu, K. C.; Gubensäk, N.; Zangger, K., Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *ChemPhysChem* **2018,** *19* (8), 895-906.
229. Craik, D. J.; Wilce, J. A., Studies of protein-ligand interactions by NMR. *Protein NMR Techniques* **1997**, 195-232.
230. McAlister, M. S.; Mott, H. R.; van der Merwe, P. A.; Campbell, I. D.; Davis, S. J.; Driscoll, P. C., NMR analysis of interacting soluble forms of the cell-cell recognition molecules CD2 and CD48. *Biochemistry* **1996,** *35* (19), 5982-5991.
231. Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J., Two-dimensional NMR lineshape analysis. *Scientific reports* **2016,** *6*, 24826.
232. Waudby, C. A.; Ouvry, M.; Davis, B.; Christodoulou, J., Two-dimensional NMR lineshape analysis of single, multiple, zero and double quantum correlation experiments. *Journal of biomolecular NMR* **2020,** *74* (1), 95-109.
233. Kleckner, I. R.; Foster, M. P., An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2011,** *1814* (8), 942-968.
234. Mittermaier, A.; Kay, L. E., New tools provide new insights in NMR studies of protein dynamics. *Science* **2006,** *312* (5771), 224-228.
235. Kai, L.; Torchia, D.; Bax, A., Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphyloccal nuclase. *Biochemistry* **1989,** *28*, 8972-8979.
236. Englander, S. W., Protein folding intermediates and pathways studied by hydrogen exchange. *Annual review of biophysics and biomolecular structure* **2000,** *29* (1), 213-238.
237. Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W., Protein dynamics viewed by hydrogen exchange. *Protein Science* **2012,** *21* (7), 996-1005.
238. Eriksson, M.; Härd, T.; Nilsson, L., On the pH dependence of amide proton exchange rates in proteins. *Biophysical journal* **1995,** *69* (2), 329-339.
239. Cordier, F.; Grzesiek, S., Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *Journal of molecular biology* **2002,** *317* (5), 739-752.
240. Zhang, Y.-Z., Protein and peptide structure and interactions studied by hydrogen exchanger and NMR. **1995**.
241. Hwang, T.-L.; van Zijl, P. C.; Mori, S., Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *Journal of biomolecular NMR* **1998,** *11* (2), 221-226.

242. Meyer, B.; Peters, T., NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition* **2003,** *42* (8), 864-890.
243. Clore, G.; Gronenborn, A., Theory and applications of the transferred nuclear Overhauser effect to the study of the conformations of small ligands bound to proteins. *Journal of Magnetic Resonance (1969)* **1982,** *48* (3), 402-417.
244. Albrand, J.; Birdsall, B.; Feeney, J.; Roberts, G.; Burgen, A., The use of transferred nuclear Overhauser effects in the study of the conformations of small molecules bound to proteins. *International Journal of Biological Macromolecules* **1979,** *1* (1), 37-41.
245. Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, G. W.; Murcko, M. A.; Moore, J. M., The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chemistry & biology* **1999,** *6* (10), 755-769.
246. Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J., *Protein NMR spectroscopy: principles and practice*. Elsevier: 1995.
247. Meiboom, S.; Gill, D., Modified spin-echo method for measuring nuclear relaxation times. *Review of scientific instruments* **1958,** *29* (8), 688-691.
248. Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W., One-dimensional relaxation-and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *Journal of the American Chemical Society* **1997,** *119* (50), 12257-12261.
249. Maity, S.; Gundampati, R. K.; Suresh Kumar, T. K., NMR Methods to Characterize Protein-Ligand Interactions. *Natural Product Communications* **2019,** *14* (5), 1934578X19849296.
250. Hamel, D. J.; Dahlquist, F. W., The contact interface of a 120 kD CheA-CheW complex by methyl TROSY interaction spectroscopy. *Journal of the American Chemical Society* **2005,** *127* (27), 9676-9677.
251. Mayer, M.; Meyer, B., Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie International Edition* **1999,** *38* (12), 1784-1788.
252. Mayer, M.; Meyer, B., Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society* **2001,** *123* (25), 6108-6117.
253. Neffe, A. T.; Bilang, M.; Grüneberg, I.; Meyer, B., Rational optimization of the binding affinity of CD4 targeting peptidomimetics with potential anti HIV activity. *Journal of medicinal chemistry* **2007,** *50* (15), 3482-3488.
254. Pickhardt, M.; Larbig, G.; Khlistunova, I.; Coksezen, A.; Meyer, B.; Mandelkow, E.-M.; Schmidt, B.; Mandelkow, E., Phenylthiazolyl-hydrazide and its derivatives are potent inhibitors of τ aggregation and toxicity in vitro and in cells. *Biochemistry* **2007,** *46* (35), 10016-10023.
