



Characterization and solution structure of multi-domain proteins and protein complexes

Vitaliyivna Kulakova, Alina

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Vitaliyivna Kulakova, A. (2019). *Characterization and solution structure of multi-domain proteins and protein complexes*. Technical University of Denmark.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Characterization and solution structure of multi-domain proteins and protein complexes

Ph.D. Thesis

Alina Kulakova

August 2019

Department of Chemistry

Technical University of Denmark

Preface

This dissertation has been submitted to the Department of Chemistry, Technical University of Denmark, in partial fulfillment of the requirement for the Ph.D. degree. The presented work was carried out in the Department of Chemistry, Technical University of Denmark, under the supervision of Pernille Harris and Günther H.J. Peters. Nano differential scanning fluorimetry and dynamic light scattering experiments were carried out in Novozymes under the supervision of Werner W. Streicher. Static and dynamic light scattering were performed in University of Manchester under the supervision of Robin Curtis and computational work under the supervision of Jim Warwicker (not included in this thesis).

Small-angle X-ray scattering experiments were performed at the EMBL P12 BioSAXS beamline at DESY (Hamburg, Germany) and at the EMBL BM29 BioSAXS beamline at ESRF (Grenoble, France). The Ph.D. project is a part of Protein-excipient Interactions and Protein-Protein Interactions in formulation consortium funded by European Union's Horizon 2020 research and innovation program (grant agreement nr 675074).

Current work resulted in the following manuscripts that are included as chapters in this thesis:

Kulakova, A., Indrakumar, S., Sønnderby, P., Gentiluomo, L., Streicher W., Roessner, D., Frieß, W., Peters G.H.J., Harris, P. Structural-Stability Studies on Recombinant Human Transferrin. **2019** (Manuscript – submitted).

Kulakova, A., Indrakumar, S., Sønnderby, P., Mahapatra, S., Streicher W., Peters G.H.J., Harris, P. Albumin-Nepriylisin Fusion Protein: Understanding Stability using Small-Angle X-ray Scattering and Molecular Dynamics Simulations. **2019.** (Manuscript – submitted).

Contribution to other publications and manuscripts

Gentiluomo, L., Roessner, D., Augustijn, D., Svilenov, H., **Kulakova, A.**, Mahapatra, S., Winter, G., Streicher, W., Rinnan, Å., Peters, G.H.J., Harris, P., Frieß, W. Application of Interpretable Artificial Neural Networks to Early Monoclonal Antibodies Development. *Eur J Pharm Biopharm.* **2019**, 141, 81-89.

Augustijn, D., Mahapatra, S., Streicher, W., Svilenov, H., **Kulakova, A.**, Pohl, C., Rinnan, Å., Novel Non-linear Curve Fitting to Resolve Protein Unfolding Transitions in Intrinsic Fluorescence Differential Scanning Fluorimetry. *Eur J Pharm Biopharm.* **2019**, 142, 506-517.

Gentiluomo, L., Svilenov, H., Augustijn, D., El Bialy, I., Greco, M.L., **Kulakova, A.**, Indrakumar, S., Mahapatra, S., Morales, M., Pohl, C., Roche, A., Tosstorff, A., Curtis, R., Nørgaard, A., Khan, T.A., Peters, G.H.J., Rinnan, Å., Streicher, W., van der Walle, C., Uddin, S., Winter, G., Roessner, D., Harris, P., Frieß, W. Boosting therapeutic protein development by publicly available datasets including comprehensive computational and biophysical characterization. **2019**. (Manuscript - submitted).

Svilenov, H., **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. **2019**. (Manuscript - submitted).

Indrakumar, S., **Kulakova, A.**, Harris, P., Peters, G.H.J. Dynamics of Human Serum Transferrin in Varying Physicochemical Conditions Explored Using Molecular Dynamics Simulations. **2019**. (Manuscript - submitted).

Augustijn, D., **Kulakova, A.**, Mahapatra, S., **Pohl, C.**, Harris, P., Rinnan, Å., Isothermal Chemical Denaturation: Data Analysis, Error Detection and Correction by PARAFAC2. **2019**. (Manuscript in preparation).

Kongens Lyngby, August 2019

Alina Kulakova

Acknowledgments

I would like to thank my supervisors, Pernille Harris and Günther Peters, for giving me opportunity to be part of this great project, and for all the help, feedback, and guidance during this journey. A special thanks to Pernille, for all the support, and for being very understanding and caring. Thanks to Günther for all constructive discussions, and for all the help and great company in the synchrotron trips. I would also like to acknowledge the European Union's Horizon 2020 research and innovation program for funding.

I am very thankful to Pernille Sønderby for all the help, advices, and suggestions with regards to interpretation of SAXS data. A special thanks to Jens Bukrinski for his help with protein classification and all the input on my work during the conferences. Thanks to Werner Streicher and Allan Nørgaard from Novozymes for being so open and helpful. A special thanks to Werner and Gitte J. Funch for helping me with nanoDSF and MST measurements and data interpretation. Thanks to Gitte for being so patient, for her advice, and all the help in the lab. I would also like to thank Robin Curtis and Jim Warwicker for my secondment in the University of Manchester. Thanks to Luke Holloway and Nicole Sibanda for helping me with SLS and DSL measurements. Many thanks to my colleagues from the PIPPI project, Christin Pohl, Sowmya Indrakumar, Sujata Mahapatra, and Dillen Augustijn, and the whole PIPPI consortium for all the scientific input and discussions. Thanks to Maria Blanner Bang for helping with practical matters in the lab and for being so supportive and understanding.

I would like to thank the beamline scientists from P12 BioSAXS in DESY and B29 BioSAXS in ESRF for all the help and assistance during our measurements and DANSCATT for funding our SAXS trips.

I would like to thank my colleagues from DTU, Tine M. Frederiksen, Line A. Ryberg, Ulf Molich, Kasper D. Tidemand, Sindrila Bannik, Natalia Skawinska, Alexandre Voute, Suk Kyu Ko, and Ida M. Vedel for being such great company. Thanks to Tine for introducing me to Matlab, and to Ulf and Alex for helping me every time when I was struggling with it. Thanks to Line and Pernille S. for helping me make the SAXS programs work. Many thanks to Line, Christin, Sujata, Natalia, Pernille S., Pernille H., and Günther for all the synchrotron trips, for delicious food, nice company, and moral support during long working nights. Thanks to Fabrizio Bottaro and Vanessa Baj, my Italian friends, for nice dinners and good times. Thank you all for all the events that we

had together: Christmas lunch, Friday breakfasts, coffee breaks with cakes and ice cream, dinners out, and many others. Many thanks to Erling Halfdan Stenby and all people from Department of Chemistry for making it such a great place to work. All of it made me feel very welcome.

I would like to express my gratitude to my parents, Leonid and Lyudmyla, for everything they have done for me. For being brave and, despite all the difficulties, leaving their home country to give me opportunity to have a better life. Thanks to my dad for translating all the textbooks for me after long working days, so I could understand and be ready for my exams. Thanks to my mom for never giving up, for being so strong and supportive, for being so optimistic and positive in difficult situations. Thanks to my little sister, Esmeralda, for being so kind and sweet. I would like to thank my grandparents for taking care of me back in Ukraine. Thanks to my grandma, for investing her time and energy in my education.

I would like to thank my boyfriend, Mads Almind, for being so supportive, for encouraging me and for fighting against my self-criticism. Thank you for being there always when I needed it, for being so understanding and caring. For cooking for me when I was busy with writing and after exhausting synchrotron trips. For making me stop working when I became too tired and annoyed. For helping me correcting my thesis and for putting all the missing articles. For helping me with Matlab and fixing my Mendeley. Thank you for always being there for me, not only when everything was going great, but also when I felt like I was failing.

I am very lucky to have all of you in my life. All of you made me person who I am now and I would not have gotten this far without you.

Alina Kulakova

Abstract

Since the introduction of recombinant human insulin in 1982, protein-based therapeutics have become an essential part of medical treatment. Every year many protein-drugs are introduced into the market and numerous are under clinical trials. Despite all the advantages of protein-based pharmaceuticals, stabilization is very challenging. Under inappropriate conditions, they might lose their activity and induce severe adverse effects. Stabilization (formulation) of protein-drugs is one of the most critical, time-consuming and expensive steps in drug development. Many promising protein-drugs have failed clinical trials due to their stability challenges. Unfortunately, no general rules for the formulation process have been reported and it is not yet possible to predict the behavior of different proteins under different conditions.

Protein-excipient Interactions and Protein-Protein Interactions in formulation (PIPPI) is a consortium (EU grant nr 675074), which goal is to improve the molecular understanding behind protein stability and thereby of the formulation process. The strategy is to combine systematic investigations of the physicochemical behavior of different types of proteins with molecular interactions. The final goal is to combine all raw and analyzed data into a database that will become publicly available for the scientific community.

This Ph.D. thesis is a part of the PIPPI project, which focus is to investigate stability of multidomain proteins using small angle X-ray scattering. In this study, three different types of multidomain proteins were investigated: transferrin, albumin-neprilysin fusion protein, and monoclonal antibodies. All of them were systematically studied under different physicochemical conditions using high-throughput techniques. These studies were used to choose conditions for small angle X-ray scattering. The resulting data were combined with other measurements, such as MD simulation to provide a better understanding of stability on the molecular level. Despite the common trends in stability, all studied multidomain proteins show different behavior and molecular interactions. Additionally, this work shows that combination of multiple methods leads to a better understanding of protein stability.

Resumé

Siden rekombinant human insulin i 1982 blev introduceret som lægemiddel er protein-baseret medicin blevet en vigtig del af medicinske behandlinger. Hvert år bliver mange protein-baserede lægemidler introduceret på markedet, og adskillige er i kliniske forsøg.

Der er mange fordele ved protein-baseret medicin, men stabilisering af proteinerne, så de kan holde sig i minimum to år, er en udfordring. Under forkerte forhold mister proteinerne og dermed medicinen sin aktivitet og kan enten blive virkningsløst eller forårsage alvorlige bivirkninger. Stabilisering (formulering) er et af de mest kritiske, tidskrævende, og dyreste trin i udviklingen af protein-baseret medicin. Mange lovende protein-baserede lægemidler fejler i kliniske test på grund af udfordringer med stabiliteten. Det er endnu ikke muligt at forudse, hvordan proteiner opfører sig under forskellige forhold, og derfor er formuleringsprocessen generisk.

Protein-excipient Interactions and Protein-Protein Interactions in formulation (PIPPI) er et EU konsortium (EU grant nr. 675074), som har det mål at forbedre forståelsen af protein-stabilitet på molekylært niveau, og derved også forbedre effektivisere formulerings-processen. Strategien i PIPPI-projektet er at kombinere systematiske undersøgelser af den fysisk-kemiske opførsel af forskellige typer proteiner med de molekylære vekselvirkninger, der forårsager deres mikroskopiske opførsel. Det endelige mål er at lægge både rå og behandlede data i en database, der er til rådighed for videre videnskabelige undersøgelser.

Denne ph.d.-afhandling er en del af PIPPI-projektet, og fokuserer på at undersøge stabiliteten af multidomæne-proteiner ved hjælp af små-vinkel røntgenspredning. Tre typer multidomæne-proteiner blev undersøgt: transferrin, albumin-nepriylsin fusions-protein, og monoklonale antistoffer. Først er de blevet undersøgt systematisk ved brug af *high throughput* metoder som termisk denaturering og isothermisk kemisk udfoldning. Resultaterne herfra blev brugt til at udvælge under hvilke betingelser der skulle laves småvinkel røntgenspredning. Resultaterne er blevet kombineret med andre undersøgelser, som for eksempel molekyle-dynamik simuleringer, for at forbedre forståelsen af stabiliteten på et molekylært niveau. På trods af de generelle tendenser i stabilitet, opfører alle de undersøgte multidomæne-proteiner sig forskelligt, men ved at kombinere de forskellige metoder er der opnået en bedre forståelse af proteinernes opførsel.

Nomenclature

$A(q)$	Amplitude of X-ray or light
apo-TrF	iron-free transferrin
BBB	blood brain barrier
C	antibody constant region
$c_{1/2}$	concentration of denaturant required to unfold 50% of the protein
C_H	antibody heavy constant region
C_L	antibody light constant region
c_x	concentration of x
DLS	Dynamic Light Scattering
D_{max}	maximum dimension
DMT1	Divalent Metal Transporter 1
<i>EOM</i>	Ensemble Optimization Method
F	fluorescence
Fab	fragment antigen-binding
Fc	fragment crystallizable region
FDA	Food and Drug Administration
Fe _C -TrF	partially open transferrin with iron bound to C-lobe
Fe _N -TrF	partially open transferrin with iron bound to N-lobe
FT	Fourier Transform
GA	Genetic Algorithm
GuHCl	guanidine hydrochloride
<i>hGH</i>	<i>human</i> growth hormone
His	histidine
HSA	<i>human</i> serum albumin
$I(q)$	intensity of X-ray at q
ICD	isothermal chemical denaturation
IgG	immunoglobulin G
K_d	dissociation constant
k_D	interaction diffusion coefficient
M	molar (mol/L)
mAb	monoclonal antibody

MALS	Multi-Angle Light Scattering
MD	Molecular Dynamics
MST	MicroScale Thermophoresis
<i>MW</i>	molecular weight
NA	Avogadro number
nanoDSF	nano differential scanning fluorimetry
NEP	neprilysin
NMR	nuclear magnetic resonance
$P(q)$	form factor
PD	polydispersity
PDB	Protein Data Bank
$p(r)$	pair-distance distance distribution
PIC	preferential interaction coefficient
q	scattering vector (momentum transfer)
R_g	radius of gyration
R_h	hydrodynamic radius
rTrF	recombinant human transferrin
$S(q)$	structure factor
SAS	small angle scattering
SAXS	small angle X-ray scattering
SEC	size exclusion chromatography
SLS	static light scattering
$T_{1/2}$	temperature of unfolding
T_{agg}	onset temperature of aggregation
T_m	melting temperature
TrF	transferrin
TrFR	transferrin receptor
V	antibody variable region
V_H	antibody heavy variable region
V_L	antibody light variable region
V_p	Porod volume
ΔG	variation of Gibbs free energy
$\Delta\rho$	electron density contrast
v	partial specific volume
λ	wavelength

Contents

Preface.....	i
Acknowledgments.....	iii
Abstract.....	v
Resumé.....	vi
Nomenclature.....	vii
Contents.....	ix
1. Introduction.....	1
2. Scientific background.....	3
Biopharmaceuticals and formulation.....	3
PIPPI project.....	7
3. Methods.....	11
Differential Scanning Fluorimetry.....	11
Isothermal Chemical Denaturation.....	12
SAXS.....	13
MicroScale Thermophoresis.....	20
4. Structural classification of protein therapeutics.....	21
Introduction.....	21
Results and discussion.....	24
Conclusions.....	27
Materials and methods.....	28
5. Transferrin.....	31

Introduction.....	31
Manuscript: Structural-Stability Studies on Recombinant Human Transferrin.....	33
Abstract.....	33
Introduction.....	34
Results.....	36
Discussion.....	43
Conclusion.....	45
Methods.....	46
Acknowledgements.....	49
6. Albumin-nepriylsin fusion protein.....	51
Introduction.....	51
Manuscript: Albumin-Nepriylsin Fusion Protein: Understanding Stability using Small-Angle X-ray Scattering and Molecular Dynamics Simulations.....	53
Abstract.....	53
Introduction.....	54
Results.....	55
Discussion.....	69
Conclusion.....	71
Methods.....	72
Acknowledgements.....	75
7. Monoclonal antibodies.....	77
Introduction.....	77
Results.....	80
Discussion.....	89
Conclusion.....	93
Methods.....	94
8. Overall discussion.....	97
9. Conclusion.....	101
References.....	103
Appendix A.....	117
Appendix B.....	125

Appendix C	137
Appendix D	145

Chapter 1

Introduction

Protein-based therapeutics have become a significant part of current medical treatment. Since the approval of human insulin, recombinant DNA technology has led to a significant increase in the number of approved biopharmaceuticals¹. In 2018, Food and Drug Administration (FDA) approved in total 59 drugs, from which 15 are protein-based therapeutics².

Protein-based therapeutics have a significant structural, functional, and pharmacological diversity. Molecular weight of the protein-drugs can range from 6 kDa (insulin) to 150 kDa (antibodies and botulinum toxin). Structurally, protein therapeutics can be single- or multidomain, mutated or linked together through recombinant DNA technology, resulting in fusion proteins. Due to their complexity, protein therapeutics are highly specific and have low potential to interfere with biological processes³, leading to the low number of side effects. However, production and manufacturing of the protein-drugs are complex and challenging.

Protein's structural complexity leads to challenges in stability: changes in secondary, tertiary, and quaternary structure can result in unfolding and/or aggregation. Those processes can lead not only to a loss of function, but also to undesirable adverse effects, *e.g.* immunogenicity⁴. One of the strategies for the protein stabilization is optimization of the protein solution (formulation). Various external factors, *e.g.* buffer, pH, and additives can influence protein stability, and therefore are tested during the formulation process⁴. No general rules for the formulation have been reported, as prediction of protein behavior under different conditions is not possible yet. Therefore, it remains one of the most expensive and time consuming steps in the protein-drug development.

PIPI (Protein-excipient Interactions and Protein-Protein Interactions in formulation)⁵ consortium is focused on understanding the protein stability on a molecular level. The overall project consists of studying structurally different proteins, which are representative of different classes of biopharmaceuticals. Initially, extensive stability studies were performed to mimic the formulation process in industry. These were later complemented with structural studies, which provide understanding of conformational changes of each protein under different conditions. Combination of different methods provides better understanding of structural-stability changes in different proteins. Understanding of rationale behind the protein stability will improve formulation

process, making it faster and cheaper. Detailed description of PIPPI project is described in chapter 2.

The current work is focused on structural-stability understanding of multidomain proteins and is divided in eight chapters. Chapter 2 covers theoretical background of the formulation process and more detailed description of the PIPPI project. Chapter 3 describes theory behind the used methods. Chapter 4 includes structural classification of biopharmaceuticals according to the SCOP database. Chapter 5, 6, and 7 cover structural-stability studies on transferrin, albumin-neprilysin fusion protein. Chapter 8 and 9 include overall discussion and conclusion.

Chapter 2

Scientific background

Biopharmaceuticals and formulation

Biopharmaceuticals are pharmaceuticals that are biological in nature and manufactured using biotechnology⁶. Those include protein-based therapeutics, gene therapy, and cell-based therapy products^{7,8}. In total, 17 gene/cell therapy products were approved by the Food and Drug Administration (FDA)⁹, whereas for protein and peptide drugs this number is around 180¹⁰, which makes them a very attractive topic for research.

Protein therapeutics are structurally complex molecules, produced by recombinant DNA technology. In their native state, protein-drugs are highly specific, which generally results in low number of side effects. However, just like all the proteins, protein-drugs do not have a static structure. Due to their flexibility, proteins exist in equilibrium between multiple conformations¹¹. Changes of environment around the protein can lead to less favorable conformational changes, which results in unfolding and/or aggregation. This can cause loss of function and immunogenicity^{1,12}. Therefore, it is important to find right conditions (formulation) in which the native protein conformation is preserved. This is a quite challenging process, because there are many different factors that can have an impact, *e.g.* pH, buffer species, and additives (excipients), such as sugars, salts, amino acids, and denaturants⁴.

Buffer

Both pH and buffer species have an effect on protein stability. Usually, proteins are stable in narrow pH range, and hence accurate control of the pH during the formulation is essential for protein stability. pH changes modulate net charge property of the protein and also modify the degree of surface hydration: depending of pI of the protein, it has different charges at different pH^{4,13}.

Several studies show that both buffer species and concentration affect protein stability. Acetate, citrate, histidine, phosphate, tris, and glycine are commonly used buffering agents, which cover pH from 3 to 10^{4,14}.

Sugars and polyols

Sugar and polyols are un-specific protein stabilizers, stabilizing protein by preferential exclusion (accumulation of water molecules around proteins)¹⁵⁻¹⁷ (see Figure 2.1). Sucrose and trehalose are the most often used as protein stabilizers. They may have different effect on different proteins. Additionally, stabilizing effect of sugars depends on its concentration⁴.

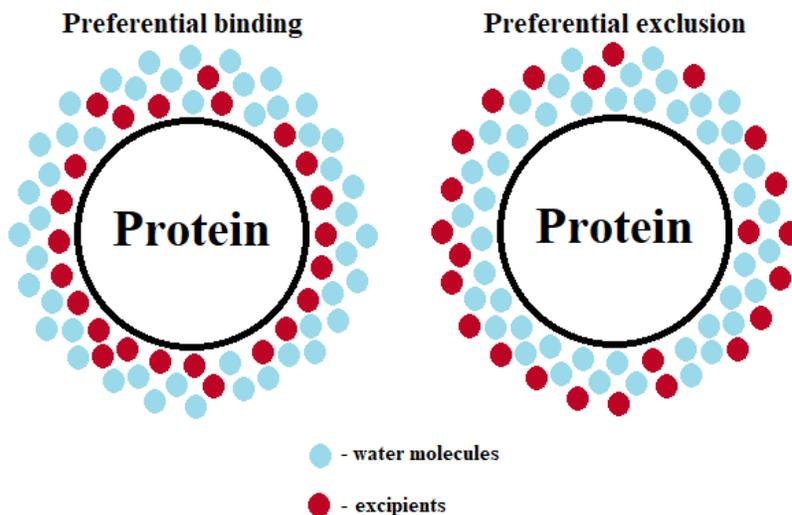


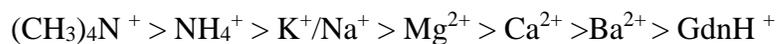
Figure 2.1: Protein stabilization by preferential binding and preferential exclusion (inspired from *Ohtake et.al*¹⁸).

Surfactants

Surfactants reduce the surface tension of proteins in solution and decrease protein adsorption/aggregation at hydrophobic surfaces. In protein formulation non-ionic surfactants are preferred over ionic surfactants, because ionic surfactants can bind to both polar and nonpolar groups in the protein and cause denaturation¹⁹. Tween 20, 40, and 80, and trixton X-100 are examples of surfactants that are used in protein formulation⁴.

Salts

Salts can stabilize, destabilize, or have no effect on proteins, depending on type and concentration of both salt and protein^{4,20}. The effect of salt at high concentration correlates with the Hoffmeister lyotropic series:



Ions listed to the right of Cl^- are chaotropes (salting-in), while salts to its left are kosmotropes (salting-out)^{21,22}. Chaotropes are known as “water structure breakers” that destabilize proteins by binding to charged groups and peptide bonds. This leads to decrease in protein surface tension, and increase in solubility and denaturation. Kosmotropes are known as “water structure makers”

that enhance hydrophobic interactions, which reduces solubility of hydrophobic groups and leads to a preferential exclusion (see Figure 2.1). This results in increasing surface tension and enhanced protein stability^{13,16}. Therefore, in the protein formulation kosmotropes salts are most commonly used.