255. Szczepina, M. G.; Zheng, R. B.; Completo, G. C.; Lowary, T. L.; Pinto, B. M., STD-NMR studies suggest that two acceptor substrates for GlfT2, a bifunctional galactofuranosyltransferase required for the biosynthesis of Mycobacterium tuberculosis arabinogalactan, compete for the same binding site. *ChemBioChem* **2009,** *10* (12), 2052-2059.
256. Angulo, J.; Enríquez-Navas, P. M.; Nieto, P. M., Ligand–receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry–A European Journal* **2010,** *16* (26), 7803-7812.
257. Klein, J.; Meinecke, R.; Mayer, M.; Meyer, B., Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS STD NMR. *Journal of the American Chemical Society* **1999,** *121* (22), 5336-5337.
258. Meinecke, R.; Meyer, B., Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin αIIbβ3. *Journal of medicinal chemistry* **2001,** *44* (19), 3059-3065.
259. Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J., Saturation-transfer difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding. *Journal of chemical Education* **2011,** *88* (7), 990-994.
260. Skinner, A. L.; Laurence, J. S., High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. *Journal of pharmaceutical sciences* **2008,** *97* (11), 4670-4695.
261. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B., WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *Journal of biomolecular NMR* **2001,** *21* (4), 349-359.
262. Dalvit, C.; Pevarello, P.; Tatò, M.; Veronesi, M.; Vulpetti, A.; Sundström, M., Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *Journal of biomolecular NMR* **2000,** *18* (1), 65-68.
263. Lagassé, H. 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*.

264. Hansel, T. T.; Kropshofer, H.; Singer, T.; Mitchell, J. A.; George, A. J., The safety and side effects of monoclonal antibodies. *Nature reviews Drug discovery* **2010,** *9* (4), 325.
265. Hudson, P. J.; Souriau, C., Engineered antibodies. *Nature medicine* **2003,** *9* (1), 129.
266. Reverdatto, S.; Burz, D. S.; Shekhtman, A., Peptide aptamers: development and applications. *Current topics in medicinal chemistry* **2015,** *15* (12), 1082.
267. Mascini, M.; Palchetti, I.; Tombelli, S., Nucleic acid and peptide aptamers: fundamentals and bioanalytical aspects. *Angewandte Chemie International Edition* **2012,** *51* (6), 1316-1332.
268. Skrlec, K.; Strukelj, B.; Berlec, A., Non-immunoglobulin scaffolds: a focus on their targets. *Trends in Biotechnology* **2015,** *33* (7), 408-418.
269. Vazquez-Lombardi, R.; Phan, T. G.; Zimmermann, C.; Lowe, D.; Jermutus, L.; Christ, D., Challenges and opportunities for non-antibody scaffold drugs. *Drug Discovery Today* **2015,** *20* (10), 1271-1283.
270. Nord, K.; Gunneriusson, E.; Ringdahl, J.; Ståhl, S.; Uhlén, M.; Nygren, P.-Å., Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nature biotechnology* **1997,** *15* (8), 772.
271. Getmanova, E. V.; Chen, Y.; Bloom, L.; Gokemeijer, J.; Shamah, S.; Warikoo, V.; Wang, J.; Ling, V.; Sun, L., Antagonists to human and mouse vascular endothelial growth factor receptor 2 generated by directed protein evolution in vitro. *Chemistry & biology* **2006,** *13* (5), 549-556.
272. Zahnd, C.; Wyler, E.; Schwenk, J. M.; Steiner, D.; Lawrence, M. C.; McKern, N. M.; Pecorari, F.; Ward, C. W.; Joos, T. O.; Plückthun, A., A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *Journal of molecular biology* **2007,** *369* (4), 1015-1028.
273. Martin, H. L.; Bedford, R.; Heseltine, S. J.; Tang, A. A.; Haza, K. Z.; Rao, A.; McPherson, M. J.; Tomlinson, D. C., Non-immunoglobulin scaffold proteins: Precision tools for studying protein-protein interactions in cancer. *New biotechnology* **2018**.
274. Retout, M.; Valkenier, H.; Triffaux, E.; Doneux, T.; Bartik, K.; Bruylants, G., Rapid and Selective Detection of Proteins by Dual Trapping Using Gold Nanoparticles Functionalized with Peptide Aptamers. *Acs Sensors* **2016,** *1* (7), 929-933.
275. Theurillat, J.-P.; Dreier, B.; Nagy-Davidescu, G.; Seifert, B.; Behnke, S.; Zürrer-Härdi, U.; Ingold, F.; Plückthun, A.; Moch, H., Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer. *Modern Pathology* **2010,** *23* (9), 1289.
276. Skerra, A., Anticalins as alternative binding proteins for therapeutic use. *Current opinion in molecular therapeutics* **2007,** *9* (4), 336-344.