Salt effect not only depends of its nature, but also on the solution pH, which determines charged state of ionizable groups⁴. Additionally, positive or negative effect on the protein stability depends of pI of the protein. When pH is close to pI, proteins are neutral and have tendency to aggregate. In this case, both anions and cations interact with the protein surface. When pH is below pI, protein is positively charged, and therefore anions are more prone to interact with the protein surface. Contrary, when pH is above pI, proteins are negatively charged and cations have more tendency to interact with a protein surface¹³.

Polymers

Presence of polymers have shown positive and negative effects on protein stability. Stabilization of polymers is usually due to surface activity, preferential exclusion, steric hindrance of protein-protein interactions, and increasing viscosity that limits protein movement⁴. Some hydrophilic polymers (*e.g.* dextran, hydroxyethyl starch, PEG4000, and gelatin) enhance protein thermal stability and suppress aggregation¹⁸. Charged polymers can stabilize proteins by electrostatic interactions⁴ however this effect is most likely protein specific¹⁸. For example, acidic fibroblast growth factor was stabilized by variety of sulfated and phosphorylated anionic polymers (*e.g.* heparin, dextran sulfate, pentosane sulfate, and others), while storage stability of lactate dehydrogenase was improved by polyethyleneimine, which is cationic polymer¹⁸.

Poly ethylene glycols (PEGs)

Due to a small non-polar moiety, PEG is different from the other polymers¹⁸. PEGs are hydrophobic in nature and interact with hydrophobic side chains, which can promote unfolding of the protein (especially at higher temperatures). However, multiple studies show PEGs stabilizing effect on certain proteins. Mechanism of action is not known yet, but it seems to be related to the protein type and PEG size⁴.

Metal ions

Multiple studies show that some metal ions stabilize proteins by binding and making them more rigid²³. Metal ion are also used to enhance activity of certain proteins⁴.

Amino acids

Amino acids are widely used in protein formulation. Certain amino acids, *e.g.* glycine, alanine, and proline are classified as compatible solutes, due to their compatibility with protein structure and function in the cells. These do not interfere with enzyme activity or protein structure, and therefore stabilize proteins without affecting their function. As well as other solutes, amino acids can be classified as chaotropes and kosmotropes (described above in salts section)²⁴. Some of

them, alone or in combination, can stabilize proteins mainly by preferential exclusion²⁵ (see Figure 2.1).

In formulation, amino acids are mainly used to increase solubility and reduce aggregation. Arginine and lysine were shown to be effective against heat stress, while methionine and histidine are often used as anti-oxidants. Stabilization of the amino acids and other additives is manifested in the increasing temperature of unfolding. Unlike other amino acids, arginine usually does not have a significant effect on temperature of unfolding and is not considered a protein stabilizer²⁴, but it is highly effective in suppressing protein aggregation^{4,18}. The most common amino acids that are used in formulation are histidine, proline, arginine, and glycine¹⁴.

Some compounds can stabilize one protein, but destabilize another. In addition, changes in concentration and combination of different factors alter protein stability. Therefore, the lack in understanding protein stability on the molecular level makes the protein formulation process long, expensive and challenging.

PIPPI project

The aim of the PIPPI project (mentioned in chapter 1) is to improve understanding of protein stability on the molecular level. In order to do so, structurally different proteins were selected to cover all existing classes of biopharmaceuticals (see Figure 2.2 and chapter 4). Stability of those proteins was studied under different conditions, varying pH, buffer species, NaCl, and in the presence of different excipients. With a number of different techniques, conformational and colloidal stability of the proteins were investigated. It was decided that chemical stability and container issues not were considered within the project.

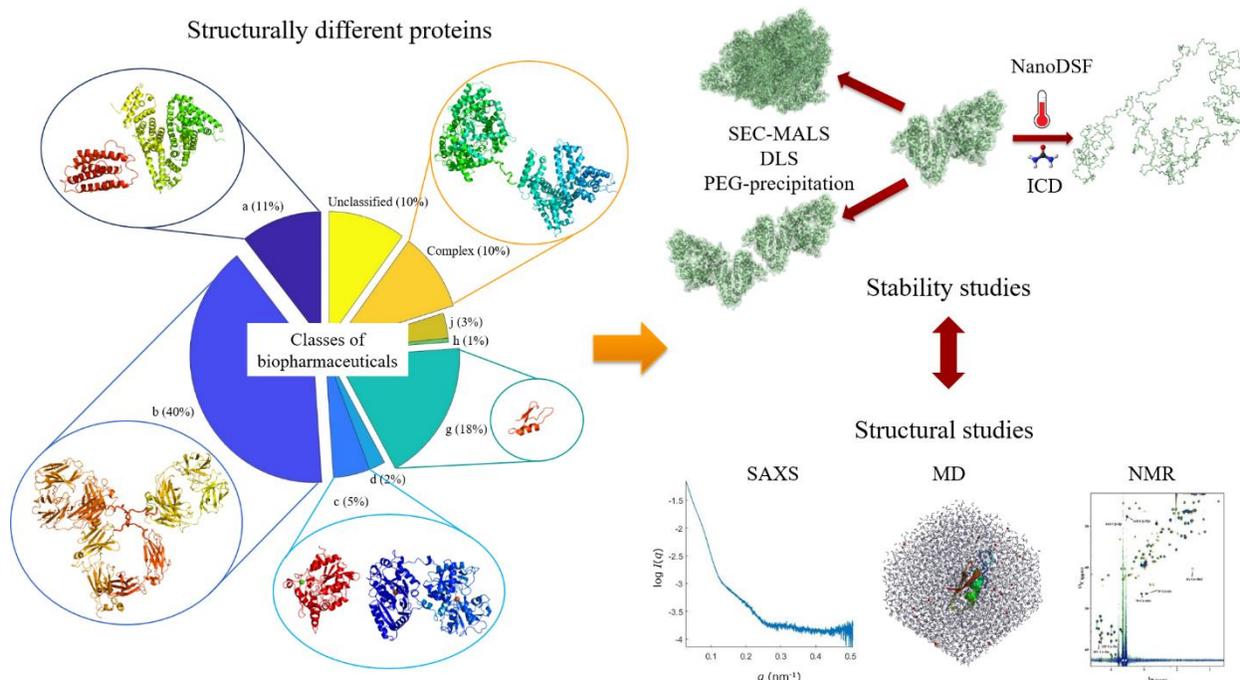


Figure 2.2: PIPPI project overview. Structurally different proteins selected to cover classes of biopharmaceuticals: albumin (in light green), interferon (in dark red), fusion protein (in cyan), antibody (in yellow/orange), lipase (in red), transferrin (in dark blue), and plectasin (in orange). Methods used for stability studies: dynamic light scattering (DLS), nanoscale differential scanning fluorimetry (nanoDSF), isothermal chemical denaturation (ICD), size exclusion chromatography with multi-angle light scattering (SEC-MALS), and PEG-precipitation. Methods used for structural studies: small angle X-ray scattering (SAXS), molecular dynamics simulations (MD), and NMR (nuclear magnetic resonance).

Stability studies were divided in two parts: screen I and screen II (see Figure 2.3). Screen I was focused on pH and salt effect, where histidine and tris buffers were selected to cover pH range from 5.0 to 7.5, and 8.0 to 9.0. NaCl was chosen to study salt effect, as it is the most often-used salt in protein formulation⁴. From the screen I, pH 5.0 and 6.5 were selected for the screen II, where effect of different buffer species and excipients was studied. For the buffer effect, acetate (pH 5.0), phosphate (pH 6.5), and histidine (pH 5.0 and 6.5) were chosen. To study effect of excipients, sucrose, arginine, and proline were selected to cover two major groups of excipients: sugars and amino acids. Additionally, chosen excipients are the most commonly used in protein formulation.

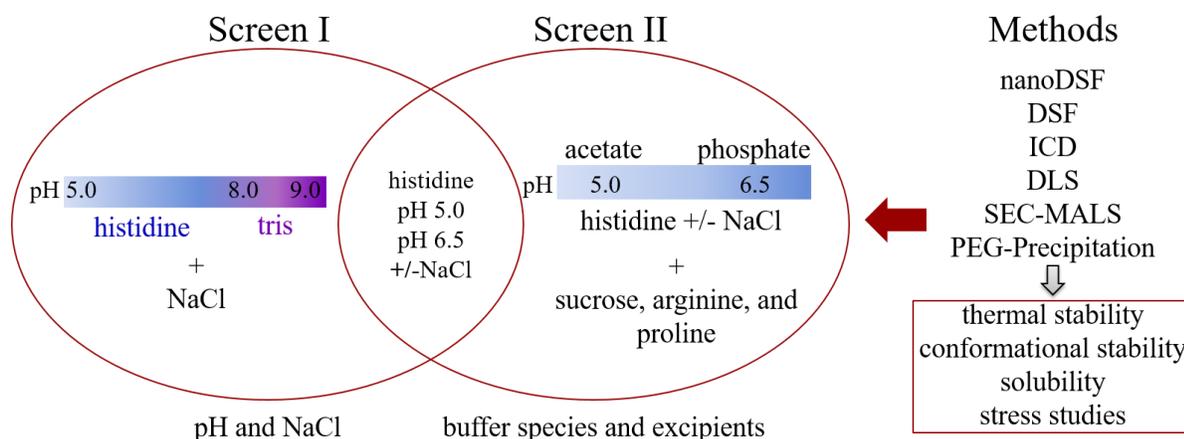


Figure 2.3: Schematic representation of stability studies.

Both screen I and screen II include a variety of methods (see Figure 2.3), which can provide complementary information and lead to a more complete picture of protein stability (see Table 2.1). All parameters that are relevant for this thesis are explained in the Chapter 3.

Table 2.1: Methods used for Screen I and Screen II.

	Method	Measured parameters*
(Nano)DSF	(Nanoscale) Differential Scanning Fluorimetry	$T_{1/2}$
DLS	Dynamic Light Scattering	R_h , PD, T_{agg} , k_D
SEC-MALS	Size Exclusion Chromatography with Multi-Angle Light Scattering	Retention time, monomer to aggregate/fragment ratio, Molecular weight (MW), PD
ICD	Isothermal Chemical Denaturation	ΔG_{unfold} , $c_{1/2}$, m -value
PEG-precipitation		Turbidity midpoint

*Definition: unfolding temperature ($T_{1/2}$), radius of hydration (R_h), polydispersity (PD), onset temperature of aggregation (T_{agg}), interaction diffusion parameter (k_D), variation of Gibbs free energy of unfolding (ΔG_{unfold}), concentration of denaturant required to unfold 50% of the protein ($c_{1/2}$), and slope of the unfolding curve (m -value).

Structural studies were performed on selected conditions from screen I and screen II using Small Angle X-ray Scattering, Molecular Dynamics, and Nuclear Magnetic Resonance. The main goal is to connect all obtained results and find structure-stability relation that will lead to a better understating of protein stability.

All obtained data, raw and analyzed, will be uploaded into a database (PIPPI-data), which will become publicly available.

My Ph.D. project is focused on structural stability studies of multi domain proteins, namely monoclonal antibodies, fusion protein, and human transferrin (see Table 2.2). It includes ICD and nanoDSF (as a part of screen I and II) and structural studies performed by SAXS.

Table 2.2: Overview of studied proteins: abbreviation and PPI number.

Protein	Abbreviation	PPI number
Human Serum Transferrin	rTrF	PPI44
Albumin-Nepilysin fusion protein	HSA-NEP	PPI18
Monoclonal Antibodies	mAb ₁ , mAb ₂ , and mAb ₃	PPI04, PPI13, PPI17

Chapter 3

Methods

Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) is a thermal denaturation-based method, which is used to analyze protein thermostability and ligand interactions^{26–28}. In DSF, the protein is heated at a constant heating rate from 20 to 95°C in the presence of environmentally sensitive fluorescent dyes²⁷. The stability of most of the proteins decreases with temperature²⁷, leading to a loss of quaternary structure and unfolding. In the folded state, hydrophobic regions are buried in the core of the protein, while in the unfolded state these are exposed to the solvent. The dyes are highly fluorescent in non-polar environments, and therefore an increase in exposed hydrophobic residues leads to a rise in fluorescence²⁸.

It is also possible to perform label-free DSF by measuring changes in intrinsic fluorescence of aromatic residues (tryptophan and/or tyrosine)²⁹. NanoDSF is a modified method that employs intrinsic fluorescence of the protein to follow the unfolding. Tryptophan and tyrosine, just as the other hydrophobic residues, are mainly located in the core of the molecule. When exposed to the solvent, emission maximum shifts towards higher wavelengths, from around 330 to 350 nm. Therefore, increase in F350/F330 is an indication of protein unfolding³⁰. NanoDSF unfolding curve, represented as F350/F330 vs. temperature usually results in a sigmoidal curve and the point of inflection corresponds to the unfolding temperature ($T_{1/2}$), where 50% of the protein is unfolded²⁷ (see Figure 3.1). Due to irreversibility of thermal denaturation, the point of inflection is defined as $T_{1/2}$ instead of melting temperature (T_m). Often, multidomain proteins show multiple $T_{1/2}$, due to distinct stability of individual domains³⁰.

Additionally, nanoDSF instruments (*e.g.* Prometheus NT.48, NanoTemper Technologies) require very low amount of protein (< 10 µg) and low sample volume (~10 µl)²⁹, which makes nanoDSF very useful for high throughput screening.

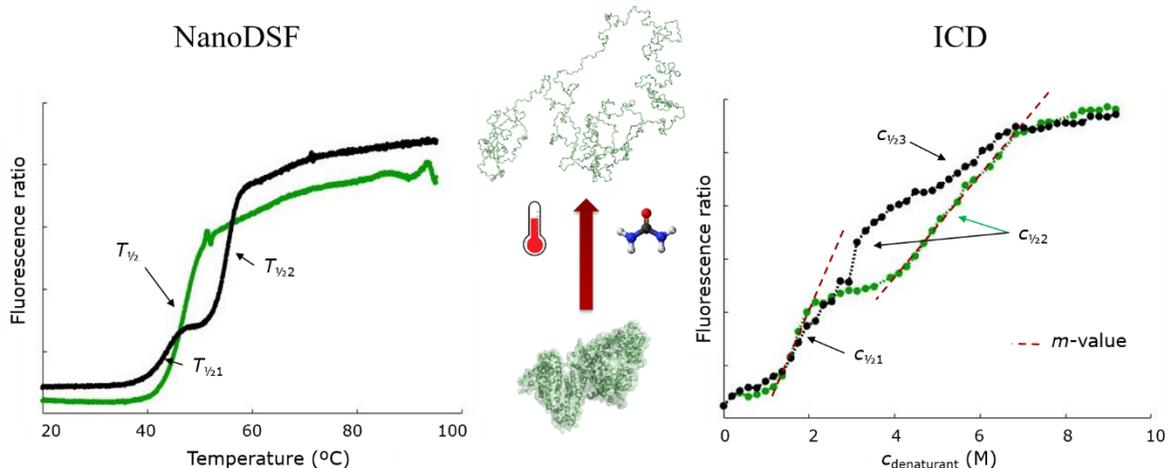


Figure 3.1: Unfolding curves from nanoDSF (on the left) and ICD (on the right).

Isothermal Chemical Denaturation

Isothermal chemical denaturation (ICD) is used to study physical stability of the proteins by inducing denaturation with a chemical agent at constant temperature³¹. From all denaturants, urea and guanidine are the most commonly used in ICD experiments^{31,32}. As for as nanoDSF, ICD can be performed in the presence of the dye and can also employ protein intrinsic fluorescence to follow the unfolding. The point of inflection in ICD curve corresponds to the $c_{1/2}$ - $c_{denaturant}$ necessary to unfold 50% of the protein (see Figure 3.1).

Contrary to nanoDSF, chemical denaturation is a reversible process, which allows to calculate the Gibbs free energy. Variation in Gibbs energy (ΔG) is universally used to determine the direction and the equilibrium of chemical reactions and physical transformations³³. ΔG follows a simple linear dependence with denaturant concentration:

$$\Delta G = \Delta G^0 - mc_{denaturant} \quad (3.1)$$

where ΔG^0 is Gibbs free energy in the absence of denaturant and m is the m -value. The m -value corresponds to the slope of the denaturation curve (see Figure 3.1) and is correlated with the change in solvent accessible area of the protein during unfolding³⁴. The relation between ΔG^0 , m -value, and $c_{1/2}$ is shown in eq. 3.2.

$$c_{1/2} = \Delta G^0 / m \quad (3.2)$$

During chemical denaturation, multidomain proteins often show multi-step unfolding process, resulting in multiple ΔG , $c_{1/2}$, and m -values³⁵, giving insight to stability of individual domains.

Despite providing valuable thermodynamic information, ICD was not used for industrial purposes, due to the lack of the automated instrumentation. Nowadays, there are several automated ICD instruments available, which perform automatic preparation, incubation, and fluorescence measurements (*e.g.* AVIA ICD 2304, Unchained Labs).

In nanoDSF and ICD experiments, proteins undergo different unfolding processes (thermal and chemical unfolding), which might result in distinct denaturation profiles. Therefore, combination of these methods can provide complementary information and better understanding of protein conformational stability.

SAXS

Small Angle X-ray Scattering (SAXS) is a powerful method that is used to study structure and structural changes of biomolecules in solution^{36,37}. The basic principles of the SAXS experiment are shown in the Figure 3.2. The experiment consists of irradiation of the protein sample with monochromatic X-rays. The incident beam is scattered by the sample components and recorded at the detector as intensity as a function of momentum transfer ($I(q)$). $I(q)$ and q are parameters that define reciprocal space, which result from Fourier transform (FT). $I(q)$ results from FT applied to spatial distribution of scattering centers in the sample and q results from FT applied to real space coordinates³⁶. The momentum transfer is defined as

$$q = \frac{4\pi \sin \theta}{\lambda} \quad (3.3)$$

where 2θ is the scattering angle and λ the wavelength of the radiation. The total scattering intensity ($I(q)$) is a product of the amplitude ($A(q)$) and its complex conjugate. The total scattered amplitude depends on the scattering lengths of individual scatters (b_i) and on the phase factors ($e^{iq \cdot r_i}$), defined by

$$A(q) = \sum_{i=1}^N b_i e^{iq \cdot r_i} \quad (3.4)$$

where N is a number of scatters that can take any orientation³⁷. Biomolecules can also be represented as continuous distribution of scattering length density $\rho(r)$, which is defined as the total scattering length of the atoms per unit volume. SAXS experiments on macromolecules in solution require separate measurement for sample and identical buffer^{36,37}. The scattering pattern of the solvent is subtracted from the scattering pattern of the sample, resulting in the net scattering from the particle ($\Delta\rho(r) = \rho(r) - \rho_s$) and can be related to the amplitude by eq. 3.5³⁷.

$$A(q) = \mathfrak{F}[\rho(r)] = \int_V \Delta\rho(r) e^{iq \cdot r} dr \quad (3.5)$$

Considering an ensemble of particles, the total scattering will depend on their distribution in the sample, therefore the total scattering density of the object is convolution of particle density distribution ($\Delta\rho(r)$) and $d(r)$, which describes position and orientation of the particles within the volume ($\Delta\rho(r)_{\text{total}} = \Delta\rho(r) * d(r)$). As a result, the amplitude is a product of $\mathfrak{F}[\Delta\rho(r)]$ and $\mathfrak{F}[d(r)]$.

The measured intensity is also a product $I(q) = P(q) * S(q)$. $P(q)$ is a form-factor that depends on the particle structure, while $S(q)$ is a structure factor that depends on the particle distribution³⁷. $P(q)$ contains information on size, shape, and structure of the molecules and molecular assemblies³⁶, while $S(q)$ describes interactions between macromolecules in solution³⁷.

SAXS data analysis requires several assumptions of ideality. As already mentioned, intensity is a product of $P(q)$ and $S(q)$ and in order to obtain structural information, it is important to measure dilute sample, where $S(q)$ can be approximated to 1 ($I(q) = P(q)$), thus neglecting intermolecular interactions. Additionally, all parameters are calculated assuming that the sample is monodisperse.

SAXS data can be analyzed by multiple analyses, namely Guinier, pair-distance distribution ($p(r)$) function, Porod, and Kratky analyses (see Figure 3.2).

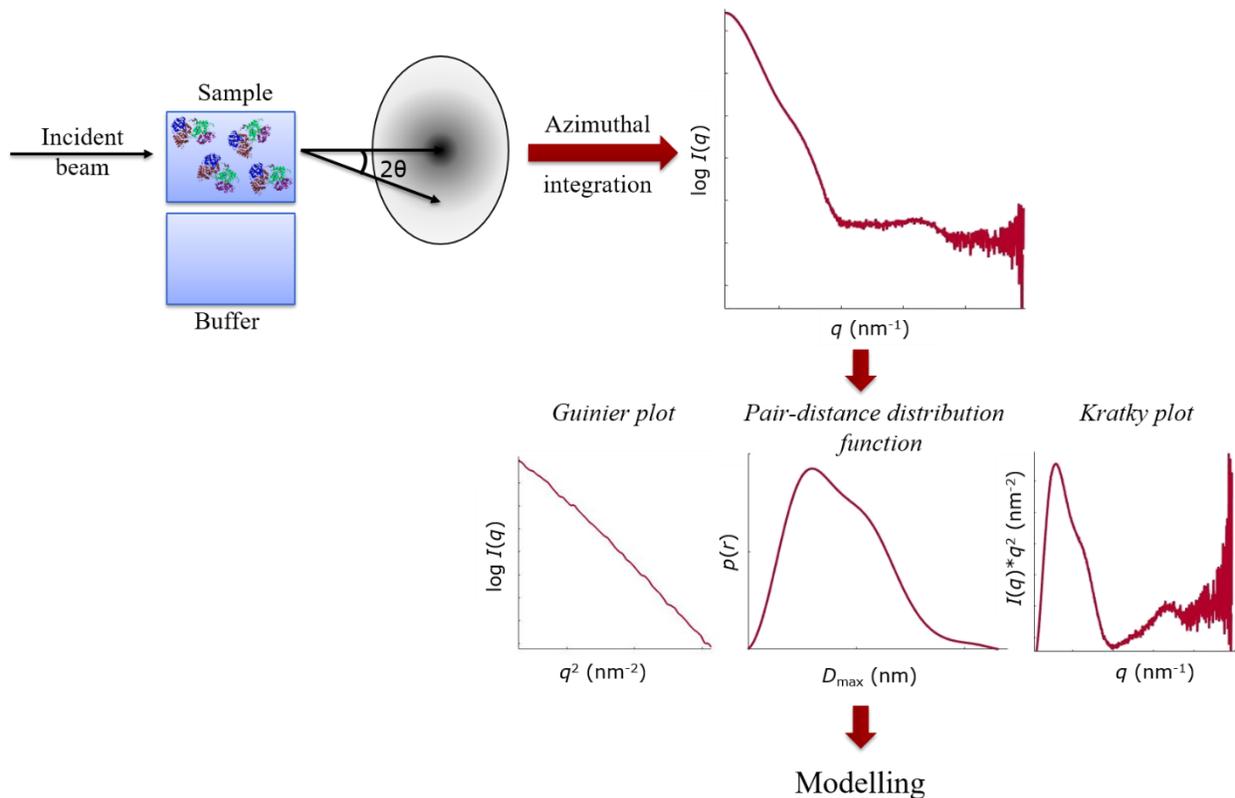


Figure 3.2: Schematic representation of SAXS. Sample scatters incident X-ray beam, which is detected on the detector and from which scattering profile of the solution is obtained. Subtracting solution scattering from buffer scattering results in protein scattering profile, which can be analyzed by multiple analyses (Guinier, pair-distance distribution, and Kratky). In case of good quality data and monodisperse solution it is possible to perform modelling based on SAXS data.

Guinier analysis ($\log(I(q))$ vs q^2) takes into consideration only initial slope of the scattering curve ($qR_g < 1.3$) and determines zero-angle intensity ($I(0)$) and R_g from the zero intercept. Moreover, $I(0)$ is proportional to the molecular weight and can be calculated according to formula below:

$$MW = \frac{I(0)N_A}{c(\Delta\rho v)^2} \quad (3.6)$$

where N_A is Avogadro's number, c is the concentration of the sample, $\Delta\rho$ is the contrast, and v is partial specific volume³⁶. In order to calculate MW it is necessary perform normalization of intensities from arbitrary units to absolute units. This can be performed by multiplication with a calibration factor, which can be determined by measuring the SAXS signal of pure water³⁶. Practically, experimental MW is often determined from well-behaved standard proteins with known MW (e.g. β -amylase and bovine serum albumin):

$$MW_{\text{exp}} = MW_{\text{st}} \frac{I(0)_{\text{exp}}/c_{\text{exp}}}{I(0)_{\text{st}}/c_{\text{st}}} \quad (3.7)$$

This requires assumption that v of standard and measured protein are similar³⁷, therefore normalization with pure water is usually more accurate.

Pair-distance distribution analysis takes into consideration more data points. The relationship between $p(r)$ function and $I(q)$ are described in eq. 3.8³⁶. The $p(r)$ provides information about the intermolecular distances and the maximum diameter (D_{max}) of the biomolecule. The $p(r)$ function strongly depends on the shape of the molecule and can vary as a function of scattering contrast between the protein and the buffer³⁶ (see Figure 3.3).

$$I(q) = 4\pi r^2 \int_0^{D_{\text{max}}} p(r) \frac{\sin(qr)}{qr} dr \quad (3.8)$$

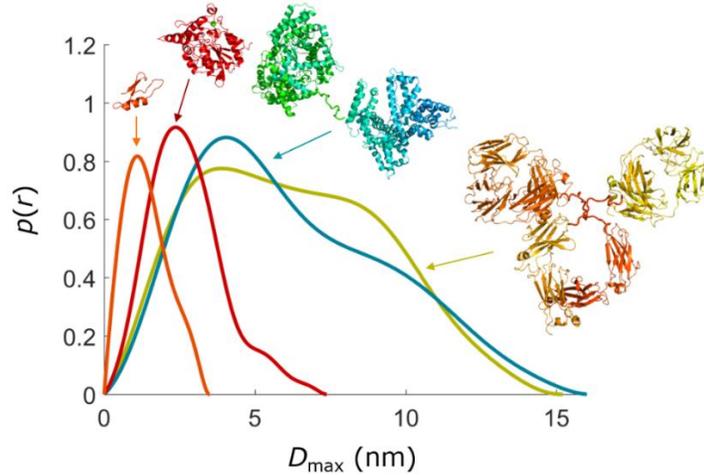


Figure 3.3: Example of $p(r)$ functions calculated from experimental data for plectasin (in orange), lipase (in red), albumin-neprilysin fusion protein (in cyan), and monoclonal antibody (in light green).

Porod analysis is used to determine hydrated volume of the biomolecules. It employs a Porod invariant - an integral of the area of the Kratky plot (see eq. 3.9) and allows to determine the Porod volume (V_p), which corresponds to 1.5-2 times the MW (in kDa). This determination does not depend on absolute scaling or protein concentration. Although, it is mostly used for globular protein, because Porod volume deviates for highly flexible and disordered proteins³⁶.

$$Q = 4\pi \int_0^{\infty} I(q) q^2 dq = 2\pi I_0 / V_p \quad (3.9)$$

Flexibility of the protein can be accessed using Kratky plot ($q^2 I(q)$ vs q), which allows to distinguish between globular and disordered protein. Globular protein has bell-shaped Kratky plot with well-defined peak, while disordered protein (flexible/unfolded) presents plateau instead (see Figure 3.4).

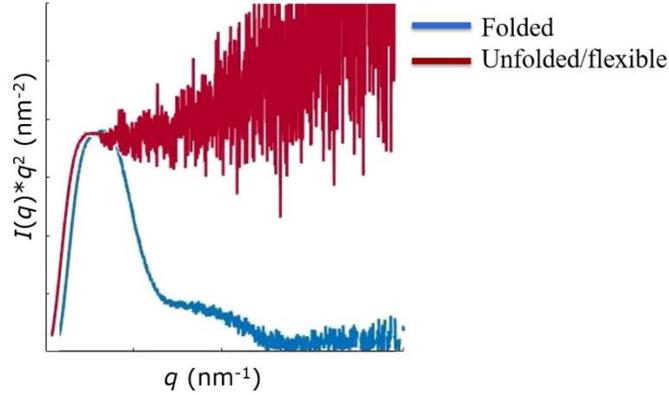


Figure 3.4: Kratky plot for folded globular protein (in blue) and unfolded/flexible protein (in red).

SAXS data at higher protein concentrations provide information about protein-protein interactions through the structure factor ($S(q)$). $S(q)$ can be calculated from the experimental intensity at concentration c using the eq. 3.10, where experimental data for low concentration (c_0) is assumed to not have any interactions³⁶.

$$S(q, c) = \frac{c_0 I_{exp}(q, c)}{c I(q, c_0)} \quad (3.10)$$

Increasing intensity with increasing protein concentration ($S(q) > 1$) is an indication of attractive interactions, while decreasing intensity ($S(q) < 1$) is an indication of repulsion (see Figure 3.5).

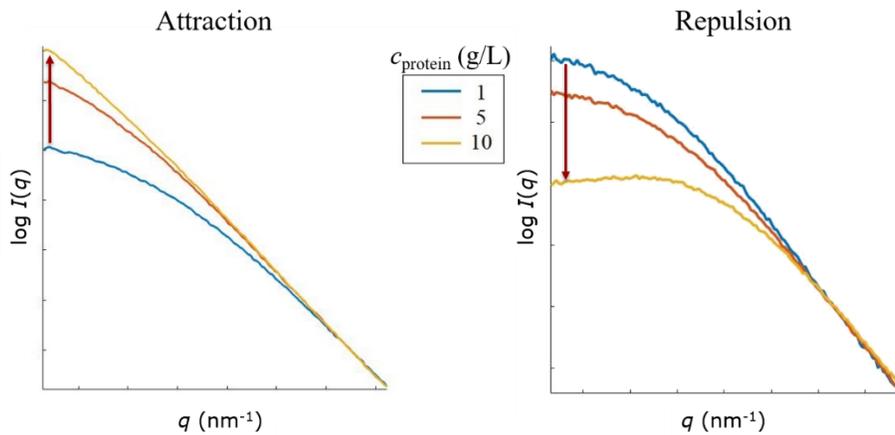


Figure 3.5: Guinier region for protein at different concentration illustrating attractive interactions (on the left) and repulsive interactions (on the right).

SAXS modelling

SAXS not only provides information about MW , R_g , D_{max} , flexibility, and interparticle interactions. It is also possible to perform modelling based on the SAXS data. The general principles of SAXS-based modelling is to find optimal parameters to describe the model that fits to an experimental data. Therefore, it is important to minimize discrepancy (χ^2) between the experimental data and computational model: acceptable χ^2 value is around 1.0; however, χ^2 highly depends on the data quality, as it will tend to be smaller, for noisy data. Therefore, it is important always to inspect how the model fits to the scattering curve. The challenge in reconstitution of three-dimensional model is that different models can provide the same scattering profile. To reduce this ambiguity, it is important to minimize function E (eq. 3.11), which includes χ^2 and a weighted penalties (P_i). Those penalties can impose physical constrains and structural restriction, known from complementary structural, computational or biochemical techniques³⁶.

$$E = \chi^2 + \sum \alpha_i P_i \quad (3.11)$$

Ab initio modelling

Ab initio modelling is used for reconstruction of the protein shape and can be done in the absence of structural information. They are two most commonly used methods, both using beads: bead modelling and dummy residues modelling³⁶.

In *DAMMIN*³⁸ (bead modelling) simulated annealing is employed to assign beads in the search volume (sphere with diameter D_{max}) either to the particle or to the solvent. The search starts with random distribution of solvent/particle beads, which randomly changes to achieve the optimal E . The penalty terms are used to ensure that the model does not consist of separate fragments. *DAMMIF*³⁹ is a newer available method with adaptive search volume and interconnectivity of the models, making the algorithm more efficient and 20-40 times faster (see example in Figure 3.6). Both *DAMMIN* and *DAMMIF* require $p(r)$ function³⁶.

In *GASBOR*⁴⁰ (dummy residues modelling) each aminoacid is represented by an averaged dummy residue centered at the approximate C_α atom position. The number of dummy residues corresponds to the number of amino acids in the protein. The search starts with a random special distribution of dummy residues and simulated annealing search is performed in spherical volume with diameter D_{max} . At each step, randomly selected dummy residue moves 3.8 Å from another dummy residue. Resulting model needs to be compatible with typical distribution of C_α atoms in proteins and theoretical scattering curve should fit to experimental data³⁶. Theoretical scattering curve are calculated using *CRYSOL*⁴¹.

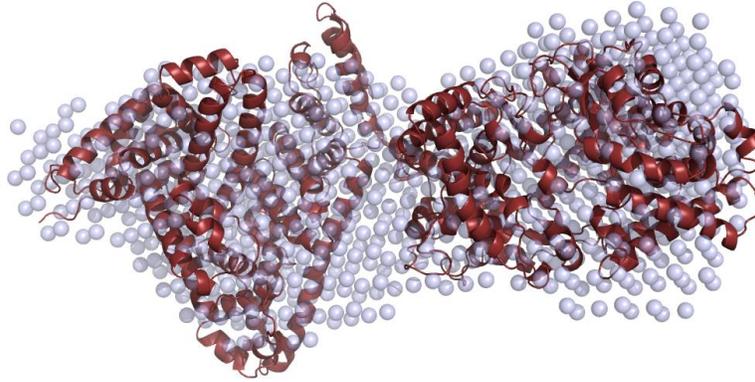


Figure 3.6: *DAMMIF* ab initio model for albumin-neprilysin fusion protein (in light blue) superposed to one of the models from *Ensemble Optimization Method* (in dark red).

Rigid body modelling

Rigid body modelling is used to model complexes and multidomain proteins, when the structures of each component is available. The relative position of individual rigid bodies (proteins/domains) are modelled to fit the SAXS experimental data (e.g. *SASREF*⁴²). In order to validate the obtained model, *CRY SOL*⁴¹ is used to calculate theoretical scattering curve of the model, which is compared with experimental SAXS curve.

When some parts of the structure are missing or protein has flexibility e/or conformational heterogeneity, hybrid modelling is applied. It consists in rigid body positioning in combination with conformational analysis of linkers using dummy residues-chain representation (e.g. *BUNCH*⁴²). In case of the missing fragments, hybrid modelling employs pre-generated library of loops (e.g. *CORAL*⁴³).

In order to perform analysis of multidomain proteins with conformational polydispersity, *Ensemble Optimization Method (EOM)*^{44,45} is often used. In *EOM* flexibility is taken into account by allowing co-existence of different conformations, which contribute to the scattering curve. Initially, *EOM* creates a large pool of possible conformation and computes theoretical scattering curves for each conformation. Then *EOM* employs a genetic algorithm (GA) to select subsets of conformation that minimizes χ^2 . It is also possible to perform pool generation and ensemble selection separately using, respectively, *RANCH* and *GAJOE* (see Figure 3.7A).

As already mentioned, most of the SAXS modelling programs are accounting for monodispersity of the sample. However, there are several programs that can be used to analyze polydisperse systems. *OLIGOMER*⁴⁶ fits an experimental scattering curve from multicomponent mixture to find volume fractions of each component. Given intensities from all the components (form-factors originated by *FFmaker*), *OLIGOMER* calculates volume fractions with minimal χ^2 (see Figure 3.7B).

It is also possible to perform rigid body modelling on polydisperse samples using *SASREFMX*⁴³, where experimental data is fitted by the linear combination of scattering profiles of oligomer and from the dissociation products (e.g. dimer and monomers) (see Figure 3.7C).

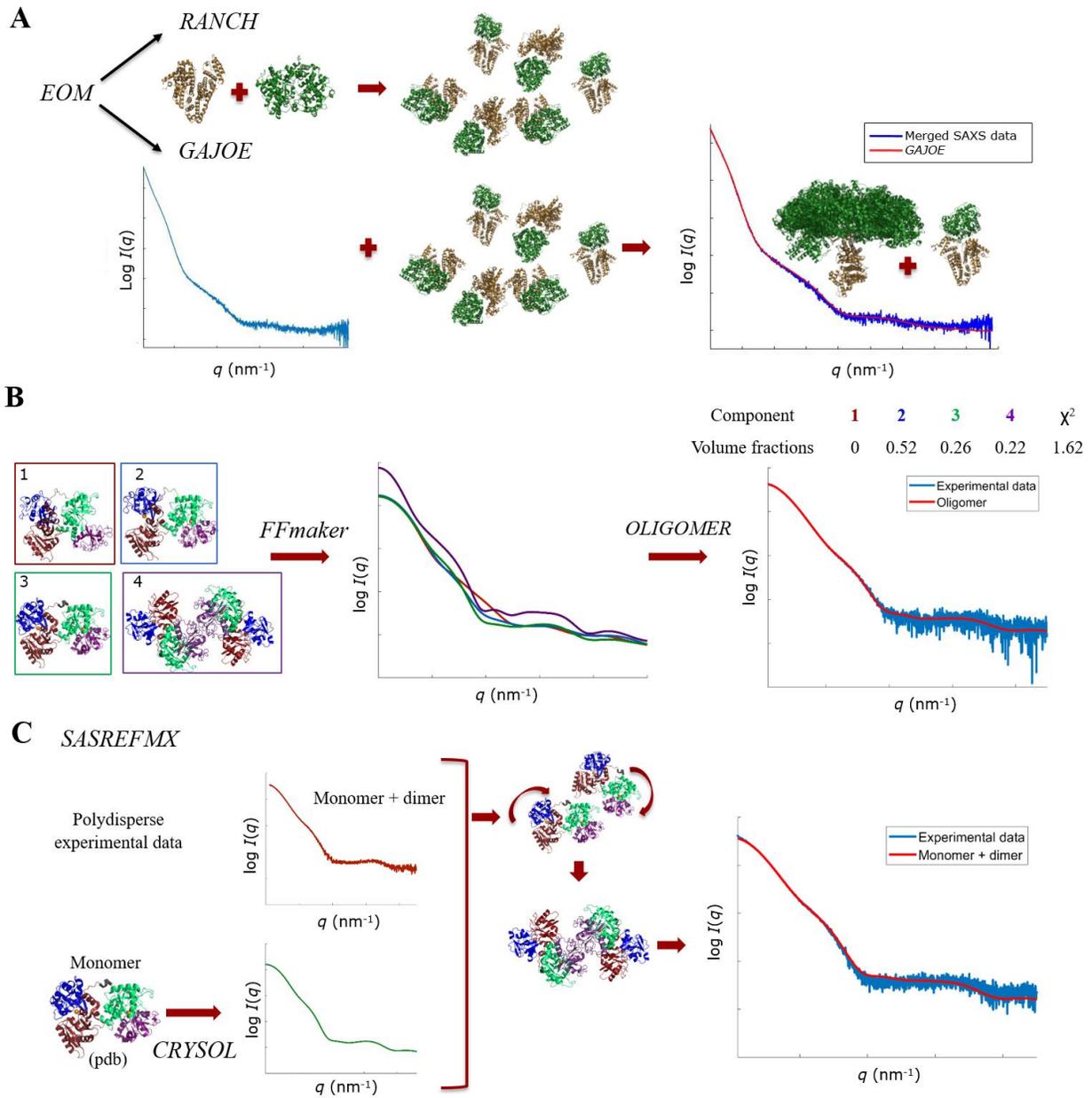


Figure 3.7: Schematic representation of SAXS data analysis for polydisperse solution. A: *EOM*, B: *OLIGOMER*, and C: *SASREFMX*.

MicroScale Thermophoresis

MicroScale thermophoresis (MST) is used to study and quantify biomolecular interactions by analyzing the direct motion of molecules through a temperature gradient⁴⁷. In thermophoresis, the temperature difference in space leads to depletion of solvated biomolecules in the region of higher temperature, which is quantified by Soret coefficient (S_T) and is described by eq. 3.12. The thermodynamic depletion depends on the interface between molecule and solvent⁴⁷⁻⁴⁹.

$$c_{hot}/c_{cold} = e^{-S_T\Delta T} \quad (3.12)$$

MST uses fluorescence to detect and quantify changes in the concentration of the molecules (c_{hot}/c_{cold})⁵⁰. Initially, homogeneously distributed molecules show a constant fluorescence. When IR-laser is turned on, it induces fast temperature changes that lead to a thermophoretic movement of labelled molecules out of the heated region. This is expressed by normalized fluorescence (F_{hot}/F_{cold}), which changes when protein is bound to the ligand, resulting in the different traces (see Figure 3.8A). Usually, MST experiment consist in titration of protein with the ligand, leading to the gradual changes that can be used to calculate binding constant (K_d)⁴⁷⁻⁴⁹ (see Figure 3.8B).

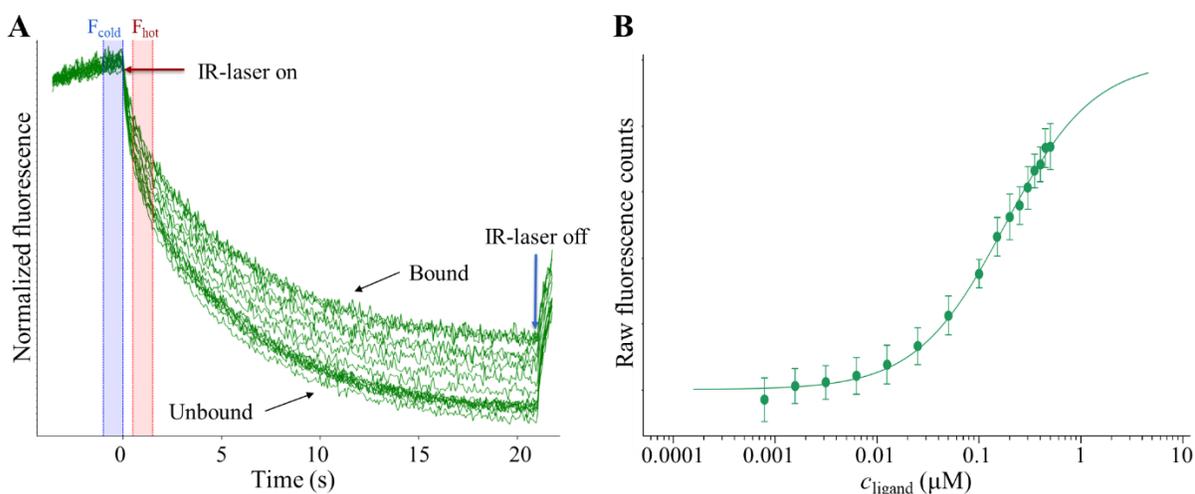


Figure 3.8: MST data representation. A: changes in normalized fluorescence with IR-laser on and off, in the absence (unbound) and presence (bound) of ligand. B: MST binding curve.

MST can detect binding of small molecules to proteins, substrates to enzymes or ligands to liposomes⁴⁷.

Chapter 4

Structural classification of protein therapeutics

Introduction

Since the introduction of the first proteinaceous drug, society and pharmaceutical industries have shown an increasing interest in protein therapeutics. The increased molecular understanding of several medical conditions led to development of protein therapeutics that now are used in the treatment of numerous diseases, *e.g.* auto-immune diseases and cancer. Every year significant number of biopharmaceuticals is approved by Food and Drug Administration (FDA) and many other are under clinical trials and will be approved in the near future^{51,52}.

One of the challenges in the development of protein-drugs is their formulation. Normally, proteins have narrow window of conditions in which they are stable. Due to the low demand, some protein-drugs need to be stored for a long time and in order to avoid the loss of function and aggregation, it is very important to choose right formulation conditions.

Protein stability, as well as every secret of protein behavior, lies in their three-dimensional fold. Changes in the structure due to protein-protein and protein-excipient interactions are the reason why certain proteins are stabilized under the certain conditions. Therefore, it is important to have an overview of the structural characterization of already existing protein-drugs. Knowledge of the protein's structure combined with knowledge of their formulation conditions will eventually lead to a connection between the two followed by an easier formulation process.

Current work is focused on structural classification of protein-drugs. Several databases were developed for the protein structural classification, such as SCOP (Structural Classification of Proteins)^{53,54}, CATH (Class Architecture Topology Homology)⁵⁵, 3DEE (The Protein Domain Database)⁵⁶, FSSP⁵⁷, and VAST^{58,59}. From all listed databases, SCOP has more stringent definition of the protein fold, where only proteins with similar core are putted into the same fold^{53,54,59}. Therefore, structural classification of existing protein-drugs was performed according SCOP database^{53,54}.

SCOP classification is based on the protein domain as fundamental unit. Protein domains are grouped at different levels according to their sequence and their structural and functional

relationship. Proteins are divided in several domains that are organized according to the following hierarchy: class, fold, superfamily, family, protein, and species (see Figure 4.1)⁵³. Here the class, the fold, and the superfamily are determined based on the 3D structure, whilst the family, the protein and species are based on the sequence: Species includes naturally occurred or artificially created variants with distinct protein sequences. Protein contains similar sequences with the same function that could be a different isoforms or originated from different species. Family contains proteins with similar sequences, but distinct functions. Superfamily includes different protein families that are inferred to be from a common ancestor and have a common structural and functional features. Fold and class are organized essentially by structural similarity: fold includes structurally similar superfamily and class is organized according to the secondary structure content.