277. Friedman, M.; Orlova, A.; Johansson, E.; Eriksson, T. L.; Hoiden-Guthenberg, I.; Tolmachev, V.; Nilsson, F. Y.; Stahl, S., Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. *J Mol Biol* **2008,** *376* (5), 1388-402.
278. Zahnd, C.; Kawe, M.; Stumpp, M. T.; de Pasquale, C.; Tamaskovic, R.; Nagy-Davidescu, G.; Dreier, B.; Schibli, R.; Binz, H. K.; Waibel, R., Efficient tumor targeting with high-affinity designed ankyrin repeat proteins: effects of affinity and molecular size. *Cancer research* **2010,** *70* (4), 1595-1605.
279. Münch, R. C.; Mühlebach, M. D.; Schaser, T.; Kneissl, S.; Jost, C.; Plückthun, A.; Cichutek, K.; Buchholz, C. J., DARPins: an efficient targeting domain for lentiviral vectors. *Molecular Therapy* **2011,** *19* (4), 686-693.
280. Hughes, D. J.; Tiede, C.; Penswick, N.; Ah-San Tang, A.; Trinh, C. H.; Mandal, U.; Zajac, K. Z.; Gaule, T.; Howell, G.; Edwards, T. A., Generation of specific inhibitors of SUMO-1–and SUMO-2/3–mediated protein-protein interactions using Affimer (Adhiron) technology. *Sci. Signal.* **2017,** *10* (505), eaaj2005.
281. Stadler, L.; Tomlinson, D.; Lee, T.; Knowles, M.; Ferrigno, P. K., The use of a neutral peptide aptamer scaffold to anchor BH3 peptides constitutes a viable approach to studying their function. *Cell death & disease* **2014,** *5* (1), e1037.
282. Jost, C.; Plückthun, A., Engineered proteins with desired specificity: DARPins, other alternative scaffolds and bispecific IgGs. *Current opinion in structural biology* **2014,** *27*, 102-112.
283. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **2007,** *372* (3), 774-797.
284. Zheng, Y.-J.; Ornstein, R. L., What happens to salt-bridges in nonaqueous environments: insights from quantum mechanics calculations. *Journal of the American Chemical Society* **1996,** *118* (45), 11237-11243.
285. Staniforth, R. A.; Giannini, S.; Higgins, L. D.; Conroy, M. J.; Hounslow, A. M.; Jerala, R.; Craven, C. J.; Waltho, J. P., Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J* **2001,** *20* (17), 4774-81.
286. Jerala, R.; Zerovnik, E., Accessing the global minimum conformation of stefin A dimer by annealing under partially denaturing conditions. *J Mol Biol* **1999,** *291* (5), 1079-89.

287. Jenko Kokalj, S.; Guncar, G.; Stern, I.; Morgan, G.; Rabzelj, S.; Kenig, M.; Staniforth, R. A.; Waltho, J. P.; Zerovnik, E.; Turk, D., Essential role of proline isomerization in stefin B tetramer formation. *J Mol Biol* **2007,** *366* (5), 1569-79.
288. Roy, A.; Kucukural, A.; Zhang, Y., I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **2010,** *5* (4), 725.
289. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y., The I-TASSER Suite: protein structure and function prediction. *Nature methods* **2015,** *12* (1), 7.
290. Zhang, Y., I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **2008,** *9* (1), 40.
291. Karaca, E.; Melquiond, A. S.; de Vries, S. J.; Kastritis, P. L.; Bonvin, A. M., Building macromolecular assemblies by information-driven docking introducing the haddock multibody docking server. *Molecular & Cellular Proteomics* **2010,** *9* (8), 1784-1794.
292. Van Zundert, G.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P.; Karaca, E.; Melquiond, A.; van Dijk, M.; De Vries, S.; Bonvin, A., The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology* **2016,** *428* (4), 720-725.
293. Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T., AMBER 2016. *University of California, San Francisco* **2016,** *1* (2), 3.
294. Case, D.; Betz, R.; Cerutti, D.; Cheatham, T.; Darden III, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N., AMBER 2016 (University of California, 2016). *Google Scholar*.
295. Hom, G. K.; Lassila, J. K.; Thomas, L. M.; Mayo, S. L., Dioxane contributes to the altered conformation and oligomerization state of a designed engrailed homeodomain variant. *Protein science* **2005,** *14* (4), 1115-1119.
296. Sanders, A.; Craven, C. J.; Higgins, L. D.; Giannini, S.; Conroy, M. J.; Hounslow, A. M.; Waltho, J. P.; Staniforth, R. A., Cystatin forms a tetramer through structural rearrangement of domain-swapped dimers prior to amyloidogenesis. *Journal of molecular biology* **2004,** *336* (1), 165-178.