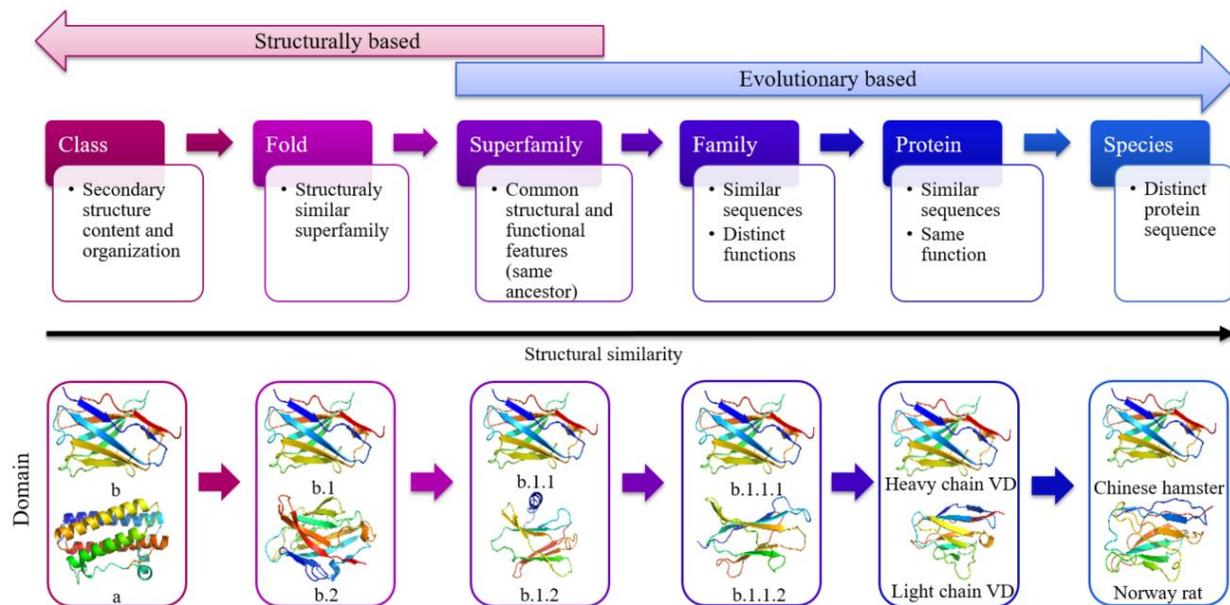


Figure 4.1: Protein classification hierarchy according to the SCOP database^{54,60}. Classification levels are illustrated by: b – protein immunoglobulin heavy chain variable domain Chinese hamster – *Cricetulus griseus* (b.1, b.1.1, b.1.1.1, Heavy chain VD, Chinese Hamster); a - Granulocyte-colony stimulating factor *Bos taurus*; b.2 - Diphtheria toxin C-terminal domain *Corynebacterium diphtheria*; b.1.2 - Erythropoietin (EPO) receptor *Homo sapiens*; b.1.1.2 – Protein beta2-microglobulin *Bos Taurus*; Light chain VL – protein immunoglobulin light chain variable domain Chinese hamster – *Cricetulus griseus*; Norway rat – protein immunoglobulin heavy chain variable domain Norway rat – *Rattus norvegicus*.

The SCOP database contains seven true classes that are organized according to secondary structure. *All alpha proteins* class contains the structures that are essentially formed by α -helices and *all beta proteins* class includes structures essentially formed by β -sheets. SCOP classification contains two different alpha and beta proteins classes: *a/b* class is composed by α -helices and mainly parallel β -sheets, whilst *a+b* class is composed by α -helices and mainly antiparallel β -sheets units. *Multi-domain* class includes proteins that have several domains that belong to different folds. The last two classes are *membrane and cell-surface protein and peptide* class and *small proteins* class. Example of the proteins for each class can be observed at the Figure 4.2.

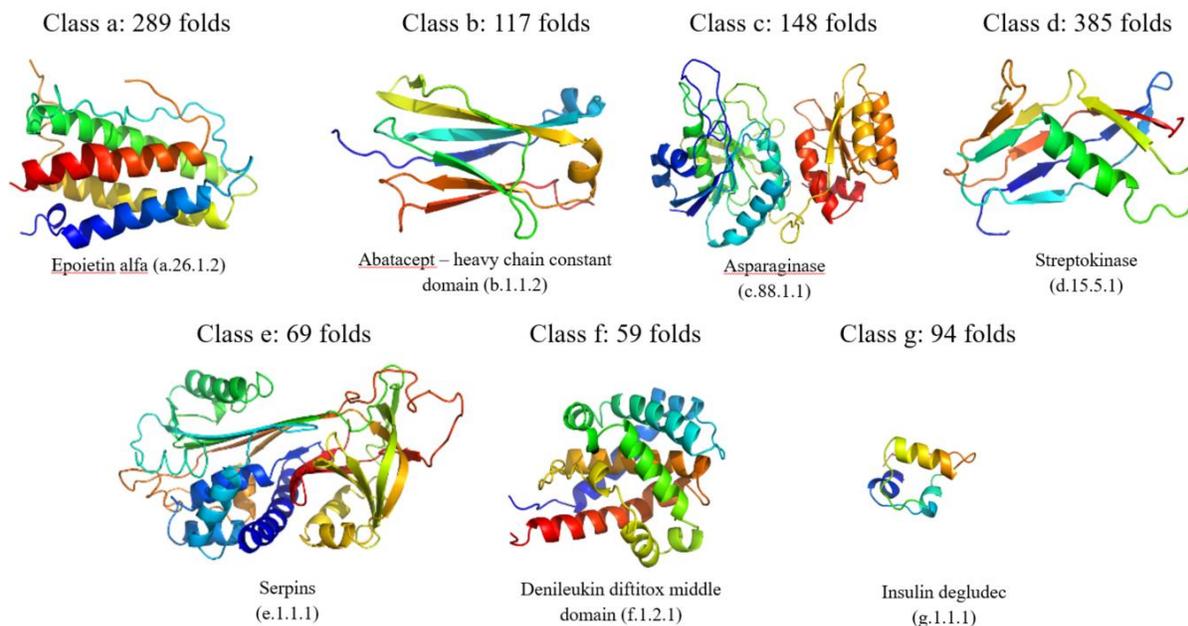


Figure 4.2: Protein examples for each true class of the SCOP database. Class a: *all alpha proteins*, class b: *all beta proteins*, class c: *alpha and beta proteins (a/b)*, class d: *alpha and beta proteins (a+b)*, class e: *multi-domain proteins*, class f: *membrane and cell-surface protein and peptide*, and class g: *small proteins*.

Moreover, the SCOP database also contains several additional not true classes: h) coiled coil proteins, i) low resolution protein structures peptides and fragments, j) peptides, k) designed proteins and l) artifacts. Those classes include proteins that were not possible classify due the information lack. Those proteins will be re-classified when more information becomes available.

Current study shows that most of protein therapeutics are distributed between three classes: *all alpha*, *all beta*, and *small proteins*. *All beta* and *small proteins* classes are mainly composed by monoclonal antibodies and insulins, which are structurally very similar within the class. Generally, protein drugs are not structurally diverse, occupying only around 2.1% of the total number of existing folds in SCOP database.

Results and discussion

Structural classification of protein therapeutics

After statistical data analysis it was possible to observe that the main part of protein therapeutics belongs to *all alpha*, *all beta*, and *small proteins* classes. As it is shown in a Figure 4.3, 40% of protein therapeutics are mainly composed by β -sheets, 11% by α -helices, and 18% are small protein therapeutics.

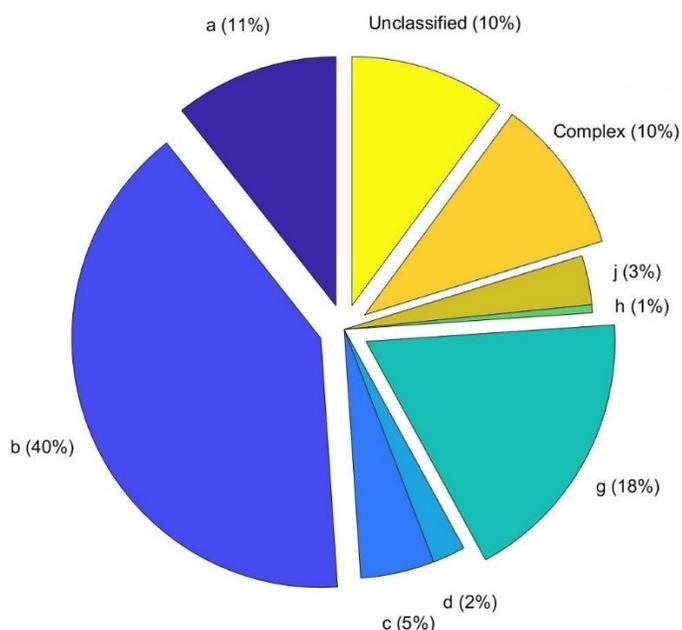


Figure 4.3: Distribution of different classes of protein therapeutics selected from FDA database (excluding diagnostic compounds). a – *all alpha proteins*; b – *all beta proteins*; c – *alpha and beta proteins (a/b)*; d – *alpha and beta proteins (a+b)*; g – *small proteins*; h – *coiled coil proteins*; j – *peptides*. Complex – includes protein-drugs with multiple active ingredients, and multidomain proteins, which domains belong to the different folds. Unclassified protein contains proteins which the sequence is not available or which is not classified in SCOP.

5% of protein therapeutics were classified in not true classes: 2% are coiled coil proteins (h class) and 3% are peptides (j class). As SCOP database contains only seven true classes, those proteins will be re-classified when more information will become available. One examples is botulinum toxin A coiled coil domain that belongs to coiled coil protein class. Also, 6% of the total number of protein therapeutics were not classified. Some of the antibodies, such as belimumab, and brentuximab vedotin, do not have amino acid sequence available. Other protein therapeutics, such as laronidase, sacrosidase, and nistrice, where sequence and structure are available, are not classified in SCOP database yet. Also, it is worth noting that the only example of a protein in the membrane protein class was denileukin diftotox middle domain. This indicates that membrane proteins have not been explored at all in terms of protein therapeutic development.

SCOP database englobes 1431 different folds and selected protein therapeutics were classified in 30 different folds, which corresponds to 2.1% of the total number of existing folds.

Almost 98% of existing folds in SCOP database are not explored in terms of protein therapeutics development.

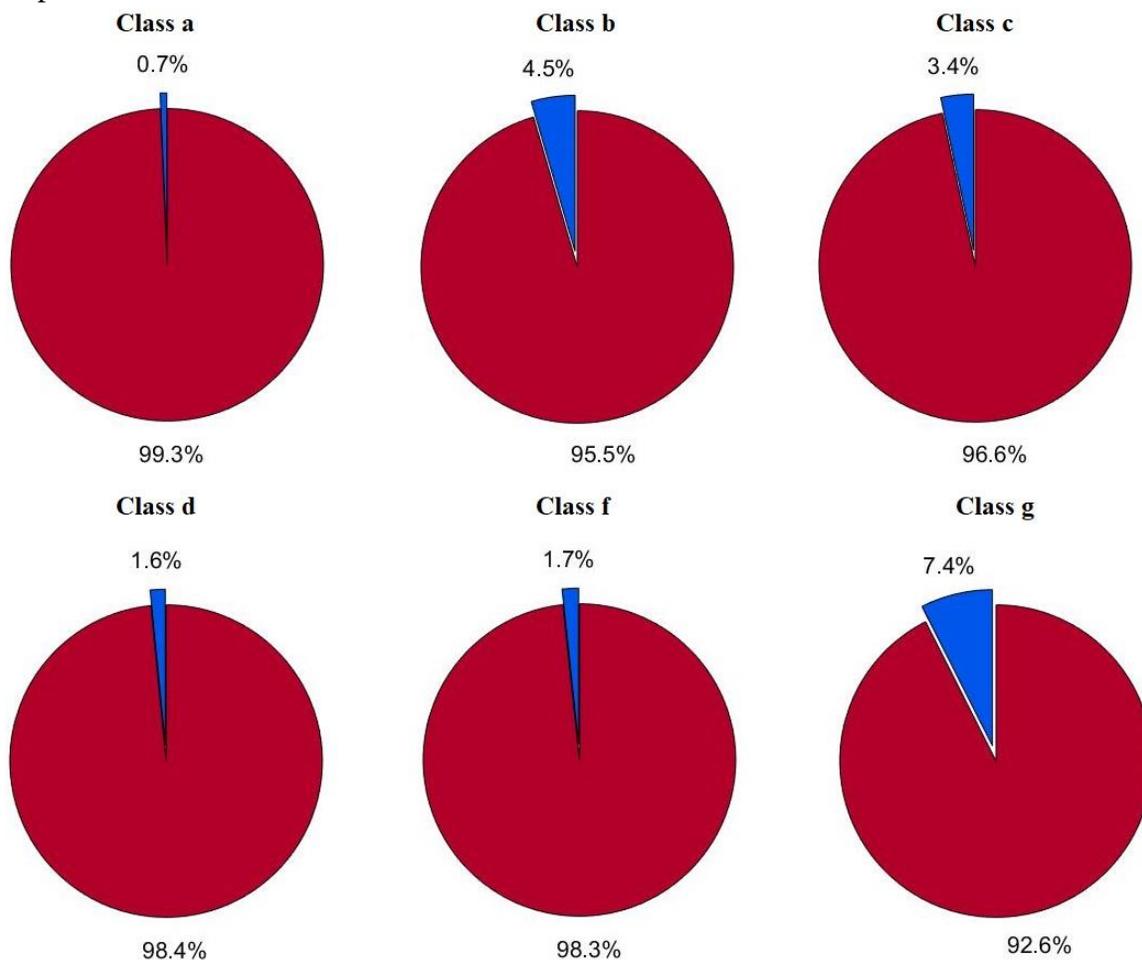


Figure 4.4: Protein therapeutic fold distribution for each class in comparison with total number of protein folds from SCOP database. In blue - folds that have at least one protein therapeutic domain included; In red - folds that have no single protein therapeutic associated.

It is possible to draft three most common folds: a.26, g.1, and b.1. a.26 is 4-helical cytokines fold that represents 13% of total protein therapeutics. g.1 is insulin like and b.1 is immunoglobulin-like beta-sandwich fold that correspond to 13% and 22% respectively.

Overview of approved protein therapeutics

Figure 4.5 shows how many new protein therapeutics have been approved the last 35 years. First generation protein therapeutics that were developed until 1980, includes mainly the small proteins, such as gonadotropin for tumor and insulin for diabetes. In the following five years, protein drug development was mainly focused on several insulin drugs, namely pork and beef insulins. Those were purified from animal pancreas and were introduced in different formulation conditions. The first recombinant human insulin was approved in 1982. Since 1996, when first insulin analog was approved, several insulin analogs were developed and introduced into the market.

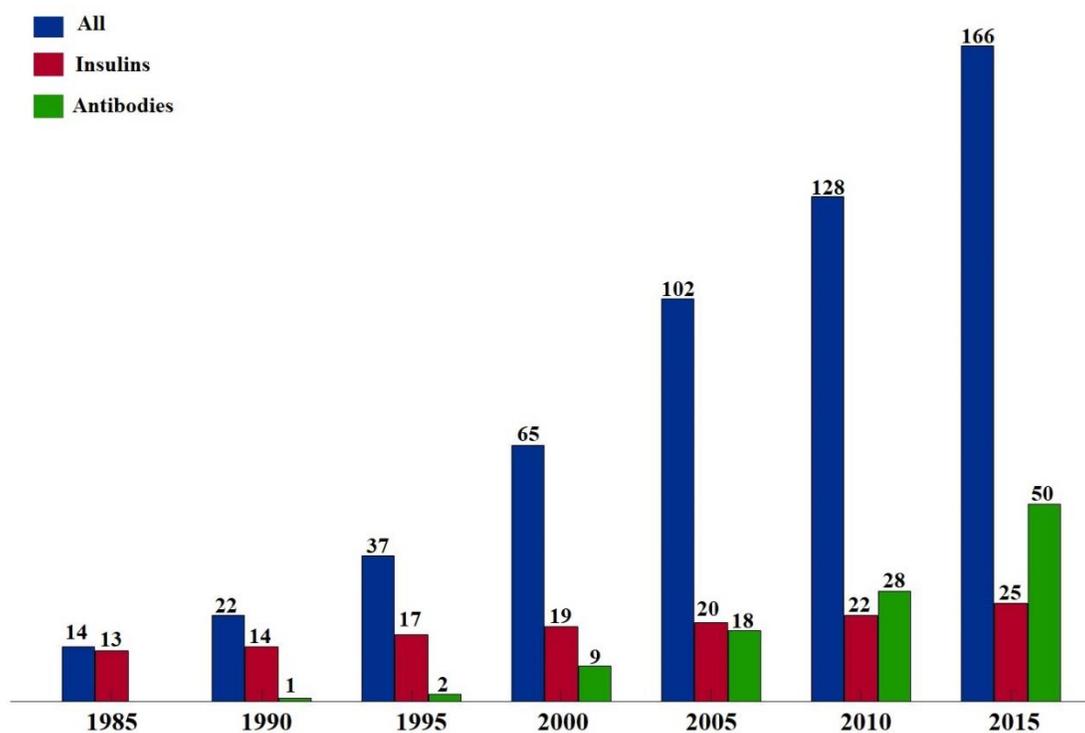


Figure 4.5: Total number of approved protein therapeutic from 1985 to 2015. All protein-drugs in blue, insulins in red, and antibodies in green.

One of the largest protein classes represented in Figure 4.3 are the monoclonal antibodies (IgGs). Muronomab-CD3 was the first mAb approved by FDA in 1986^{61,62}. However, a significant increase in the number of approved antibodies can be noticed from 2000 to 2005 where it has increased twice (see Figure 4.5). Reflecting that, this drug type has had a lot of focus in pharmaceutical industries due to its large variety of use.

Additionally, antibodies can be obtained from different sources and are classified as murine, chimeric, mixed, humanized, and human (see Figure 4.6)⁶³. Initially, several murine antibodies were introduced into the market^{61,64,65}, but due to their immunogenicity were later replaced by chimeric, humanized, and fully human. In the last 5 years, mainly human and humanized antibodies were approved.

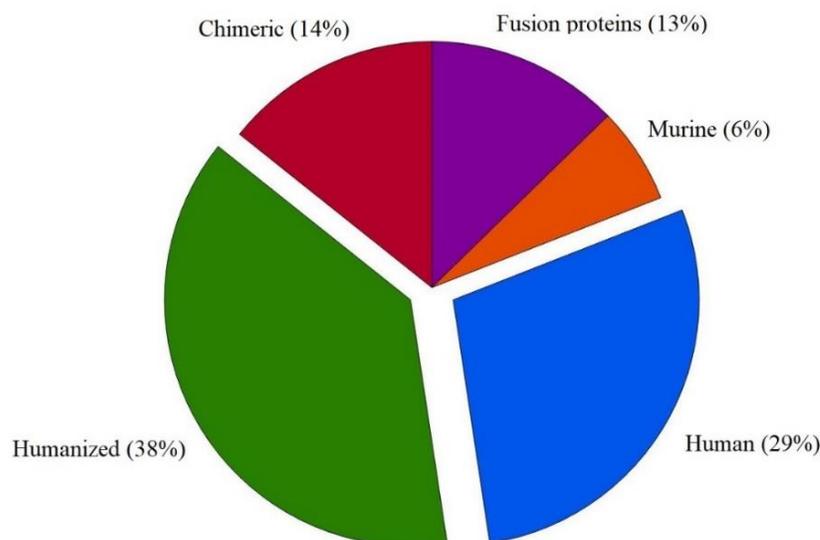


Figure 4.6: Distribution of different antibody types and fusion proteins approved up to 2016. Murine antibodies in orange, chimeric antibodies in red, humanized antibodies in green, human antibodies in blue, and fusion proteins in purple.

Conclusions

Current investigation showed that all protein therapeutics are mainly distributed between three classes: *all alpha*, *all beta*, and *small proteins*, where last two are mainly composed by antibodies and insulins. The number of represented folds/classes is very small when compared to the total number of folds listed in the SCOP database, including only 2.1%. Therefore, it is clear that there must be more protein therapeutics to be discovered.

This study lay the basis for the choice of the PIPPI database that was chosen to contain *all alpha*, *all-beta*, *alpha and beta* proteins (*a/b* and *a+b*), *small proteins*, and *multidomain proteins* (see Table 4.1), covering 5 classes that comprise 76% of all biopharmaceuticals (see Figure 4.3).

Table 4.1: Overview of selected proteins for PIPPI database.

	Class	Chosen protein	PPI nr
a	<i>All alpha</i>	HSA-NEP, interferon alpha, HSA	PPI18, PPI30, PPI49
b	<i>All beta</i>	Monoclonal antibodies	PPI01, PPI02, PPI03, PPI04, PPI08, PPI10, PPI11, PPI13, and PPI17
c	<i>Alpha and beta (a/b)</i>	Transferrin, lipases (<i>Thermomyces lanuginosus</i> and <i>Rhizomucor miehei</i>)	PPI44, PPI45, PPI46
d	<i>Alpha and beta (a+b)</i>	HSA-NEP	PPI18
e	<i>Multidomain proteins</i>	HSA-NEP	PPI18
g	<i>Small proteins</i>	Plectasins	PPI40, PPI41, PPI42, PPI43

Materials and methods

For the current study, it was decided to exclude diagnostic compounds and focus on the protein therapeutics with functions that are related to disease treatment.

Drug classification was started with selection of protein therapeutics from FDA database that also provided information regarding approval year, marketing status, drug target and disease. In order to complete this information and obtain details about amino acid sequence and CAS number, KEGG⁶⁶, drugs.com and drugbank.ca⁶⁷ databases were used. Some of the protein sequences were searched through SciFinder⁶⁸ database using the CAS number. For some proteinaceous drugs, approval year was found in DrugPatentWatch database⁶⁹.

For the structural classification several steps were performed as shown in Figure 4.7. In the first step each sequence was analyzed by protein BLAST⁷⁰ tool and it was possible identify PDB structure for some of protein therapeutics. In the second step structural characterization was performed. Each PDB structure was searched in SCOP^{53,54} database. For protein therapeutics, that do not have PDB file or where the structure is not classified in SCOP, the structure with highest similarity was used (>30%). Amino acid sequence of each domain was aligned to original protein therapeutic sequence using EXPASY⁷¹ tool. This step allowed to perform more accurate classification, especially for multi-domain proteins: in many cases protein therapeutic domains are 100% aligned to domains that belong to different proteins. Some of most common examples are fusion proteins, and humanized and chimeric antibodies.

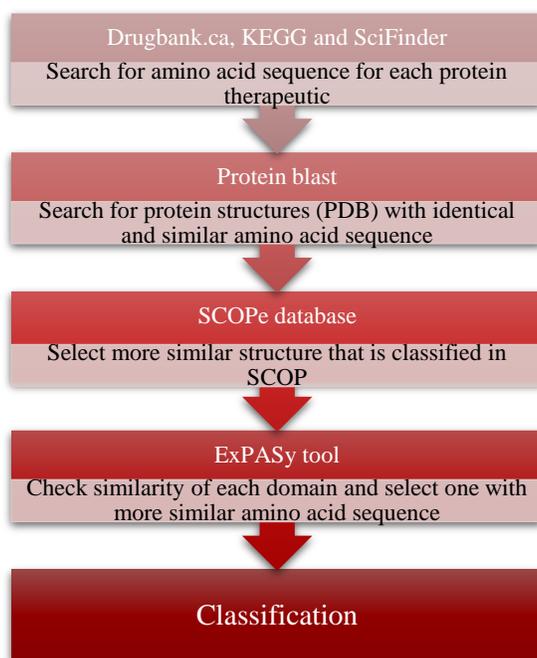


Figure 4.7: Protein classification strategy scheme. Drugbank.ca¹⁶, KEGG⁶⁶ and SciFinder⁶⁸ databases were used to complete information obtained from FDA database and obtain amino acid sequence. Blast⁷⁰ tool was used to search PDB structures with identical/similar amino acid sequence. SCOP^{53,54} database was used for classification of protein structures. Only identical sequences or sequences with similarity higher than 30% were considered. ExPASy⁷¹ tool was used to determine the level of similarity for each domain.

The structure classification includes class, fold, superfamily and family. Protein therapeutic with several active ingredients or multi-domain proteins, where the individual domains belong to different families, were classified in a separate sheet. Multi-domain protein drugs were not classified in class “e”, due to absence of information in SCOPe. Multidomain proteins, which individual domains belong to the same fold (*e.g.* antibodies) were classified in respective class, while the once, which domains belong to the different classes, were classified in “Complex”. That is why some protein therapeutic were considered multiple times for the statistical analysis.

One of the most common protein therapeutics is insulin. Despite the fact that all insulin drugs belong to the same family, they were separated according to amino acid sequence. The protein therapeutics, which have amino acid sequence identity lower than 30%, were included in unclassified group, as well as proteins which sequences were not available.

Chapter 5

Transferrin

Introduction

Human Serum Transferrin (TrF) is one of the most abundant proteins in human plasma⁷³. It is synthesized in the liver and is considered to be a major iron-carrying protein in the blood⁷⁴.

TrF is an 80 kDa multidomain protein composed by two similar domains/lobes: the N-lobe and the C-lobe, each carrying an iron(III) ion, and is connected by a linker peptide (see Figure 5.1). Each lobe is divided into two subdomains: N-lobe: N1 and N2, C-lobe: C1 and C2. Opening and closing of each lobe is restricted to the twisting of N2 and C2 subdomains⁷⁵. Those conformational changes are required for iron binding and release, which is essential for iron delivery to the cells.

The mechanism of iron delivery is illustrated in the Figure 5.2. It starts with binding of iron-loaded TrF to the transferrin receptor (TrFR) located on the membrane (Figure 5.2, step 1). The TrF-TrFR complex enters into the cell via endocytosis (Figure 5.2, steps 2/3). The acidic pH of the endosome contributes to the release of iron from the TrF, which is exported from the endosome by divalent metal transporter 1 (DMT1) (Figure 5.2, step 4). Finally, the TrF-TrFR complex is recycled to the cell surface (Figure 5.2, step 5)⁷⁶.

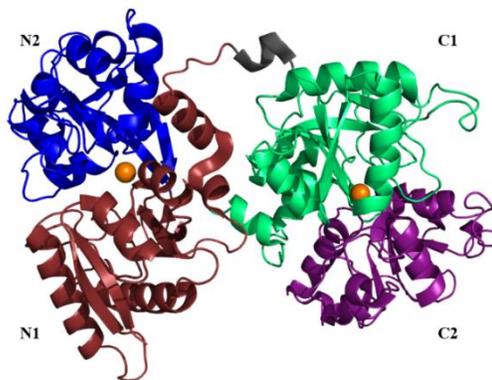


Figure 5.1: Crystal structure of TrF (pdbid: 3V83).

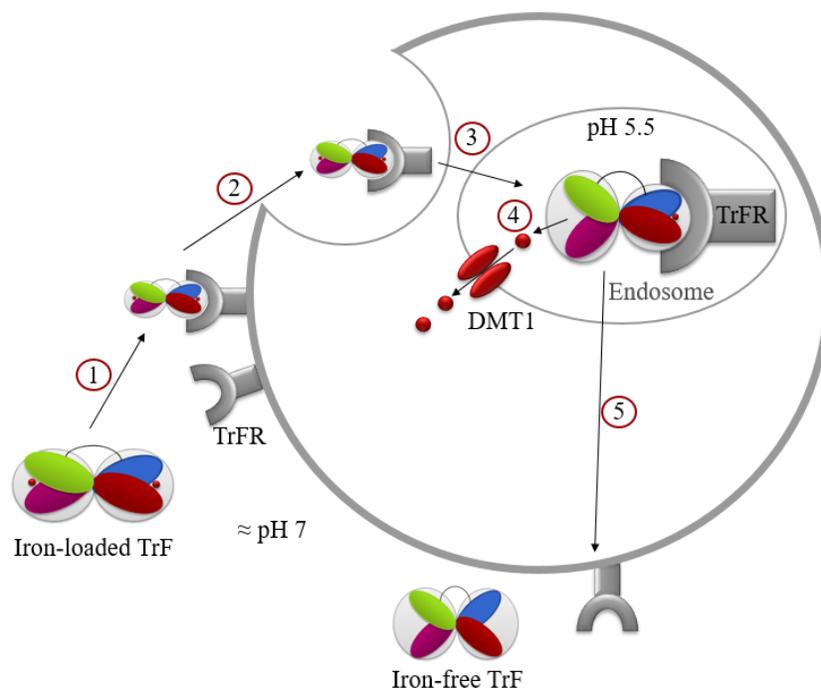


Figure 5.2: TrF mechanism of iron delivery (adapted from figure 1)⁷⁶.

Besides being essential for iron delivery in human organism, TrF has multiple therapeutic applications⁷⁷. In cases of atransferrinemia (genetic TrF deficiency), replacement therapy can take place by infusion of iron-free TrF (apo-TrF)⁷⁸. Administration of apo-TrF can decrease oxidative stress by binding to free iron, which will improve the condition of patients with ischemia reperfusion injury and acute renal failure⁷⁹. This is because presence of free iron catalyzes production of oxygen free-radicals, which leads to oxidative stress and, consequently, to inflammation and cell death⁸⁰. In some cases of diabetes, oxidative damage leads to decreasing of TrF levels by around 10%. Therefore, administration of apo-TrF can recover TrF levels resulting in reduced oxidative damage and reduced frequency of cardiovascular diseases⁸¹. Removing free iron also prevents growth of pathogens⁸². For instance, bacterial infections are common in bone marrow transplantation and usually lead to increased iron levels, which exceed binding capacity of TrF⁸³. Therefore, infusion of apo-TrF increases the binding capacity of free iron. Some studies suggest that apo-TrF can be used in cancer therapy. In combination with other drugs, TrF can promote cytotoxicity and proliferation in lymphokine-activated killer and natural killer cells⁸⁴. Finally, TrF can be used in targeted drug delivery. Different compounds, such as metal ions, small molecules proteins, and genes can be attached to TrF and can, through the TrF-TrFR pathway (see Figure 5.2) be delivered into the cells⁸⁵ and across the blood brain barrier⁸⁶.

Due to its potential applications, TrF is seen as an attractive protein to be used for therapeutic purposes. Therefore, understanding of its stability is crucial for potential drug development and the formulation process. We decided to investigate the overall stability of TrF and connect it to the conformational changes, which will lead to understanding the stability on the molecular level.

STRUCTURAL-STABILITY STUDIES ON RECOMBINANT HUMAN TRANSFERRIN

Alina Kulakova^a, Sowmya Indrakumar^a, Pernille Sønderby^a, Lorenzo Gentiluomo^{b,c}, Werner Streicher^d, Dierk Roessner^b, Wolfgang Frieß^c, Günther H.J. Peters^a, Pernille Harris^a

^aDepartment of Chemistry, Technical University of Denmark, Kemitorvet 207, 2800 Kgs. Lyngby, Denmark

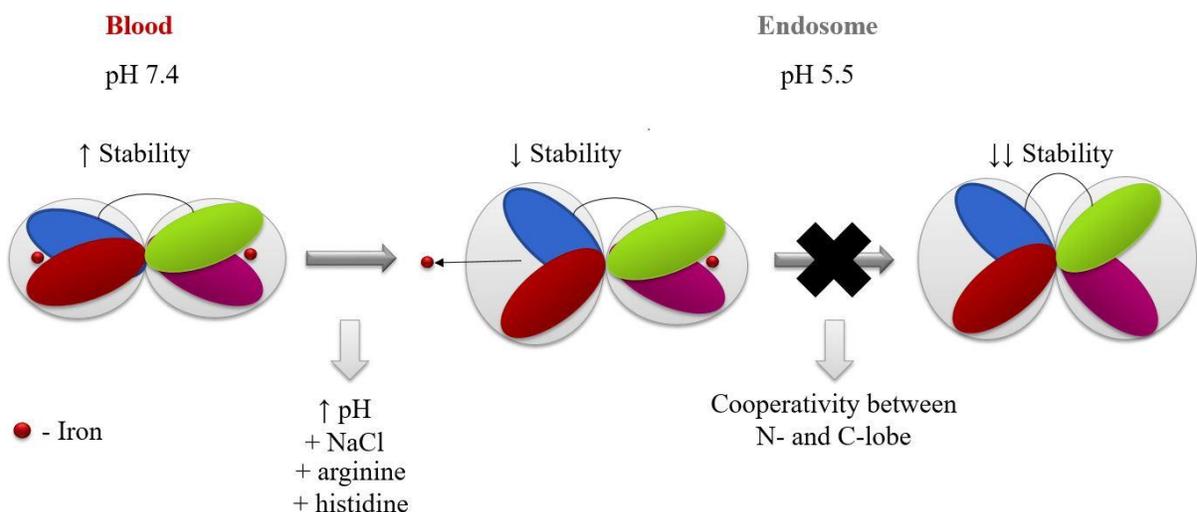
^bWyatt Technology Europe GmbH, Hochstrasse 18, 56307 Dernbach, Germany

^cDepartment of Pharmacy, Pharmaceutical Technology and Biopharmaceutics; Ludwig-Maximilians-University of Munich, Butenandtstrasse 5, 81377 Munich, Germany

^dNovozymes A/S, Biologiens Vej 2, 2800 Kgs. Lyngby, Denmark

Abstract

Transferrin is an attractive candidate for drug delivery due to its ability to cross the blood brain barrier. However, in order to be able to use it for therapeutic purposes, it is important to investigate how its stability depends on different formulation conditions. Combining high-throughput thermal and chemical denaturation studies with small angle X-ray scattering (SAXS) and molecular dynamics (MD) simulations, it was possible to connect the stability of transferrin with its conformational changes. The release of iron induces opening of transferrin, which results in a negative effect on its stability. Presence of NaCl, arginine, and histidine leads to opening of the transferrin N-lobe and has a negative impact on the overall stability.



Introduction

Over the last decades, the number of approved protein-based therapeutics has increased significantly and these drugs have become essential for the treatment of various diseases, such as diabetes, hemophilia, hepatitis C, and cancer⁶⁷. This is because, compared to small molecules, protein-based therapeutics show higher specificity and therefore, generally, have less side effects. However, due to their high complexity, protein-drugs are less stable and require special conditions (formulations) that will preserve their stability during production and storage. Under inappropriate conditions, proteins have a high tendency to unfold, which may lead not only to a loss of activity, but also to aggregation⁴. Unfortunately, no general rules for formulation have been reported, because it is not yet possible to predict the behavior of different proteins under different conditions. Therefore, it is important to obtain a detailed molecular understanding of the rationale behind protein stability and conformational changes.

Typically, the protein-drugs cannot cross the blood brain barrier (BBB), which is essential for treatment of certain diseases, such as Alzheimer's disease and brain cancer. One of the strategies to overcome this problem is to attach protein-drugs to a protein that is able to cross the BBB. Therefore, transferrin is an attractive candidate for drug delivery⁸⁷⁻⁸⁹, since it is one of the most abundant and stable proteins in human plasma⁷³, and it is able to cross the BBB through receptor-mediated endocytosis⁹⁰.

Human serum transferrin (TrF) is a major iron-carrying protein in the blood. TrF regulates iron levels in biological fluids, and not only supplies the cells with ferric iron, but also prevents production of radicals in the blood by removing free iron⁷⁴. TrF is a multidomain protein, which is composed of two similar lobes: the N- and the C-lobe each of them binding an iron ion hereafter referred to iron. The lobes alter between open and closed conformation by binding and releasing iron. TrF has been crystallized in three different conformations: open⁹¹, partially open⁹², and closed⁷⁵ (see Figure 5.3). It has an open conformation when both lobes are free of iron. In the partially open conformation, in the presence of the transferrin receptor iron is believed to be bound to the N-lobe⁹³ (Fe_N-TrF) with the C-lobe open. However, in the absence of the transferrin receptor, iron release is faster in N-lobe⁹³. Therefore, the crystal structure for the partially open conformation has the N-lobe open⁹² with iron bound to the C-lobe (Fe_C-TrF). Finally, TrF has a closed conformation when iron is bound to both lobes.

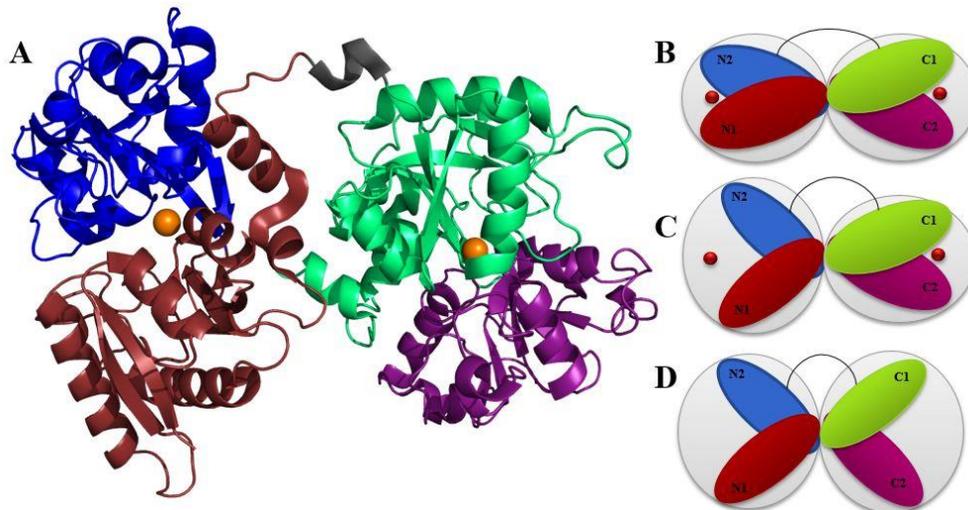


Figure 5.3: A: Crystal structure of the closed conformation of TrF (pdbid: 3V83)⁹¹ and conformational representations for B: closed, C: partially open, and D: open conformations.

Iron release and conformational changes have been studied using a variety of techniques, including small angle scattering (SAS). Before TrF's crystal structures became available, SAS studies indicated that TrF has spheroidal shape⁹⁴. In the presence and absence of iron, TrF showed differences in SAS curves and distance distribution functions that pointed towards conformational differences⁹⁵. Reported values for the radius of gyration were lower in the presence of iron, which suggested a more compact conformation^{96–98}. It has also been reported that the release of iron is pH-dependent and is induced by decreasing pH⁹⁸. Moreover, kinetic studies have shown that iron release is influenced by sodium chloride (NaCl). At neutral pH, chloride ions retards iron release, while at acidic pH, it accelerates iron release⁹⁹. In addition, it has been shown that the mechanism of iron release is a complex process that involves cooperativity between the lobes^{100–102}.

This study is focused on thermal and chemical denaturation of recombinant human transferrin (rTrF) in different pH and salt concentrations and with different co-solutes. These studies are combined with structural analyses performed by Small-Angle X-ray Scattering (SAXS) and molecular dynamics (MD) simulations. The SAXS results confirmed previously reported results on the effect of pH and NaCl on the conformation: rTrF shifts towards an open conformation with decreasing pH⁹⁸ and with addition of NaCl at low pH⁹⁹, which has a negative impact on overall stability. Moreover, it was shown that arginine, which is used as common stabilizer in protein formulation, binds to rTrF destabilizing the protein as indicated by an up to 20°C decrease in the temperature of unfolding ($T_{1/2}$). MD simulations are in agreement with the SAXS results and show that NaCl, arginine, and histidine induce opening of the N-lobe.

Results

The overall conformational stability of recombinant transferrin (rTrF) was analyzed by thermal denaturation using nano differential scanning fluorimetry (nanoDSF) and isothermal chemical denaturation (ICD). Two different denaturants, urea and guanidine hydrochloride (GuHCl) were tested. Due to the high conformational stability of rTrF, urea was not strong enough to unfold it completely (see Figure 5.4C), therefore, only GuHCl unfolding data were analyzed. The initial screen was performed as a function of pH (5-9) and ionic strength (0, 70, and 140 mM NaCl). NanoDSF thermal unfolding shows a single two-step transition (from folded to unfolded state), while chemical denaturation curves demonstrate the presence of an intermediate state, resulting in a three-state transition. In addition, the intermediate state is better defined at lower pH values (see Figure 5.4). Only the first transition in the chemical unfolding curves was further considered, as this is where unfolding process is initiated. Due to the poorly intermediate state, the calculated ΔG_{unfold} shows high standard deviations and was therefore not considered for analysis.

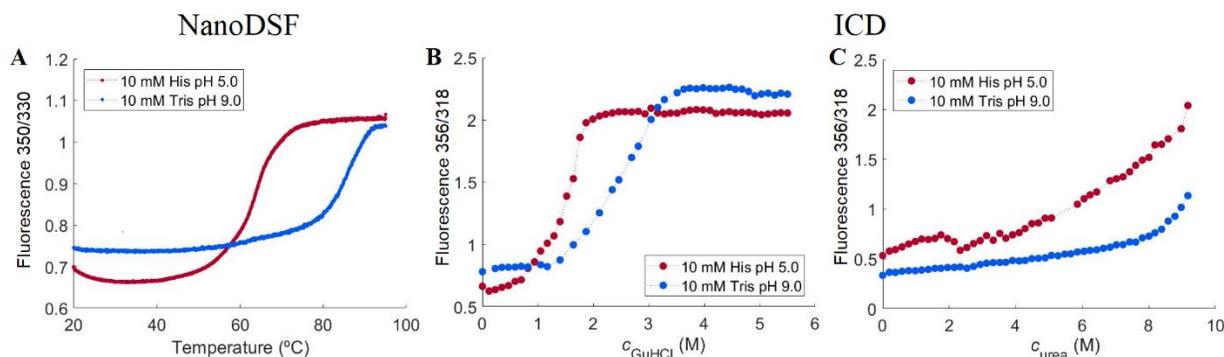


Figure 5.4: rTrF thermal and chemical unfolding curves. A: thermal unfolding curves from nanoDSF, B: chemical unfolding curves in the presence of GuHCl from ICD and C: chemical unfolding curves in the presence of urea from ICD. rTrF in 10 mM histidine (His) pH 5.0 (red), and 10 mM tris pH 9.0 (blue).

pH dependence

$T_{1/2}$ measured by nanoDSF is shown in Figure 5.5A. An increase in $T_{1/2}$ with increasing pH is seen, meaning that the thermal stability is higher at higher pH values. Likewise, chemical denaturation shows an increase in the amount of GuHCl needed to unfold 50% of the protein ($c_{1/2}$) with increasing pH values (see Figure 5.5B).

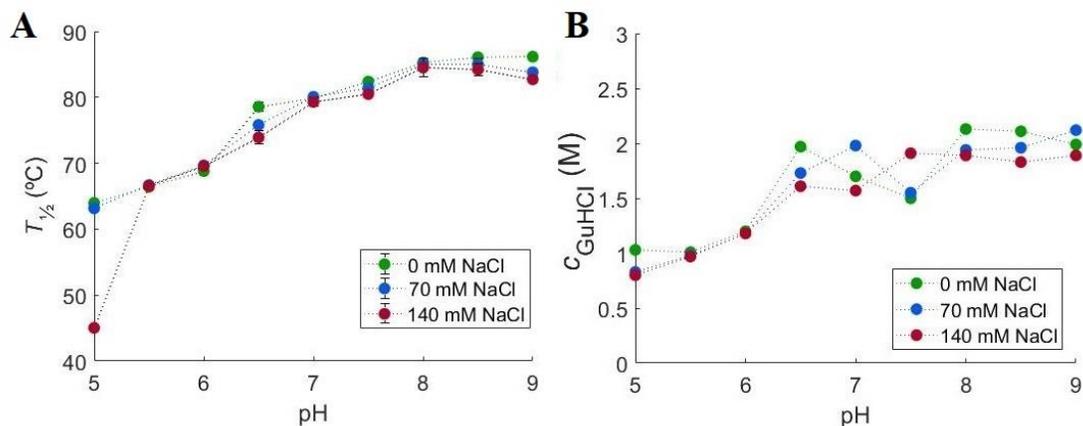


Figure 5.5: Initial stability studies performed by NanoDSF and ICD at different pH and ionic strengths. A: changes in $T_{1/2}$ and B: changes in $c_{1/2}$ with pH in the presence of 0 mM (green), 70 mM (blue), and 140 mM (red) NaCl.

In order to study conformational changes, SAXS concentration series data were collected at pH 4.0, 5.0, 6.5, and 8.0 with 0 mM NaCl (see Table S5.1 in Appendix B). All scattering curves and SAXS data analyses are shown in the Appendix B. At pH 6.5 and 8.0, the curves coincide and the intensity at low q -values decreases with increasing rTrF concentration, indicating a repulsive system. Contrary to this, at pH 5.0 the intensity increases with rTrF concentration, which is characteristic for aggregation. Both aggregation and repulsion are observed at pH 4.0 (see Figure S5.1A in Appendix B). Moreover, four representative curves (shown in Figure 5.6) differ in shape depending on pH, which indicates conformational dissimilarity. Finally, we observe that the estimated molecular weight (MW) at 1 g/L for most of the conditions is higher than the expected: 75 kDa (see Table S5.3 in Appendix B), which means that a substantial fraction of the protein molecules form larger species.

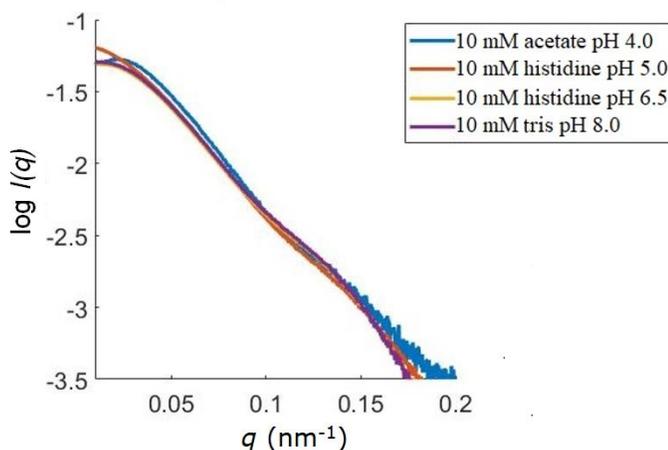


Figure 5.6: Comparison of SAXS curves from $c_{\text{rTrF}} \sim 10$ g/L collected at different pH. Blue: acetate pH 4.0; red: histidine pH 5.0; yellow: histidine pH 6.5; and purple: tris pH 8.0.

In order to characterize the size of the larger species, SEC-MALS was performed (see Table S5.5 and Figure S5.5 in Appendix B), confirming the presence of approximately 12% dimer and 2% trimer at all tested conditions. Additionally, static light scattering was performed as a function of protein concentration showing a concentration independent *MW* (data not shown).