297. Jurczak, P.; Groves, P.; Szymanska, A.; Rodziewicz-Motowidlo, S., Human cystatin C monomer, dimer, oligomer, and amyloid structures are related to health and disease. *FEBS letters* **2016,** *590* (23), 4192-4201.
298. Jenko, S.; Škarabot, M.; Kenig, M.; Gunčar, G.; Muševič, I.; Turk, D.; Žerovnik, E., Different propensity to form amyloid fibrils by two homologous proteins—human stefins A and B: searching for an explanation. *Proteins: Structure, Function, and Bioinformatics* **2004,** *55* (2), 417-425.
299. Ekiel, I.; Abrahamson, M., Folding-related dimerization of human cystatin C. *Journal of Biological Chemistry* **1996,** *271* (3), 1314-1321.
300. Perlenfein, T. J.; Mehlhoff, J. D.; Murphy, R. M., Insights into the mechanism of cystatin C oligomer and amyloid formation and its interaction with β-amyloid. *Journal of Biological Chemistry* **2017,** *292* (27), 11485-11498.
301. Rousseau, F.; Schymkowitz, J.; Wilkinson, H.; Itzhaki, L., Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proceedings of the National Academy of Sciences* **2001,** *98* (10), 5596-5601.
302. Taler-Verčič, A.; Hasanbašić, S.; Berbić, S.; Stoka, V.; Turk, D.; Žerovnik, E., Proline residues as switches in conformational changes leading to amyloid fibril formation. *International journal of molecular sciences* **2017,** *18* (3), 549.
303. Abrahamson, M.; Grubb, A., Increased body temperature accelerates aggregation of the Leu-68--> Gln mutant cystatin C, the amyloid-forming protein in hereditary cystatin C amyloid angiopathy. *Proceedings of the National Academy of Sciences* **1994,** *91* (4), 1416-1420.
304. Byeon, I.-J. L.; Louis, J. M.; Gronenborn, A. M., A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *Journal of molecular biology* **2003,** *333* (1), 141-152.
305. Louis, J. M.; Byeon, I.-J. L.; Baxa, U.; Gronenborn, A. M., The GB1 amyloid fibril: Recruitment of the peripheral β-strands of the domain swapped dimer into the polymeric interface. *Journal of molecular biology* **2005,** *348* (3), 687-698.
306. Malevanets, A.; Sirota, F. L.; Wodak, S. J., Mechanism and energy landscape of domain swapping in the B1 domain of protein G. *Journal of molecular biology* **2008,** *382* (1), 223-235.
307. Tan, Y. S.; Lane, D. P.; Verma, C. S., Stapled peptide design: principles and roles of computation. *Drug Discovery Today* **2016,** *21* (10), 1642-1653.
308. Fairlie, D. P.; Dantas de Araujo, A., Stapling peptides using cysteine crosslinking. *Peptide Science* **2016,** *106* (6), 843-852.
309. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **2010,** *43* (1), 186-190.
310. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (4), 658-674.

311. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (4), 486-501.
312. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography* **2012,** *68* (4), 352-367.
313. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (1), 12-21.
314. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015,** *31* (8), 1325-7.
315. Hyberts, S. G.; Robson, S. A.; Wagner, G., Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *Journal of biomolecular NMR* **2013,** *55* (2), 167-178.
316. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015,** *11* (8), 3696-3713.
317. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic acids research* **2005,** *33* (suppl_2), W368-W371.
318. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research* **2012,** *40* (W1), W537-W541.
319. Price, D. J.; Brooks III, C. L., A modified TIP3P water potential for simulation with Ewald summation. *The Journal of chemical physics* **2004,** *121* (20), 10096-10103.
320. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79* (2), 926-935.
321. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993,** *98* (12), 10089-10092.
322. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995,** *103* (19), 8577-8593.
323. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.
324. Roe, D. R.; Cheatham III, T. E., PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **2013,** *9* (7), 3084-3095.
325. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research* **2000,** *33* (12), 889-897.
326. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *The Journal of Physical Chemistry B* **2000,** *104* (15), 3712-3720.
327. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004,** *55* (2), 383-394.
328. Gohlke, H.; Kiel, C.; Case, D. A., Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of molecular biology* **2003,** *330* (4), 891-913.
329. Fraczkiewicz, R.; Braun, W., Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of computational chemistry* **1998,** *19* (3), 319-333.
330. Skerra, A., Alternative non-antibody scaffolds for molecular recognition. *Curr Opin Biotechnol* **2007,** *18* (4), 295-304.
331. Zalar, M.; Indrakumar, S.; Levy, C. W.; Tunnicliffe, R. B.; Peters, G. H. J.; Golovanov, A. P., Studies of the oligomerisation mechanism of a cystatin-based engineered protein scaffold. *Sci Rep* **2019,** *9* (1), 9067.