It is known that rTrF exists in different conformations: open, partially open, and closed, which is related to iron binding and release⁹⁶. In order to evaluate the rTrF conformation at different pH values, *OLIGOMER*⁴⁶ analysis was performed using pdbid: 2HAU⁷⁵ for the completely open conformation, pdbid: 3QYT⁹² for the partially open conformation (Fe_C-rTrF) and pdbid: 3V83⁹¹ for the closed conformation as input, while the dimer was modelled by *SASREFMX*⁴³. Due to the very small amounts of trimer, this species was not taken into consideration.

The *OLIGOMER* analysis is shown in Figure 5.7 (see also Table S5.4 in Appendix B). It shows that at 10 mM histidine pH 6.5 and 10 mM tris pH 8.0 rTrF is in the closed conformation (Figure 5.7D and E). At 10 mM histidine pH 5.0 and 10 mM acetate pH 5.0 rTrF is present in closed and partially open conformation (Figure 5.7C). By changing to 10 mM acetate buffer pH 4.0, it allowed us to detect a small amount of the open conformation. At pH 4.0, rTrF shifts from partially open to open conformation, while the fraction of the closed conformation remains constant.

NaCl dependence

Overall thermal stability is independent of the NaCl concentration, except at pH 5.0, where 140 mM NaCl cause a decrease in $T_{1/2}$ from 65°C to 45°C (see Figure 5.5A). At higher pH values the addition of NaCl does not show any effect. The ICD experiments did not show an NaCl effect on $c_{1/2}$.

SAXS experiments were performed at 5 g/L rTrF in 10 mM histidine pH 5.0 and 6.5 with increasing NaCl concentrations. At pH 5.0, a gain in $I(0)$ is seen with increasing c_{NaCl} due to rising *MW* (up to 100 kDa), which points to the presence of aggregates (see Table S5.3 in Appendix B). The *OLIGOMER* output shown in Figure 5.7 shows an increase in the volume fraction of higher *MW* species. The observation of larger aggregates is in agreement with NanoDSF and ICD results showing lower conformational stability at pH 5.0 with increasing salt concentration. At pH 6.5, the repulsion decreases when NaCl is added, leading to higher $I(0)$ (see Figure S5.1H in Appendix B).

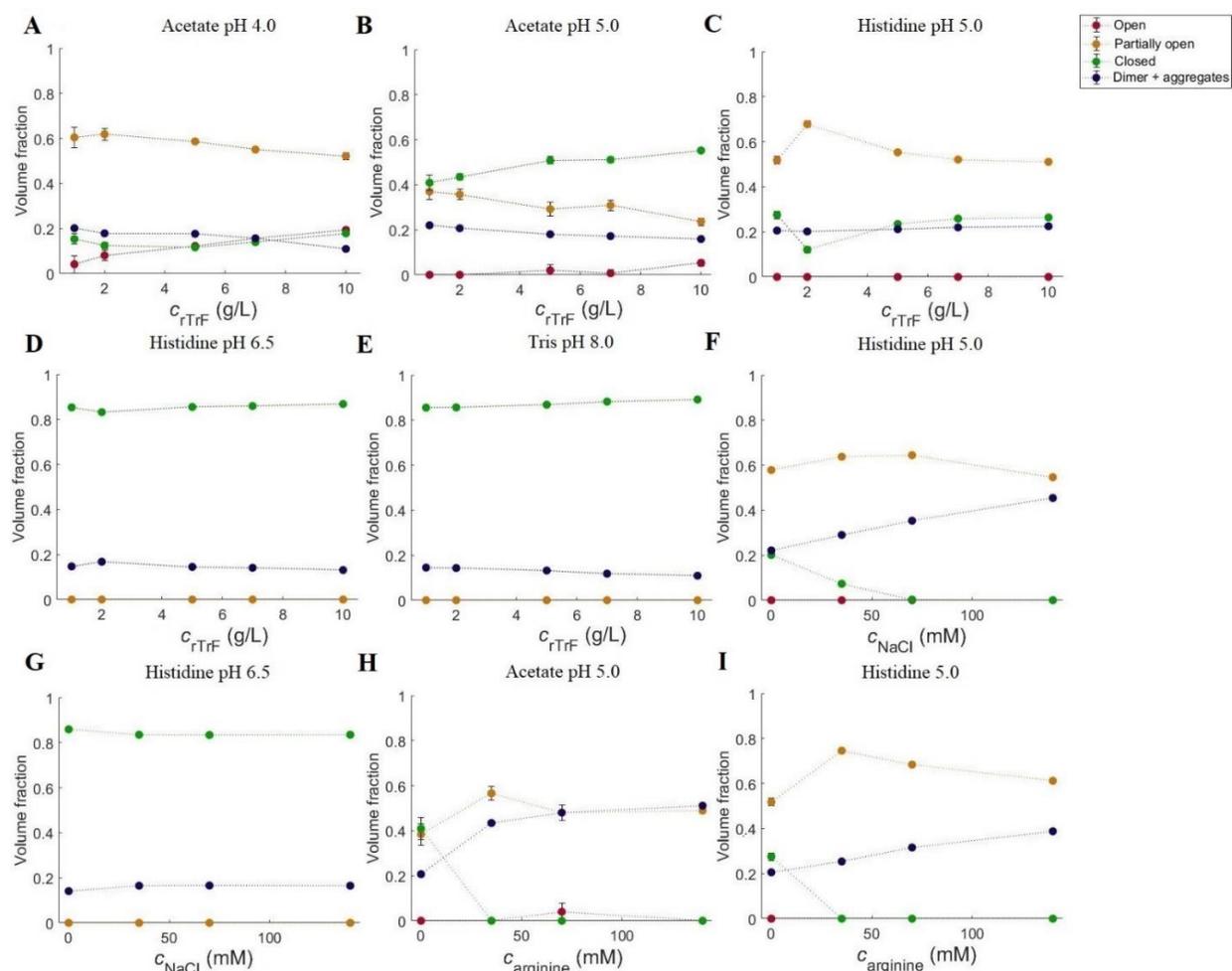


Figure 5.7: Fraction of different monomer conformations (open (in red), partially open (in orange), and closed (in green), and dimer (in blue)). A: acetate pH 4.0; B: acetate pH 5.0; C: histidine pH 5.0; D: histidine pH 6.5; E: tris pH 8.0; F: 5 g/L rTrF, histidine pH 5.0 with NaCl; G: 5 g/L rTrF, histidine 6.5 with NaCl; H: 5 g/L rTrF, acetate 5.0 with arginine; I: 5 g/L rTrF, histidine 5.0 with arginine.

Buffer and excipient dependence

For the investigations of excipient and buffer effects histidine buffer at pH 5.0 and 6.5 with 0 or 140 mM NaCl, as well as acetate pH 5.0 and phosphate pH 6.5 were selected. Furthermore, three different excipients, 280 mM sucrose, 140 mM arginine, and 280 mM proline, were tested.

$T_{1/2}$ from nanoDSF and $c_{1/2}$ from ICD are shown in Figure 5.8. With respect to the buffer dependence effect, it is seen that at pH 5.0 rTrF has a 5°C higher $T_{1/2}$ in acetate buffer compared to histidine buffer, while addition of 140 mM NaCl to the histidine buffer decreases $T_{1/2}$ by 15°C. The ICD measurements show a somewhat different picture for histidine buffer as the addition of NaCl does not influence $c_{1/2}$, which is already significantly lower in the histidine buffer compared to the acetate buffer.

At pH 6.5, $T_{1/2}$ is about 15°C higher in histidine buffer than in phosphate buffer. Addition of 140 mM NaCl to the histidine buffer decreases $T_{1/2}$ by 5°C. The ICD results are similar as $c_{1/2}$ is

reduced by 1 M in phosphate buffer, while in histidine buffer, addition of NaCl only leads to a very small decrease.

Adding sucrose or proline at pH 5.0 leads to minor effects on $T_{1/2}$. In contrast, arginine has enormous impact on rTrF stability. Addition of arginine leads to a decrease of $T_{1/2}$ by 20-25°C at pH 5.0 except when 140 mM NaCl is present. Both arginine and 140 mM NaCl reduce $T_{1/2}$ by around 20°C, however, adding both of them together does not alter this already low $T_{1/2}$ (see Figure 5.8A). Chemical denaturation shows a different picture, where addition of excipients in histidine buffer does not have an effect on $c_{1/2}$. In acetate buffer $c_{1/2}$ is 1.6 M and addition of arginine or proline leads to a decrease of $c_{1/2}$ by approximately 1 M.

At pH 6.5 arginine decreases $T_{1/2}$ by 5-10°C in histidine buffer, but has no effect in phosphate buffer, where $T_{1/2}$ is already low. Excipients do not show an effect on $c_{1/2}$ at pH 6.5 in both buffers (see Figure 5.8D).

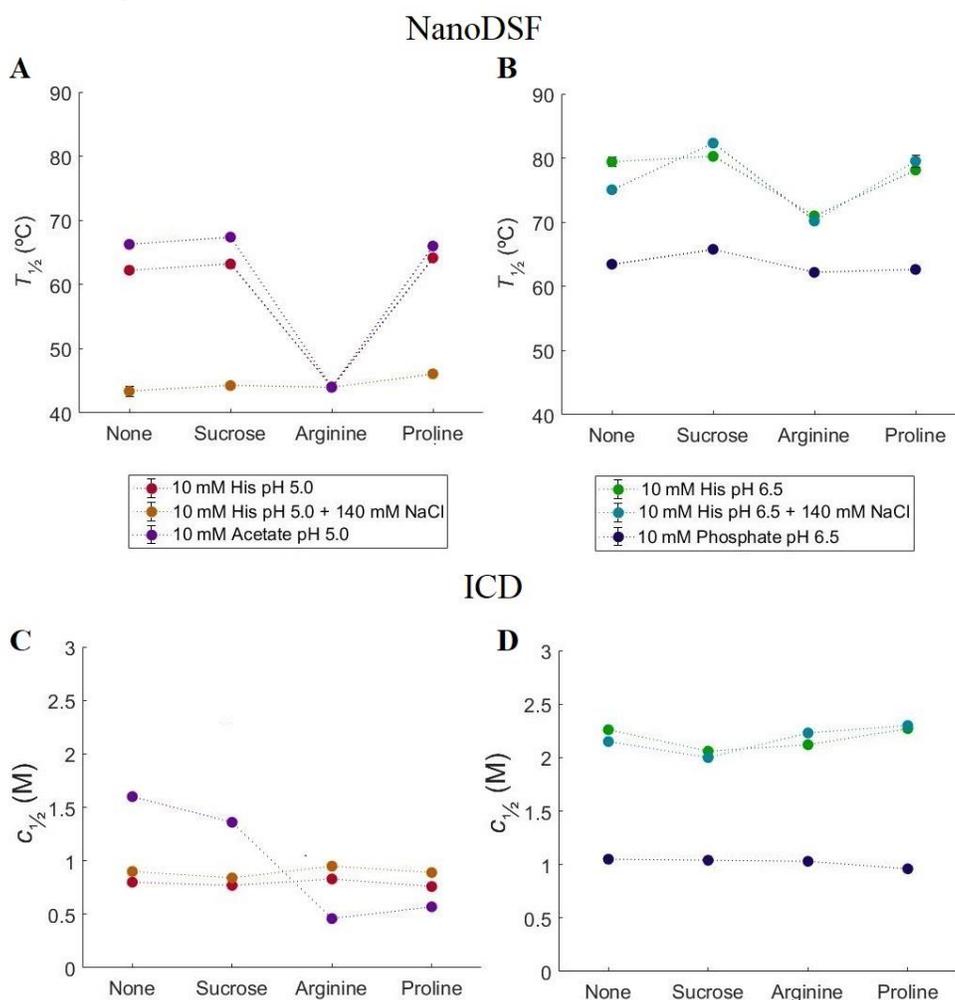


Figure 5.8: NanoDSF and ICD stability studies using different buffers and excipients. Purple: 10 mM acetate pH 5.0; red 10 mM histidine pH 5.0; orange: 10 mM histidine pH 5.0 with 140 mM NaCl; blue: 10 mM phosphate pH 6.5; green: 10 mM histidine pH 6.5; cyan: 10 mM histidine pH 6.5 with 140 mM NaCl. A: changes in $T_{1/2}$ at histidine (0 and 140 mM NaCl) and acetate pH 5.0; B: changes in $T_{1/2}$ at histidine (0 and 140 mM NaCl) and phosphate pH 6.5; C: changes in $c_{1/2}$ at histidine (0 and 140 mM NaCl) and acetate pH 5.0; D: changes in $c_{1/2}$ at histidine (0 and 140 mM NaCl) and phosphate pH 6.5.

The negative effect of arginine on the thermal rTrF stability can be explained by arginine binding to the protein. This was tested by performing MST using proline as a negative control, which does not have a significant effect on thermal stability. MST results show that arginine binds weakly to rTrF with K_d of 0.180 M (see Figure 5.9).

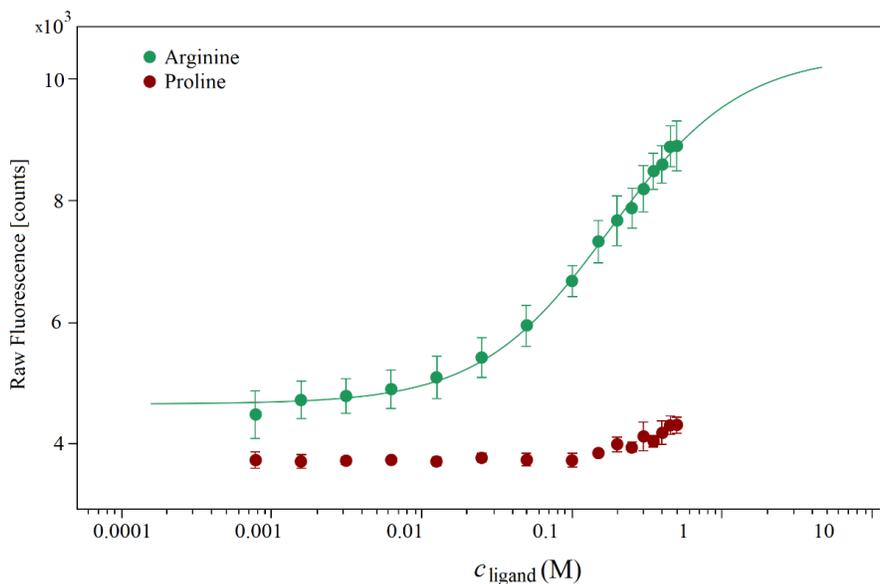


Figure 5.9: MST binding curve for arginine (in green) from K_d -fit using proline (in red) as a negative control.

SAXS data were collected for rTrF in histidine and acetate buffers at pH 5.0 (see Figure S5.1 in Appendix B) and analyzed using *OLIGOMER* (see Figure 5.7B and C; and Table S5.4 in Appendix B). In both buffers, partially open and closed conformations are present, with the volume fraction of the partially open conformation decreasing with increasing rTrF concentration. However, acetate buffer shows higher volume fractions of closed conformation.

Molecular Dynamics simulations

In order to understand the interactions between rTrF and the other components in the solution at selected pH and buffer conditions, MD simulations were performed in the presence of NaCl, histidine, arginine, acetate, and phosphate.

MD simulations were performed for 100 ns in the presence of bound Fe^{3+} and bicarbonate (CO_3^{2-}) in both lobes. All systems reached a constant root mean square deviation after 3 ns (data not shown). Patches that comprise at least three residues structurally close on the protein surface and have moderately strong interactions are colored based on $P(I_{\text{score}})$ (see Figure 5.10). $P(I_{\text{score}})$ given to a residue helps to deduct the preference of different additives on the protein surface, which in turn leads to an understanding of the different mechanism related to stabilization and iron release. At pH 5.0, both arginine and histidine are positively charged (+1), while acetate is negatively charged (-1). Even though excipients interact with the protein at several regions, only few patches interacting with the additives are relatively large and strong. Generally, arginine,

histidine, and NaCl bind more strongly in the C-lobe as compared to acetate, which binds stronger in the N-lobe. A particular region on the C-lobe that is a common interaction site for the different buffer components consists of residue D416 and D420 (see Figure 5.10A, B, and C).

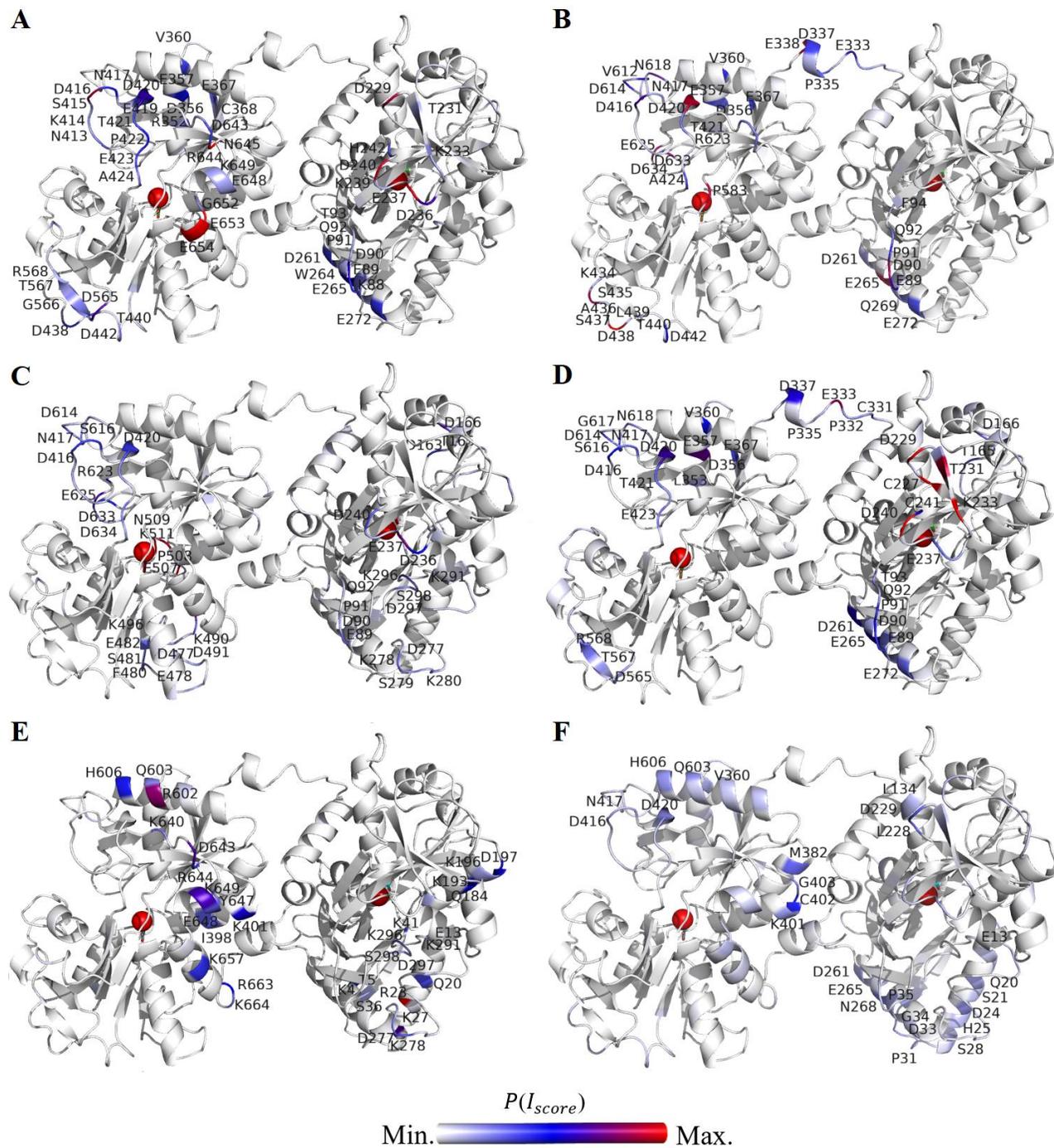


Figure 5.10: The closed structure of rTrF colored at the patches interacting strongly with buffer components based on the $P(I_{score})$ calculated at pH 5.0. A: arginine, B: histidine, C: NaCl, and D: acetate; and at pH 6.5 for E: phosphate, and F: histidine. The C-lobe is shown on the left and the N-lobe on the right. Single letter code for amino acids is used.

At pH 6.5, phosphate shows many weak interactions with single amino acids, and relatively large, but weak interacting patches on the protein surface that are highlighted in the Figure 5.10E. In contrast, although interactions with histidine are spread across the protein surface, there are few prominent patches with strong interactions (see Figure 5.10F).

Discussion

pH effect

pH-dependent conformational changes in rTrF are aligned with its physiological function, which is binding and transporting iron into the cells. Due to the high toxicity of iron, it is important that rTrF remains in the closed conformation in blood⁷⁴. Blood has a pH of 7.4 where according to the SAXS data only the closed conformation is present. However, rTrF should be able to supply cells with iron. Iron release occurs under acidic condition in the endosome^{103,104}, where the pH is around 5¹⁰⁵. SAXS data analysis shows presence of the partially open conformation at pH 5.0, which supports conclusions of previous studies that prove conformational changes of rTrF being pH-dependent⁹⁸. Our SAXS studies do not show evidence of the presence of a fully open conformation at pH 5.0. At pH 4.0, a small fraction of fully open conformation was detected, accompanied by increasing aggregation (see Figure S5.1A in Appendix B). This suggests that the presence of the fully open conformation may induce aggregation and therefore, it is important to have a mechanism that reduces the possibility of full opening of rTrF. This data supports that iron release is not only pH-dependent but also involves cooperativity between the N- and the C-lobe¹⁰⁰⁻¹⁰².

Thermal and chemical denaturation showed a decrease in physical stability with decreasing pH, at which the partially open conformation is present. Since this conformation has a higher solvent-accessible surface area than the closed conformation, it may unfold more easily.

NaCl effect

With addition of NaCl, $T_{1/2}$ at pH 5.0 decreases by 20°C and SAXS data indicate aggregation of rTrF. As already noted, at pH 5.0 both partially open and closed conformation are present. Addition of NaCl decreases the volume fraction of the closed conformation, and increases the one of the partially open conformation with the N-lobe open (see Figure 5.7F). Opening of the N-lobe can be explained by interactions of NaCl on the protein surface. Especially, the loop regions (residues 89-94, 277-280, and 296-298), which are close to the iron binding cleft, are prone to strongly interact with salt ions. Previous studies have shown that crosstalk between the lobes leads to iron release first occurring from the N-lobe^{93,100}. At this end, strong interaction in the C-lobe loop region (D416, D420) might be inducing conformational changes that result in iron release from the N-lobe. (see Figure 5.10C). This is in agreement with previous studies, where NaCl has been proven to accelerate iron release at acidic pH, due to the higher anion-binding affinity of rTrF⁹⁹ when compared to higher pH. In addition, presence of the partially open conformation, which has lower stability, contributes to the aggregation process.

At higher pH values the presence of NaCl does not induce significant changes in thermal stability. The SAXS studies confirmed that rTrF is only present in the compact conformation at pH 6.5, and that the addition of NaCl has no impact (see Figure 5.7G). In a previous study, chloride was shown to slow down iron release at neutral pH⁹⁹.

Excipient effect

Amongst the tested excipients, arginine has a pronounced negative effect on the stability of rTrF, especially at pH 5.0 where $T_{1/2}$ decreases by 20°C and $c_{1/2}$ is reduced by 1M. According to SAXS results obtained with higher arginine concentration the fraction of partially open conformation increases at the expense of the closed conformation (see Figure 5.7H and I and Table S5.4 in Appendix B), leading to an aggregation. MST confirms weak arginine binding to rTrF at pH 5.0, whereas MD simulations show strong interactions in both C- and N-lobes (see Figure 5.10A). Adding proline in acetate has a slightly destabilizing effect seen in $c_{1/2}$, but not in $T_{1/2}$. Furthermore, MST did not show binding of proline.

Buffer effect

The buffer type has a clear effect on the protein stability⁴. At pH 5.0, replacing histidine by acetate buffer positively affects rTrF stability, while at pH 6.5, histidine buffer is preferable over phosphate buffer.

SAXS studies show that in acetate buffer at pH 5.0 the volume fractions of the closed and partially open conformations are around 0.4 (see Figure 5.5G and Table S5.4 in Appendix B). In histidine buffer, the conformation of rTrF is shifted towards partially open, increasing the volume fraction to 0.5 (see Figure 5.7I and Table S5.4 in Appendix B). According to the MD studies, histidine has stronger binding to the C-lobe as compared to the acetate, which binds stronger in the N-lobe. However, both have a common loop region (89-94) around the iron binding site in the N-lobe, where Y96 coordinates with iron⁹². Previous studies have shown that in the absence of the TrF receptor, the mechanism of iron release starts with opening of the N-lobe⁹³. This suggests that histidine, due to the stronger interactions with residues 89-94 in the N-lobe loop region, induces its opening and shifts equilibrium towards the partially open conformation, resulting in the lower stability.

As already mentioned, at pH 5.0, histidine, arginine, and NaCl shift the equilibrium from the closed to the partially open conformation. All of them have strong interactions with regions in the N-lobe, particularly around the iron binding site comprising of loop region (89-94), which might cause a change in electrostatic field around the region leading to conformational changes that induces the opening. Additionally, they have one common interaction patch comprising D416 and D420 and others residues around (Figure 5.10A, B, and C), pointing to a crucial role in rTrF's conformational changes. These residues are present on the loop region close to C-lobe cleft, but not directly connecting the two subdomains. However, this loop is prone to high fluctuations as reflected in MD simulations (data not shown) and might be involved in the cooperativity between

two lobes, since conformational changes in this region can lead to lobe-lobe interaction, and contribute to the iron release from N-lobe.

At pH 6.5 both phosphate and histidine have weak to negligible interactions in the loop region of the C-lobe and also in the loop region (89-94) of the N-lobe (see Figure 5.10E and F). In both buffers, rTrF is present only in the closed conformation, pointing to the involvement of these two loops (89-94, 416-420) in the iron release mechanism. Phosphate has a destabilizing effect compared to histidine, which might be due to few patches interacting strongly with phosphate on the protein surface. Contrary, histidine interacts weakly over many small patches on the protein surface. The overall charge of the phosphate at pH 6.5 is -2, while histidine is neutral, making phosphate more likely to interact strongly with the exposed hydrophilic patches as compared to histidine. Additionally, the preferential interaction coefficient values are higher for phosphate as compared to histidine implying higher preference of the protein surface for phosphate (see Figure S5.6 in Appendix B).

Conclusion

The presented work is a systematic study of the overall physical behavior of rTrF in a variety of different buffer conditions combined with structural studies using SAXS and MD simulations. The increase of $T_{1/2}$ and $c_{1/2}$ are both indicators of increased conformational stability. Although, some of the trends seem to be similar for these two indicators, some specific differences are seen probably because in one experiment temperature increases and in the other experiment a chemical compound is added (GuHCl). However, combining denaturation results with volume fractions of closed and partially open conformations seen in the SAXS studies (see Figure 5.11), it is possible to observe a decrease in volume fraction of the partially open conformation with increasing $T_{1/2}$ and $c_{1/2}$, and a corresponding increase in volume fraction of the closed conformation. Several conditions, such as the presence of arginine, NaCl, buffers, and pH changes can lead to opening and, consequently, to a decrease in rTrF stability. MD simulations indicate that this occurs due to the binding of the additives to the loop regions of the C-lobe, causing its opening for iron release.

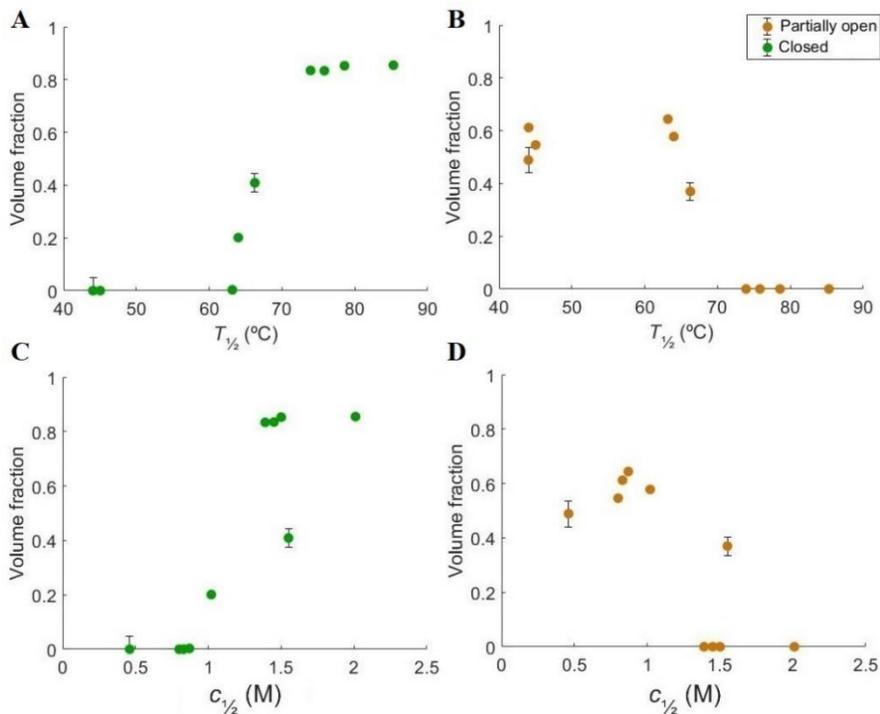


Figure 5.11: Volume fractions of different rTrF species correlated to thermal and chemical denaturation studies. A and B: volume fractions correlated to $T_{1/2}$; C and D: volume fractions correlated to $c_{1/2}$. Partially open conformation colored in orange and closed conformation colored in green.

Methods

Dialysis and formulation

Recombinant human transferrin (rTrF) was provided by AlbuMedix Ltd. in 20 g/L solution and was dialyzed into 10 mM histidine pH 5.5, 7.0, and 10 mM tris pH 8.5 for pH and NaCl screening. Concentration of rTrF was measured on a Nanodrop™ 1000 (Thermo Fisher Scientific, Waltham, USA) using extinction coefficient calculated from the primary sequence⁷¹ (see Table S5.2). For stability studies with different buffers and excipients, dialysis was performed at 10 mM histidine pH 5.0 and 6.5, 10 mM acetate pH 5.0, and 10 mM phosphate pH 6.5. Final solutions were obtained by diluting rTrF into the right buffer (with $\text{pH} \pm 0.1$) (see Figure 5.12).

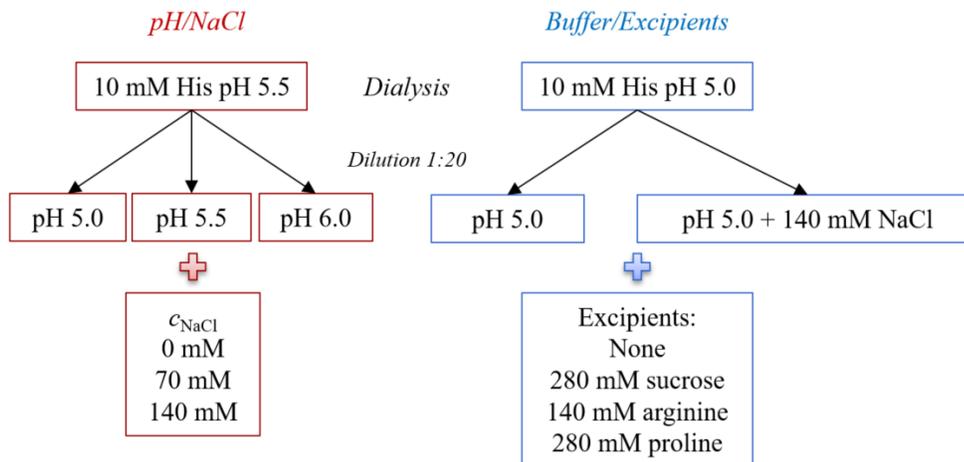


Figure 5.12: Schematic representation of dialysis and formulation process.

Thermal stability studies

Thermal denaturation studies were performed with the Prometheus NT.48 (NanoTemper Technologies, Munich, Germany). NanoDSF Grade Standard Capillaries were manually loaded with 10 μ l of protein at 1 g/L in the final conditions. All experiments were performed from 20 to 95°C with a linear thermal ramp using the heating rate of 1°C/min. Protein intrinsic fluorescence was measured and the unfolding process was monitored by looking at the shift in the fluorescence spectra (350/330 nm). All measurements were done in triplicates and the data analysis was performed using PR.Control v1.12.2 software (NanoTemper Technologies, Munich, Germany).

Chemical Denaturation

All chemical denaturation studies were performed on Unchained Labs HUNK system - AVIA ICD 2304 (Unchained Labs, Pleasanton, USA). The excitation wavelength was 285 nm, and emission intensities were recorded from 300 nm to 450 nm. The gain setting was set for 100, based on a previously performed gain test. From the incubation test, 162 min of additional incubation time was used. 48-point linear gradient of denaturant was automatically generated for each condition. For the first screening urea and guanidine hydrochloride (GuHCl) were used as denaturants, while for the second screening GuHCl was selected. 10 M urea and 6 M GuHCl stock solutions were prepared in each tested condition. Protein stock solutions were prepared at 1 g/L and were subsequently diluted 12.5 times to the final condition. Data collection and analysis were performed using Formulator software v3.02 (Unchained Labs, Pleasanton, USA). Protein intrinsic fluorescence was measured and the unfolding process was monitored by looking at the shift in the fluorescence spectra (356/318 nm) with increasing GuHCl concentration. In order to minimize the error, a secondary fit was performed for each pH value combining different NaCl concentrations. Free energy of unfolding (ΔG_{unfold}), $c_{1/2}$, and m -values were calculated for both transitions.

Microscale Thermophoresis

MicroScale Thermophoresis (MST) was performed using Monolith NT.115 Label Free system through the MO.Control software (NanoTemper Technologies, Germany). All measurements were carried out in 10 mM acetate pH 5.0 at 25°C and two different ligands were chosen: arginine and proline with stock concentration of 1 M each. Each standard capillary was manually loaded with 10 µl of protein at 1 µM with different ligand concentrations, covering the concentration range from 500 to 0.78 mM. All measurement were carried out at 20% excitation power. Data analysis was performed using the software MO. Affinity analysis. Initial fluorescence was used for data evaluation. For the arginine binding curve and K_d calculation eight independent experiments were performed. Proline binding affinity experiments were done in triplicates.

Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)

A Vanquish Horizon™ UPLC system with a variable wavelength UV detector was operated at 280 nm (Thermo Fischer Scientific, Waltham, USA). All experiments were performed at 4°C and temperature was controlled by autosampler. The separation was performed with a Superdex 200 increased 10/30 GL column. The aqueous mobile phase consisted of 38 mM NaH₂PO₄, 12 mM Na₂HPO₄, 150 mM NaCl, and 200 ppm NaN₃ at pH 7.4 dissolved in HPLC-grade water. The mobile phase was filtered through Durapore VVPP 0.1 µm membrane filters (Millipore Corporation, Billerica, MA, USA). All the samples were centrifuged and injected in duplicates at a volume of 25 µl. Immediately after exiting the column, samples passed through the UV detector followed by static light scattering apparatus, a TREOS MALS detector (Wyatt Technology, Santa Barbara, USA), and differential refractive index detector (Optilab T-rEX, Wyatt Technology, Santa Barbara, USA). Data collection and processing were performed using the ASTRA® software V7.2 (Wyatt Technology, Santa Barbara, USA).

Small Angle X-ray Scattering

Data collection was performed at the P12 beamline at the Petra III storage ring (DESY, Hamburg DE)¹⁰⁶ and at the BM29 beamline (ESRF, Grenoble FR)¹⁰⁷ (see Table S5.2 in Appendix B). Radius of gyration (R_g) and maximum dimension (D_{max}) were derived from the experimental data with the graphical data analysis program *PRIMUSQT*⁴³.

The rTrF crystal structures are available in three conformations in the protein data bank¹⁰⁸, *i.e.*, partially open (PDB ID: 3QYT⁹²), closed (PDB ID: 3V83⁹¹), and open (PDB ID: 2HAU⁷⁵) conformations.

Rigid body modelling of the dimer was performed using *SASREFMX*⁴³. In order to calculate the volume fractions of each component in the mixture, the data program *OLIGOMER*⁴³ was used. *FFMAKER*⁴³ was used to create an input file for *OLIGOMER* with a form factor for each component (open, partially open, and closed conformations retrieved from the protein data bank¹⁰⁸ and dimer from *SASREFMX* as input).

Molecular Dynamics simulation

The closed rTrF crystal structure was obtained from the protein data bank¹⁰⁸ (PDB ID: 3V83⁹¹). This conformation was used as a start structure for molecular dynamics (MD) simulations. The Fe^{3+} ion and bicarbonate (CO_3^{2-}) molecules were considered during the simulations. The structure was initially prepared at pH 5.0 and pH 6.5 using the H++ server (<http://biophysics.cs.vt.edu/H++>)¹⁰⁹ which accounts for the protonation state of the titratable residues. Full details of the setup of the MD simulations has been described previously¹¹⁰. The excipients, namely acetate, phosphate, arginine, histidine, and sodium chloride were included in the study. Structures were obtained from PubChem¹¹¹ and Zinc Database¹¹². These molecules were prepared at the desired pH using ligprep tool in Schrödinger 2016-3 suite (Schrödinger, LLC, New York, NY, USA)¹¹³. Parameter file for the excipients and bicarbonate were prepared using the antechamber¹¹⁴ module in Amber 16 at desired pH. Charges were estimated using the AM1-BCC¹¹⁵ charge method. Using the 12-6-4 LJ-type nonbonded^{116,117} model in the amber force field, parameters for Fe^{3+} were obtained. All-atom classical constant pH molecular dynamics simulations¹¹⁸ in explicit solvent were carried out with the Amber 16 program¹¹⁹ employing the amber force field ff99SB¹²⁰ for proteins. Titratable residues such as Asp, His, Lys, Tyr, surrounding the Fe^{3+} and the bicarbonate were titrated during the simulations. Ionic strength for each of the excipients was adjusted to 140 mM by additions of 124 solute molecules to the solvated system containing approximately 48000 water molecules. Finally, constant pH simulations were performed for 100 ns and coordinates were saved every 10 ps. Analyses were performed with CPPTRAJ¹²¹ in Amber 16, and VMD version 1.9.3¹²². Preferential interaction coefficient (PIC) for the specific simulated system was calculated using the method described previously¹¹⁰. Furthermore, an interaction score per $P(I_{score})$ was calculated to estimate binding capacity of co-solute to residues on protein surface as described. Centre of mass of the co-solute was used for the determination of PIC and $P(I_{score})$.

Acknowledgements

DanScatt for funding SAXS trip.

EMBL P12 DESY and EMBL B29 ESRF for providing beam time for performing the SAXS experiments.

Albumedix Ltd for kindly providing us with recombinant transferrin.

European Union's Horizon 2020 research and innovation program for funding (grant agreement nr 675074).

Simulations were performed at the high performance computing (HPC) services at DTU and in-house CPU/GPU cluster facilities at DTU Chemistry.

Chapter 6

Albumin-neprilysin fusion protein

Introduction

In protein-drug development, an optimization of protein biochemical characteristics is essential to achieve desirable biotechnological applications. This process is particularly challenging for novel biotherapeutics, as many different properties need to be optimized. One possible strategy is fusion technology, which can be used not only in development of novel protein drugs, but also in the optimization of already existing biotherapeutics.

The first fusion protein, etanercept, was marketed in 1998 under the name Enbrel®. It fuses the ligand-binding portion of the human tumor necrosis factor receptor (TNFR) and the Fc portion of human IgG₁. Etanercept is still used for the treatment of rheumatoid arthritis and chronic plaque psoriasis¹²³. Since then, multiple fusion proteins have been approved⁵¹, and many other are under clinical trials⁵².

Fusion proteins are created by joining two or more genes that originally coded for separate proteins⁵². Like most other protein-drugs, fusion proteins are produced by recombinant DNA technology (see Figure 6.1), resulting in a single polypeptide chain with functional properties of both proteins¹²⁴.

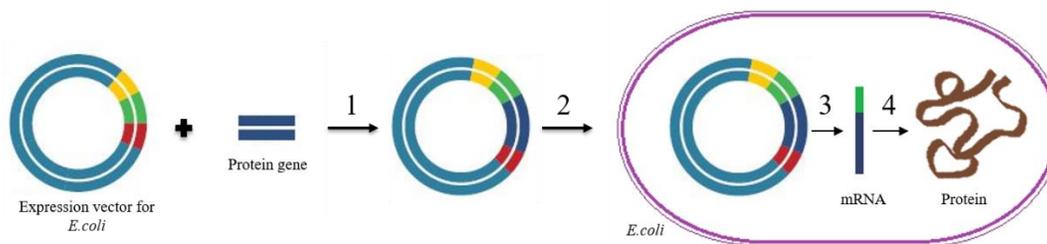


Figure 6.1: Recombinant DNA technology (inspired from Life, The Science of Biology¹²⁴). 1 - DNA that codifies for the protein of interest is inserted into the expression vector. 2 - *E. coli* cells are transformed with the vector. 3 and 4 – the protein is produced in large amounts¹²⁴.

Typically, proteins for fusion are selected based on their functionalities: one part is involved in molecular recognition, while other part has additional functions, like extended half-life, stability, cytotoxicity, and novel targeting or delivery routes¹²⁵.

Multiple advantages make fusion technology an attractive tool for protein-drug development. One fusion protein combines multiple functionalities, which might simplify manufacturing and drug discovery. Combination of two proteins in a single molecular entity results in identical distribution profiles. Moreover, it is possible to add new functionalities that are necessary for therapeutic purposes, that are lacking in natural proteins: like longer half-life or targeting specificity⁵².

Despite these advantages, fusion proteins also give rise to multiple challenges. Combination of two different proteins can result in increased difficulty in manufacturing due to noncompatible properties, which can lead to aggregation and misfolding of one of the domains. It is challenging to control relative amounts of each component to reach the dose with optimal efficacy and safety⁵². Furthermore, fusion proteins have a high potential for immunogenicity due to novel epitopes that are formed at the junction between two proteins⁵². Finally, fusion proteins can be challenging to formulate because of conflicting stability requirements⁵².

The current study addresses some of the challenges encountered in the formulation of these proteins. Understanding of stability on the molecular level is crucial for formulation of fusion proteins. Therefore, we decided to investigate the overall stability of an albumin-neprilysin fusion protein, HSA-NEP, at various physicochemical conditions and relate the results to the conformational changes providing a molecular understanding of the interplay between protein properties and stability.

ALBUMIN-NEPRILYSIN FUSION PROTEIN: UNDERSTANDING STABILITY USING SMALL-ANGLE X-RAY SCATTERING AND MOLECULAR DYNAMICS SIMULATIONS

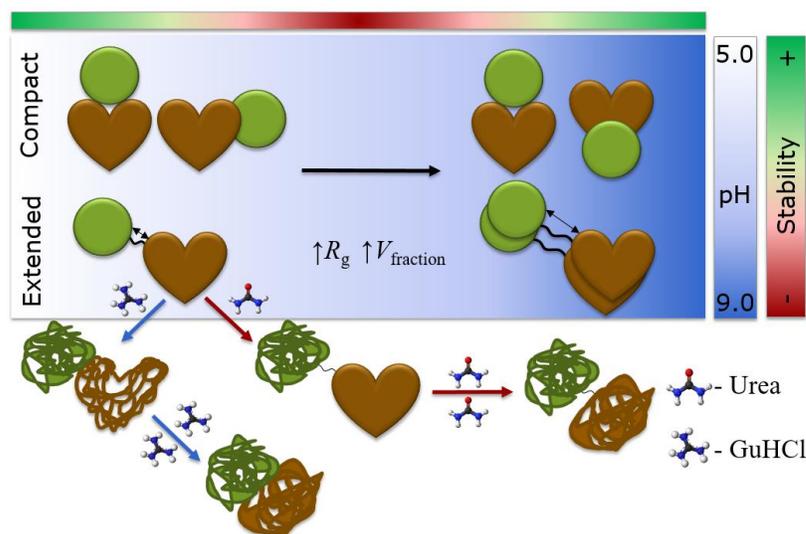
Alina Kulakova^a, Sowmya Indrakumar^a, Pernille Sønderby^a, Sujata Mahapatra^b, Werner Streicher^b, Günther H.J. Peters^a, Pernille Harris^a

^aDepartment of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

^bNovozymes A/S, Biologiens Vej 2, 2800 Kgs. Lyngby, Denmark

Abstract

Fusion technology is widely used in protein-drug development to increase activity, stability, and bioavailability of protein therapeutics. Fusion proteins, like any other type of biopharmaceuticals, need to remain stable during production and storage. Due to the high complexity and additional intramolecular interactions, it is not possible to predict the behavior of fusion proteins based on the behavior the individual proteins. Therefore, understanding the stability of fusion proteins on the molecular level is crucial for the development of biopharmaceuticals. The current study on the albumin-neprilysin (HSA-NEP) fusion protein uses a combination of thermal and chemical unfolding with small-angle X-ray scattering and molecular dynamics simulations to show a correlation between decreasing stability and increasing repulsive interactions, which is unusual for most biopharmaceuticals. It is also seen that HSA-NEP is not fully flexible: it is present in both compact and extended conformations. Additionally, the volume fraction of each conformation changes with pH. Finally, the presence of NaCl and arginine increases stability at pH 6.5, but decreases stability at pH 5.0.



Introduction

Stability and efficacy of protein-drugs are essential to achieve desirable biopharmaceutical applications. One of the strategies in optimization of protein-therapeutics is fusion technology, which consists of linking a target protein to a more stable protein. This approach has shown to improve catalytic efficiency, activity, stability, and solubility of protein-drugs¹²⁶. Moreover, the fusion-protein approach is used to prevent fast renal clearance by connecting a target protein to a protein with a longer half-life. Human serum albumin (HSA), which is used in this study, and immunoglobulin Gs are proteins widely used in fusion technology due to their long half-life.