332. Turk, V.; Stoka, V.; Turk, D., Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci* **2008,** *13* (5), 5406-20.
333. Garai, K.; Frieden, C., The association-dissociation behavior of the ApoE proteins: kinetic and equilibrium studies. *Biochemistry* **2010,** *49* (44), 9533-41.
334. Kuzmic, P., Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Analytical biochemistry* **1996,** *237* (2), 260-73.

335. Schreiber, G.; Haran, G.; Zhou, H. X., Fundamental aspects of protein-protein association kinetics. *Chem Rev* **2009,** *109* (3), 839-60.
336. Škerget, K.; Vilfan, A.; Pompe-Novak, M.; Turk, V.; Waltho, J. P.; Turk, D.; Žerovnik, E., The mechanism of amyloid-fibril formation by stefin B: Temperature and protein concentration dependence of the rates. *Proteins* **2009,** *74* (2), 425-436.
337. Geering, K.; Jaunin, P.; Jaisser, F.; Merillat, A. M.; Horisberger, J. D.; Mathews, P. M.; Lemas, V.; Fambrough, D. M.; Rossier, B. C., Mutation of a conserved proline residue in the beta-subunit ectodomain prevents Na^+-K^+-ATPase oligomerization. *The American journal of physiology* **1993,** *265* (4 Pt 1), C1169-74.
338. Nilsson, M.; Wang, X.; Rodziewicz-Motowidlo, S.; Janowski, R.; Lindstrom, V.; Onnerfjord, P.; Westermark, G.; Grzonka, Z.; Jaskolski, M.; Grubb, A., Prevention of domain swapping inhibits dimerization and amyloid fibril formation of cystatin C - Use of engineered disulfide bridges, antibodies, and carboxymethylpapain to stabilize the monomeric form of cystatin C. *J Biol Chem* **2004,** *279* (23), 24236-24245.
339. Zerovnik, E.; Virden, R.; Jerala, R.; Turk, V.; Waltho, J. P., On the mechanism of human stefin B folding: I. Comparison to homologous stefin A. Influence of pH and trifluoroethanol on the fast and slow folding phases. *Proteins* **1998,** *32* (3), 296-303.
340. Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G., Frustration in biomolecules. *Quarterly reviews of biophysics* **2014,** *47* (4), 285-363.
341. Mascarenhas, N. M.; Gosavi, S., Protein Domain-Swapping Can Be a Consequence of Functional Residues. *The journal of physical chemistry. B* **2016,** *120* (28), 6929-38.
342. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2014,** *31* (8), 1325-1327.
343. Carpenter, J. F.; Randolph, T. W.; Jiskoot, W.; Crommelin, D. J.; Middaugh, C. R.; Winter, G.; Fan, Y. X.; Kirshner, S.; Verthelyi, D.; Kozlowski, S., Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. *Journal of pharmaceutical sciences* **2009,** *98* (4), 1201-1205.
344. Moussa, E. M.; Panchal, J. P.; Moorthy, B. S.; Blum, J. S.; Joubert, M. K.; Narhi, L. O.; Topp, E. M., Immunogenicity of therapeutic protein aggregates. *Journal of pharmaceutical sciences* **2016,** *105* (2), 417-430.
345. Chatani, E.; Imamura, H.; Yamamoto, N.; Kato, M., Stepwise organization of the β-structure identifies key regions essential for the propagation and cytotoxicity of insulin amyloid fibrils. *Journal of Biological Chemistry* **2014,** *289* (15), 10399-10410.
346. Chatani, E.; Inoue, R.; Imamura, H.; Sugiyama, M.; Kato, M.; Yamamoto, M.; Nishida, K.; Kanaya, T., Early aggregation preceding the nucleation of insulin amyloid fibrils as monitored by small angle X-ray scattering. *Scientific reports* **2015,** *5*, 15485.
347. Ziaunys, M.; Sneideris, T.; Smirnovas, V., Self-inhibition of insulin amyloid-like aggregation. *Physical Chemistry Chemical Physics* **2018,** *20* (43), 27638-27645.
348. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B., Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proceedings of the National Academy of Sciences* **2003,** *100* (1), 330-335.
349. Finder, V. H.; Glockshuber, R., Amyloid-β aggregation. *Neurodegenerative Diseases* **2007,** *4* (1), 13-27.
350. Ferrone, F. A.; Rotter, M. A., Crowding and the polymerization of sickle hemoglobin. *Journal of Molecular Recognition* **2004,** *17* (5), 497-504.
351. Frare, E.; Mossuto, M. F.; de Laureto, P. P.; Tolin, S.; Menzer, L.; Dumoulin, M.; Dobson, C. M.; Fontana, A., Characterization of oligomeric species on the aggregation pathway of human lysozyme. *Journal of molecular biology* **2009,** *387* (1), 17-27.
352. Muschol, M.; Hill, S. E.; Mulaj, M., Multiple Pathways of Lysozyme Aggregation. In *Bio-nanoimaging*, Elsevier: 2014; pp 389-396.
353. Vetri, V.; Librizzi, F.; Leone, M.; Militello, V., Thermal aggregation of bovine serum albumin at different pH: comparison with human serum albumin. *European Biophysics Journal* **2007,** *36* (7), 717-725.
354. Juárez, J.; Taboada, P.; Mosquera, V., Existence of different structural intermediates on the fibrillation pathway of human serum albumin. *Biophysical journal* **2009,** *96* (6), 2353-2370.
355. Stirpe, A.; Pantusa, M.; Rizzuti, B.; Sportelli, L.; Bartucci, R.; Guzzi, R., Early stage aggregation of human serum albumin in the presence of metal ions. *International journal of biological macromolecules* **2011,** *49* (3), 337-342.
356. Tosstorff, A.; Svilenov, H.; Peters, G. H.; Harris, P.; Winter, G., Structure-based discovery of a new protein-aggregation breaking excipient. *European Journal of Pharmaceutics and Biopharmaceutics* **2019,** *144*, 207-216.

357. Tosstorff, A.; Menzen, T.; Winter, G., Exploring Chemical Space for new Substances to stabilize a therapeutic Monoclonal Antibody. *Journal of pharmaceutical sciences* **2019**.
358. Nicoud, L.; Cohrs, N.; Arosio, P.; Norrant, E.; Morbidelli, M., Effect of polyol sugars on the stabilization of monoclonal antibodies. *Biophysical chemistry* **2015,** *197*, 40-46.
359. Barata, T. S.; Zhang, C.; Dalby, P. A.; Brocchini, S.; Zloh, M., Identification of protein–excipient interaction hotspots using computational approaches. *International journal of molecular sciences* **2016,** *17* (6), 853.
360. Arakawa, T.; Kita, Y.; Ejima, D.; Tsumoto, K.; Fukada, H., Aggregation suppression of proteins by arginine during thermal unfolding. *Protein and peptide letters* **2006,** *13* (9), 921-927.
361. Arakawa, T.; Tsumoto, K., The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochemical and biophysical research communications* **2003,** *304* (1), 148-152.
362. Chang, B. S.; Kendrick, B. S.; Carpenter, J. F., Surface-induced denaturation of proteins during freezing and its inhibition by surfactants. *Journal of pharmaceutical sciences* **1996,** *85* (12), 1325-1330.
363. Bam, N. B.; Cleland, J. L.; Yang, J.; Manning, M. C.; Carpenter, J. F.; Kelley, R. F.; Randolph, T. W., Tween protects recombinant human growth hormone against agitation-induced damage via hydrophobic interactions. *Journal of pharmaceutical sciences* **1998,** *87* (12), 1554-1559.
364. Timasheff, S. N., Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proceedings of the National Academy of Sciences* **2002,** *99* (15), 9721-9726.
365. Chi, E. Y.; Krishnan, S.; Kendrick, B. S.; Chang, B. S.; Carpenter, J. F.; Randolph, T. W., Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Science* **2003,** *12* (5), 903-913.
366. Patching, S. G., Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2014,** *1838* (1), 43-55.
367. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q., Insights into protein–ligand interactions: mechanisms, models, and methods. *International journal of molecular sciences* **2016,** *17* (2), 144.
368. Svilenov, H.; Winter, G., The ReFOLD assay for protein formulation studies and prediction of protein aggregation during long-term storage. *European Journal of Pharmaceutics and Biopharmaceutics* **2019,** *137*, 131-139.
369. King, A. C.; Woods, M.; Liu, W.; Lu, Z. J.; Gill, D.; Krebs, M. R. H., High-throughput measurement, correlation analysis, and machine-learning predictions for pH and thermal stabilities of Pfizer-generated antibodies. *Protein Science* **2011,** *20* (9), 1546-1557.
370. Svilenov, H.; Winter, G., Rapid sample-saving biophysical characterisation and long-term storage stability of liquid interferon alpha2a formulations: Is there a correlation? *International Journal of Pharmaceutics* **2019,** *562*, 42-50.
371. Maddux, N. R.; Iyer, V.; Cheng, W.; Youssef, A. M.; Joshi, S. B.; Volkin, D. B.; Ralston, J. P.; Winter, G.; Middaugh, C. R., High throughput prediction of the long-term stability of pharmaceutical macromolecules from short-term multi-instrument spectroscopic data. *Journal of pharmaceutical sciences* **2014,** *103* (3), 828-839.
372. Chakroun, N.; Hilton, D.; Ahmad, S. S.; Platt, G. W.; Dalby, P. A., Mapping the aggregation kinetics of a therapeutic antibody fragment. *Molecular pharmaceutics* **2016,** *13* (2), 307-319.
373. Robinson, M. J.; Matejtschuk, P.; Bristow, A. F.; Dalby, P. A., T m-values and unfolded fraction can predict aggregation rates for granulocyte colony stimulating factor variant formulations but not under predominantly native conditions. *Molecular pharmaceutics* **2017,** *15* (1), 256-267.
374. Zhang, J.; Frey, V.; Corcoran, M.; Zhang-van Enk, J.; Subramony, J. A., Influence of arginine salts on the thermal stability and aggregation kinetics of monoclonal antibody: Dominant role of anions. *Molecular pharmaceutics* **2016,** *13* (10), 3362-3369.
375. Gentiluomo, L.; Svilenov, H. L.; Augustijn, D.; El Bialy, I.; Greco, M. L.; Kulakova, A.; Indrakumar, S.; Mahapatra, S.; Morales, M. M.; Pohl, C., Advancing therapeutic protein discovery and development through comprehensive computational and biophysical characterization. *Molecular Pharmaceutics* **2019**.
376. Singh, S. M.; Bandi, S.; Jones, D. N.; Mallela, K. M., Effect of Polysorbate 20 and Polysorbate 80 on the higher-order structure of a monoclonal antibody and its fab and fc fragments probed using 2D nuclear magnetic resonance spectroscopy. *Journal of pharmaceutical sciences* **2017,** *106* (12), 3486-3498.
377. Matsuo, H.; Walters, K. J.; Teruya, K.; Tanaka, T.; Gassner, G. T.; Lippard, S. J.; Kyogoku, Y.; Wagner, G., Identification by NMR spectroscopy of residues at contact surfaces in large, slowly exchanging macromolecular complexes. *Journal of the American Chemical Society* **1999,** *121* (42), 9903-9904.

378. Hammes, G. G.; Tallman, D. E., A nuclear magnetic resonance study of the interaction of L-epinephrine with phospholipid vesicles. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1971,** *233* (1), 17-25.
379. Maurer, T., NMR studies of protein-ligand interactions. In *Protein-Ligand Interactions*, Springer: 2005; pp 197-213.
380. Jahnke, W.; Floersheim, P.; Ostermeier, C.; Zhang, X.; Hemmig, R.; Hurth, K.; Uzunov, D. P., NMR reporter screening for the detection of high-affinity ligands. *Angewandte Chemie International Edition* **2002,** *41* (18), 3420-3423.
381. Meiby, E.; Simmonite, H.; le Strat, L.; Davis, B.; Matassova, N.; Moore, J. D.; Mrosek, M.; Murray, J.; Hubbard, R. E.; Ohlson, S., Fragment screening by weak affinity chromatography: comparison with established techniques for screening against HSP90. *Analytical chemistry* **2013,** *85* (14), 6756-6766.
382. Urick, A. K.; Calle, L. P.; Espinosa, J. F.; Hu, H.; Pomerantz, W. C., Protein-observed fluorine NMR is a complementary ligand discovery method to 1H CPMG ligand-observed NMR. *ACS chemical biology* **2016,** *11* (11), 3154-3164.
383. Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C., Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno [2, 3-d] pyrimidine inhibitors of the Hsp90 molecular chaperone. *Journal of medicinal chemistry* **2009,** *52* (15), 4794-4809.
384. Winter, A.; Higueruelo, A. P.; Marsh, M.; Sigurdardottir, A.; Pitt, W. R.; Blundell, T. L., Biophysical and computational fragment-based approaches to targeting protein–protein interactions: applications in structure-guided drug discovery. *Quarterly reviews of biophysics* **2012,** *45* (4), 383-426.
385. Cimmperman, P.; Baranauskienė, L.; Jachimovičiūtė, S.; Jachno, J.; Torresan, J.; Michailovienė, V.; Matulienė, J.; Sereikaitė, J.; Bumelis, V.; Matulis, D., A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophysical journal* **2008,** *95* (7), 3222-3231.
386. Zhuravleva, A.; Korzhnev, D. M., Protein folding by NMR. *Progress in nuclear magnetic resonance spectroscopy* **2017,** *100*, 52-77.
387. Horváth, G.; Biczók, L.; Majer, Z.; Kovács, M.; Micsonai, A.; Kardos, J.; Toke, O., Structural insight into a partially unfolded state preceding aggregation in an intracellular lipid-binding protein. *The FEBS journal* **2017,** *284* (21), 3637-3661.
388. Wang, W.; Ohtake, S., Science and art of protein formulation development. *International journal of pharmaceutics* **2019**, 118505.
389. Manning, M. C.; Liu, J.; Li, T.; Holcomb, R. E., Rational design of liquid formulations of proteins. In *Advances in protein chemistry and structural biology*, Elsevier: 2018; Vol. 112, pp 1-59.
390. Barnett, G. V.; Razinkov, V. I.; Kerwin, B. A.; Blake, S.; Qi, W.; Curtis, R. A.; Roberts, C. J., Osmolyte effects on monoclonal antibody stability and concentration-dependent protein interactions with water and common osmolytes. *The Journal of Physical Chemistry B* **2016,** *120* (13), 3318-3330.
391. Courtenay, E.; Capp, M.; Anderson, C.; Record, M., Vapor pressure osmometry studies of osmolyte− protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro. *Biochemistry* **2000,** *39* (15), 4455-4471.
392. Matulis, D.; Kranz, J. K.; Salemme, F. R.; Todd, M. J., Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **2005,** *44* (13), 5258-5266.
393. Holdgate, G. A.; Anderson, M.; Edfeldt, F.; Geschwindner, S., Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: matching high information content with high throughput. *Journal of structural biology* **2010,** *172* (1), 142-157.
394. Dalvit, C., NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug discovery today* **2009,** *14* (21-22), 1051-1057.
395. Svilenov, H. L.; Kulakova, A.; Zalar, M.; Golovanov, A. P.; Harris, P.; Winter, G., Orthogonal Techniques to Study the Effect of pH, Sucrose, and Arginine Salts on Monoclonal Antibody Physical Stability and Aggregation During Long-Term Storage. *Journal of pharmaceutical sciences* **2020,** *109* (1), 584-594.
396. Lubach, J. W.; Padden, B. E.; Winslow, S. L.; Salsbury, J. S.; Masters, D. B.; Topp, E. M.; Munson, E. J., Solid-state NMR studies of pharmaceutical solids in polymer matrices. *Analytical and bioanalytical chemistry* **2004,** *378* (6), 1504-1510.
397. Zalar, M.; Svilenov, H. L.; Golovanov, A. P., Binding of excipients is a poor predictor for aggregation kinetics of biopharmaceutical proteins. *European Journal of Pharmaceutics and Biopharmaceutics* **2020**.
398. Panjwani, N.; Hodgson, D. J.; Sauvé, S.; Aubin, Y., Assessment of the effects of pH, formulation and deformulation on the conformation of interferon alpha-2 by NMR. *Journal of pharmaceutical sciences* **2010,** *99* (8), 3334-3342.

399. Rembert, K. B.; Okur, H. I.; Hilty, C.; Cremer, P. S., An NH moiety is not required for anion binding to amides in aqueous solution. *Langmuir* **2015,** *31* (11), 3459-3464.
400. Tadeo, X.; Pons, M.; Millet, O., Influence of the Hofmeister anions on protein stability as studied by thermal denaturation and chemical shift perturbation. *Biochemistry* **2007,** *46* (3), 917-923.
401. Rembert, K. B.; Paterová, J.; Heyda, J.; Hilty, C.; Jungwirth, P.; Cremer, P. S., Molecular mechanisms of ion-specific effects on proteins. *Journal of the American Chemical Society* **2012,** *134* (24), 10039-10046.
402. Zalar, M.; Golovanov, A. P., New Disulphide Bond in Cystatin-Based Protein Scaffold Prevents Domain-Swap-Mediated Oligomerization and Stabilizes the Functionally Active Form. *ACS Omega* **2019,** *4* (19), 18248-18256.
403. Venkitakrishnan, R. P.; Benard, O.; Max, M.; Markley, J. L.; Assadi-Porter, F. M., Use of NMR saturation transfer difference spectroscopy to study ligand binding to membrane proteins. In *Membrane Protein Structure and Dynamics*, Springer: 2012; pp 47-63.
404. Fielding, L.; Rutherford, S.; Fletcher, D., Determination of protein–ligand binding affinity by NMR: observations from serum albumin model systems. *Magnetic resonance in Chemistry* **2005,** *43* (6), 463-470.
405. Tosstorff, A.; Menzen, T.; Winter, G., Exploring Chemical Space for new Substances to stabilize a therapeutic Monoclonal Antibody. *Journal of pharmaceutical sciences* **2020,** *109* (1), 301-307.
406. Kalayan, J.; Henchman, R. H.; Warwicker, J., A Model for Counterion Binding and Charge Reversal on Protein Surfaces. *Molecular Pharmaceutics* **2019**.
407. Nau, W. M.; Assaf, K.; Das, D., Applications of Cucurbiturils in Medicinal Chemistry and Chemical Biology. *Frontiers in chemistry* **2019,** *7*, 619.
408. Lau, Y. H.; De Andrade, P.; Wu, Y.; Spring, D. R., Peptide stapling techniques based on different macrocyclisation chemistries. *Chemical Society Reviews* **2015,** *44* (1), 91-102.
409. Klein, M., Stabilized helical peptides: overview of the technologies and its impact on drug discovery. *Expert opinion on drug discovery* **2017,** *12* (11), 1117-1125.

BLANK PAGE