HSA is a highly abundant and well-studied serum protein with a half-life around 19-22 days¹²⁷. Currently, multiple albumin fusion proteins are under clinical trials, and two are already accepted by the Food and Drug Administration (FDA)¹²⁸. One of them is albiglutide: an albumin fusion protein connected to a glucagon like peptide-1 receptor agonist¹²⁹. This therapeutic is administered for treatment of type 2 diabetes. The second albumin fusion drug is albutrepenonacog alfa, linked to the recombinant coagulation factor IX, which is used for treatment of hemophilia B¹³⁰. In both therapeutics, the presence of HSA contributes to a significant increase in half-life: from 1.5-5 min¹³¹ up to 3.6-8 days for albiglutide, and from 17-34 h up to 92 h for albutrepenonacog alfa^{130,52}.

As shown above, fusion technology can be used to address a variety of problems in protein-drug development. However, as any other type of biopharmaceutical, fusion proteins require special conditions (formulation) that will preserve their stability during production and storage. As it is not yet possible to predict the behavior of different proteins under different conditions, formulation remains a long and expensive process in protein-drug production. Formulation is particularly challenging for fusion proteins⁵², as additional intramolecular interactions lead to a change in stability, and therefore it is hard to predict the behavior of fusion proteins based on stability of individual proteins.

The aim of this study is to provide a better understanding on fusion protein stability and relate it to conformational changes. Specifically, by investigating the stability of albumin fused to neprilysin (HSA-NEP) (see Figure 6.2). NEP is widely distributed in mammalian tissues and is involved in the inactivation of a variety of signaling peptides^{132,133}. Additionally, it is involved in the degradation of amyloid β peptides, which makes it an attractive candidate as a protein-drug for treatment of Alzheimer's disease¹³⁴.

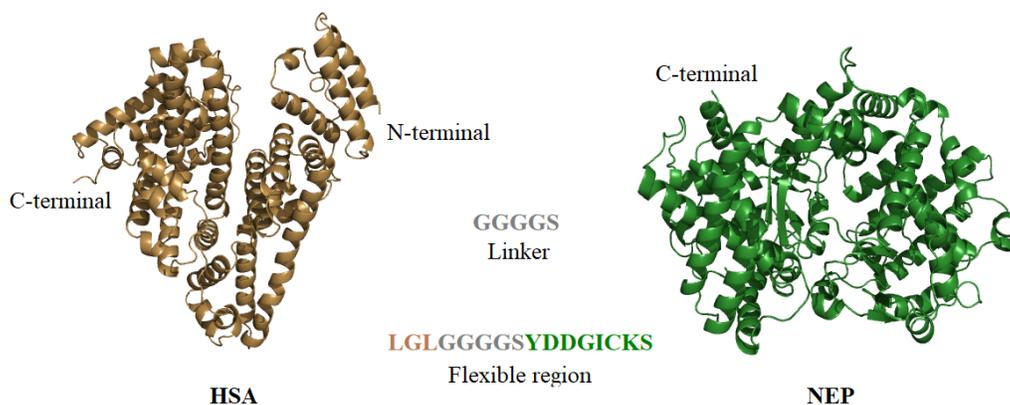


Figure 6.2: HSA N-terminus fused with the NEP C-terminus via GGGGS linker.

In this study, the overall stability of HSA-NEP was investigated by thermal and chemical denaturation by varying pH and buffer composition, complemented by small-angle X-ray scattering (SAXS) and molecular dynamics (MD) simulations. In combination, the stability studies and SAXS show a correlation between increasing protein repulsion and decreasing conformational stability. This contradicts conclusions from a previous study on a similar fusion protein: albumin fused with human growth hormone (HSA-*hGH*)¹³⁵, where repulsion is concluded to have a stabilizing effect at certain conditions. Additionally, combining SAXS results with MD simulation results provided a molecular understanding of the determinants that cause the HSA-NEP conformational changes.

Results

The overall stability of HSA-NEP was analyzed by thermal and chemical denaturation using nano differential scanning fluorimetry (nanoDSF) and isothermal chemical denaturation (ICD). Initially, ICD was performed using two different denaturants: urea and guanidine hydrochloride (GuHCl). GuHCl is a strong denaturant that starts to unfold HSA-NEP at low concentrations. Therefore, urea was used for subsequent studies, in order to obtain well-defined denaturation curves. The initial analysis was performed as a function of pH (5-9) and NaCl concentration (0, 70, and 140 mM).

pH dependence

The denaturation curves from thermal unfolding using nanoDSF are shown in Figure 6.3A. HSA-NEP has a single two-state unfolding (from folded to unfolded state) at pH 5.0, which is shifted towards a three-state unfolding with increasing pH (with the presence of the intermediate state). However, from pH 7.5 to pH 8.5 the thermal unfolding shifts back to a two-state (see Figure 6.3A). Chemical denaturation results in a multi-state unfolding (with two intermediate states), which is shifted towards a three-state unfolding with increasing pH (see Figure 6.3B). This means that the first intermediate state has a clear plateau at pH 5.0, less clear at pH 7.5, and not apparent at pH

8.5. The second intermediate state is not well-defined at pH 5.0, but more pronounced at pH 7.5 and 8.5. Hence, thermal and chemical denaturation of HSA-NEP show different unfolding mechanisms.

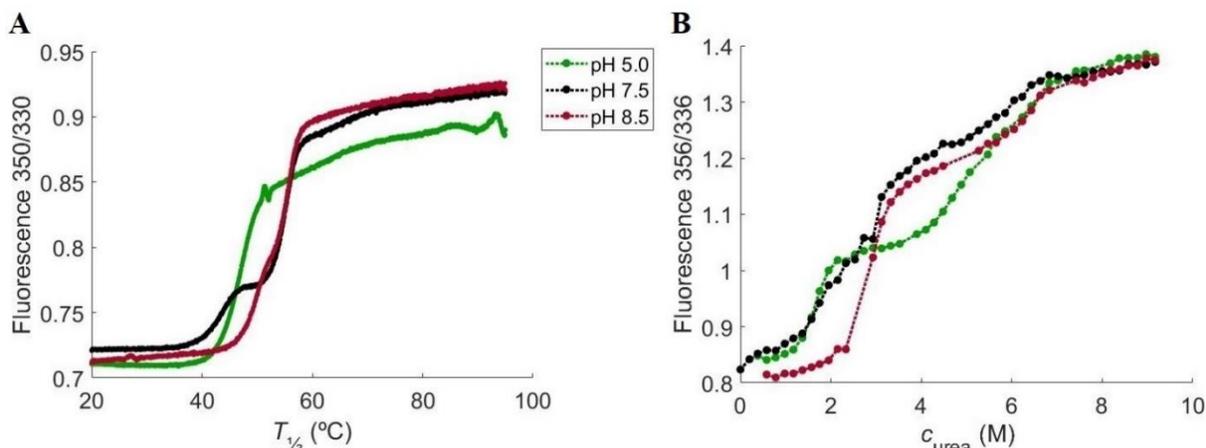


Figure 6.3: NanoDSF (A) and ICD (B) curves at 10 mM histidine pH 5.0 (green), 10 mM histidine pH 7.5 (black), and 10 mM tris pH 8.5 (red).

The temperature of unfolding ($T_{1/2}$) and the denaturant needed to unfold 50% of the protein ($c_{1/2}$) are shown in the Figure 6.4. At 0 mM NaCl, $T_{1/2}$ decreases from pH 5.5 to 7.5, with the higher values at lower (pH 5.0 and 5.5) and higher pH (from pH 8.0 to pH 9.0) (see Figure 6.4A). The $c_{1/2}$ increases from pH 5.0 to pH 6.0, where it reaches a plateau. It increases again around pH 8.0 (see Figure 6.4B).

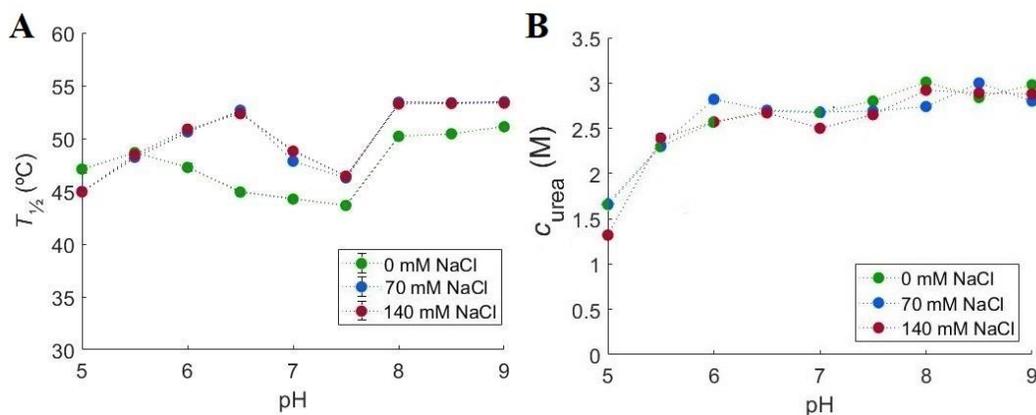


Figure 6.4: Initial stability studies performed using A: NanoDSF and B: ICD at different pH and ionic strengths.

In order to study the associated conformational changes, SAXS concentration series data were collected at 10 mM histidine at pH 5.0, 5.5, 6.5, and 7.5, and 10 mM tris at pH 8.5 with 0 mM NaCl. All scattering curves and SAXS data analysis can be seen in Appendix C (see Table S6.3 and Figure S6.1). The intensity at low q -values decreases with increasing HSA-NEP concentration at all pH values except pH 5.0, indicating a repulsive system (see Figure 6.5).

Moreover, repulsive interactions increase from pH 5.5 to 7.5 and decrease from pH 7.5 to 8.5, which correlates with the observed changes in $T_{1/2}$ (see Figure 6.4A).

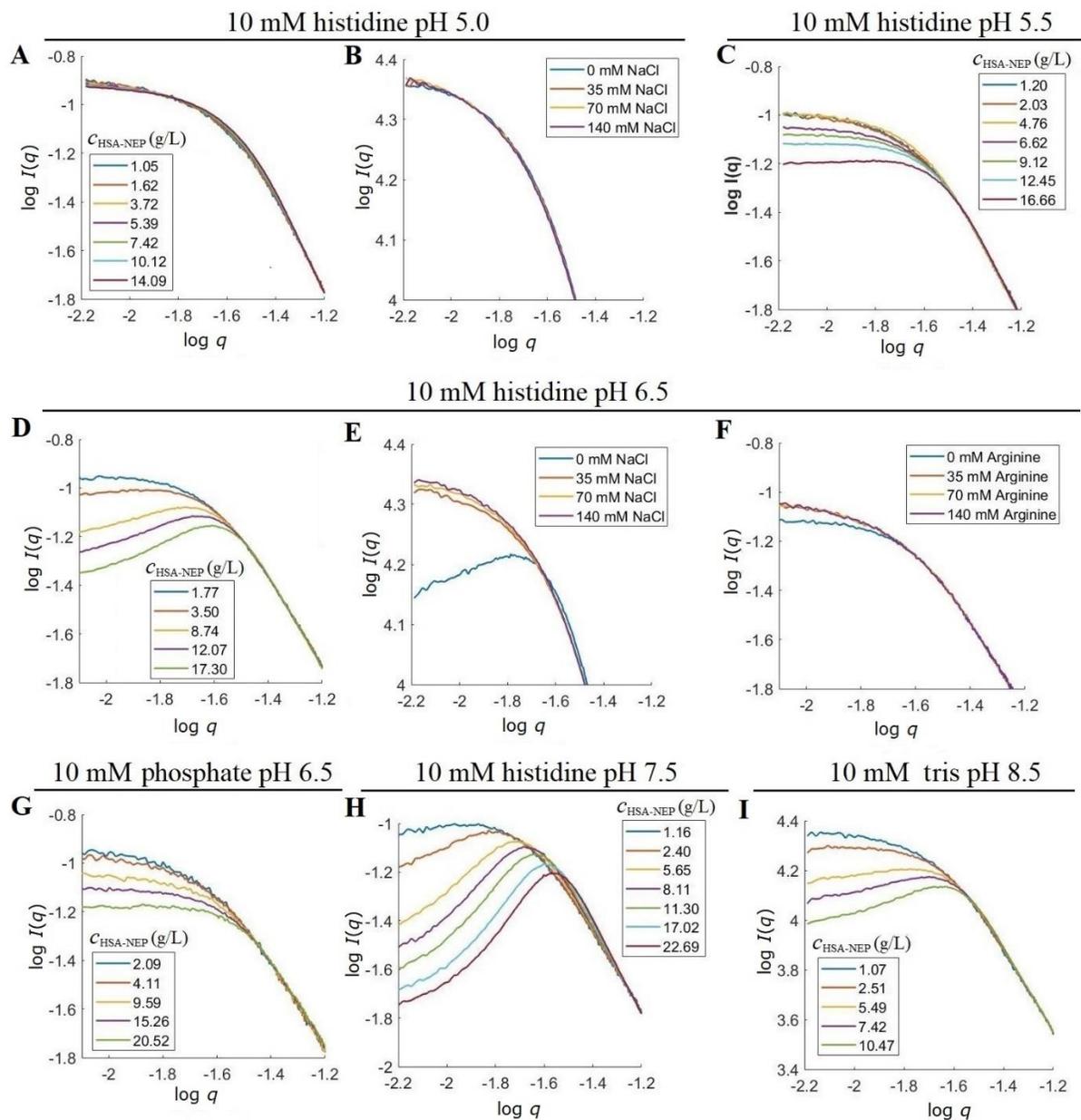


Figure 6.5: SAXS scattering curves for concentration series at A: 10 mM histidine pH 5.0; C: 10 mM histidine pH 5.5; D: 10 mM histidine pH 6.5; G: 10 mM phosphate pH 6.5; H: 10 mM histidine pH 7.5; I: 10 mM tris pH 8.5. SAXS scattering curves varying C_{NaCl} at B: 10 mM histidine pH 5.0 and E: 10 mM histidine pH 6.5, with $C_{\text{HSA-NEP}}$ around 5.5-6 g/L; and varying C_{Arginine} at F: 10 mM histidine pH 6.5 with $C_{\text{HSA-NEP}}$ around 2 g/L.

The Kratky plots of HSA-NEP show increase of the scattering at higher angles, which is characteristic for flexible systems³⁷ (see Figure S6.2 in Appendix C). Therefore, conformational changes of HSA-NEP were studied using *Ensemble Optimization Method (EOM)*^{45,136}, which accounts for flexibility. The *EOM* results were analyzed by looking at the distribution of the radius of gyration (R_g), the high-resolution models, and their volume distribution at different conditions. The R_g distributions shown in Figure 6.6A and B, indicate the presence of two overall populations: a more compact conformation with R_g around 4 nm and a more extended conformation with R_g around 5 nm. The R_g distribution around 4 nm shows multiple peaks, indicating the presence of multiple compact conformations. The detailed SAXS analysis of the high-resolution models shows the presence of three different compact conformations (see Figure 6.6E), whose distribution is pH dependent (see Figure 6.6C).

The volume fraction of the extended conformation increases with increasing pH. Also, R_g increases with increasing pH. The analysis of the high-resolution models shows many different possible conformations, which is an indication of high flexibility of the extended conformation. This is illustrated in Figure 6.6E, where HSA is kept fixed, and the position of NEP is seen to vary amongst the suggested structures.

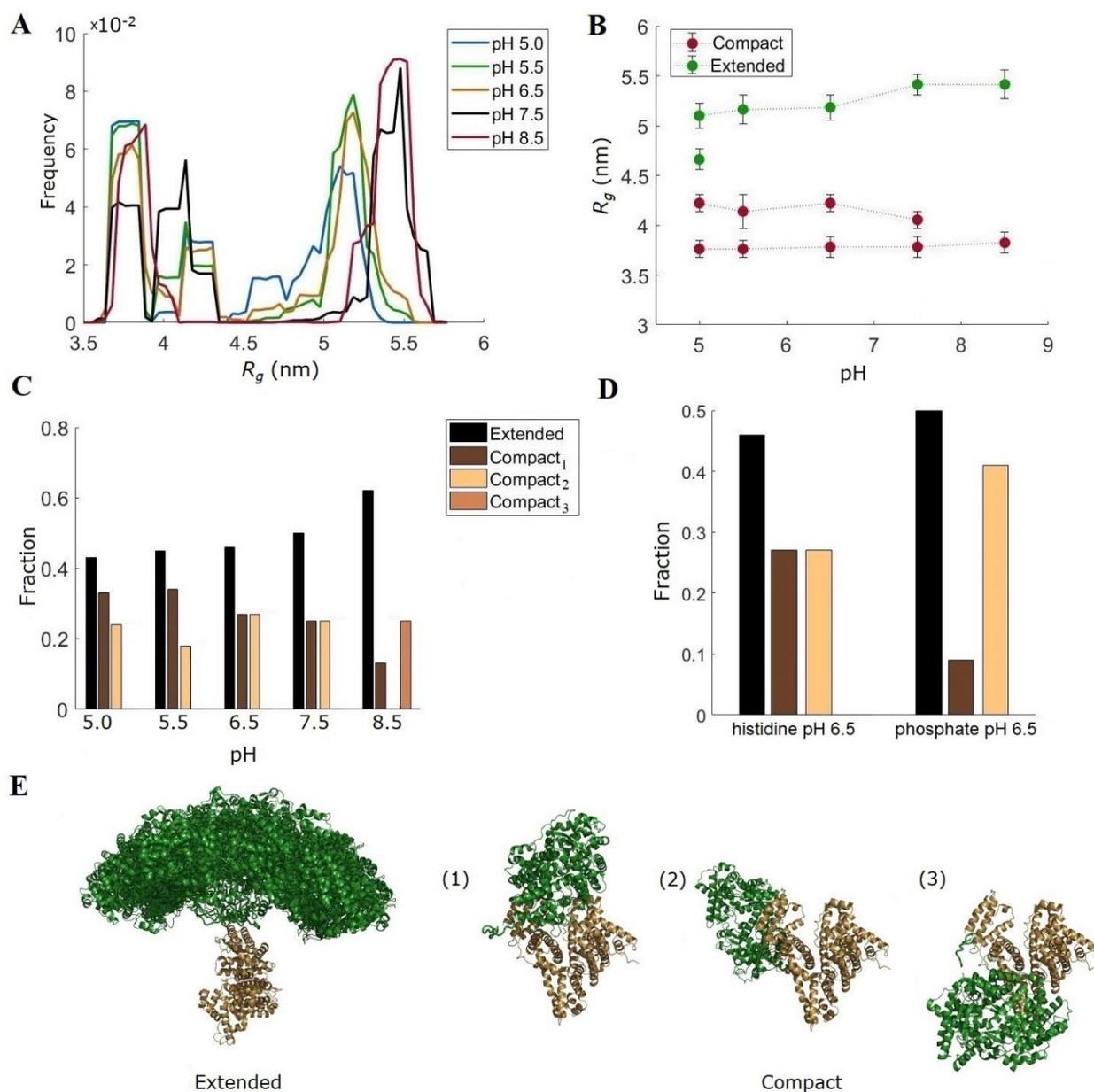


Figure 6.6: Analysis from *EOM*. A and B: R_g distributions, C: volume fractions of different conformations in 10 mM histidine at pH 5.0, 5.5, 6.5, and 7.5, and 10 mM tris pH 8.5, D: volume fractions of different conformations at pH 6.5 in histidine and phosphate buffers. E: high resolution models for extended and compact conformations.

NaCl dependence

With the addition of NaCl, the thermal denaturation studies show an increase in $T_{1/2}$ above pH 5.5 which points to an increase in thermal stability (see Figure 6.4A). The ICD studies do not show a clear trend for NaCl effect. A SAXS data was collected in the presence of NaCl at 10 mM histidine pH 5.0 and 6.5 (see Figure 6.5B and E). At pH 5.0, the scattering curves do not change with the presence of NaCl. However, at pH 6.5 the intensity at low q -values increases with addition of NaCl and calculated molecular weight (MW) shifts from 136 to 150 kDa (see Table S6.3 in Appendix

C), which is closer to the real MW of HSA-NEP. These results show that addition of NaCl screens the repulsive interactions present at pH 6.5, 0 mM NaCl.

Buffer and excipients dependence

Relative to the histidine buffer, both the acetate and phosphate buffers give rise to higher $T_{1/2}$ and $c_{1/2}$ (see Figure 6.7), which points to a higher conformational stability of the protein. Sucrose, arginine, and proline were selected as excipients and tested in the different buffers.

pH 5.0

At pH 5.0, both sucrose and proline have a weak positive effect on the HSA-NEP thermal stability, while the effect of arginine is negative and more pronounced. In acetate buffer, the presence of arginine is destabilizing, as it causes a decrease in $T_{1/2}$ and $c_{1/2}$ (see Figure 6.7A and C and Table 6.1). In histidine pH 5.0, arginine does not have a significant effect on the thermal stability, but the ICD studies show a decrease in $c_{1/2}$ by around 1 M in the presence of NaCl.

pH 6.5

The phosphate buffer at pH 6.5 was selected for SAXS measurements, as it has a clear positive effect in both ICD and nanoDSF experiments (see Figure 6.6B and D). Like in the histidine buffer, at low q -values the intensity decreases with increasing $c_{\text{HSA-NEP}}$ (see Figure 6.5D and G), which points to the presence of repulsive interactions. However, this decrease is less pronounced in phosphate buffer, which means that the system is less repulsive. At ~ 3 g/L, the calculated $MW_{\text{HSA-NEP}}$ is lower in phosphate buffer (166 kDa) than in histidine buffer (180 kDa) (see Table S6.3 in Appendix C), which points to a lower amount of larger species/aggregates in phosphate buffer. Moreover, the volume fractions of the compact conformations differs significantly: in histidine buffer both conformations, compact_1 and compact_2 , are present in equal amounts (0.27), while in phosphate buffer, the amount of compact_1 has decreased to 0.13 and the amount of compact_2 has increased to 0.41 (see Figure 6.6D).

At pH 6.5, the three excipients affect the system differently. Sucrose is slightly stabilizing in phosphate and histidine buffers, but in combination with NaCl it has a destabilizing effect, decreasing $T_{1/2}$ by 4.8°C (see Figure 6.7B and Table 6.1). Proline has some stabilizing effect in histidine buffer, but in combination with NaCl it decreases $T_{1/2}$ by 6.8°C, which points to a decrease in stability. In phosphate buffer, both proline and arginine have a different effect in thermal and chemical denaturation studies: they have negative effect on thermal stability, but seem to protect HSA-NEP from chemical denaturation (see Figure 6.7B and D and Table 6.1). In histidine buffer at pH 6.5 arginine increases $T_{1/2}$ by 5.4°C, which means an increase in HSA-NEP stability. Additionally, SAXS data show a decrease in the intermolecular repulsion (see Figure 6.5F). Addition of arginine in combination with NaCl has low effect.

NanoDSF

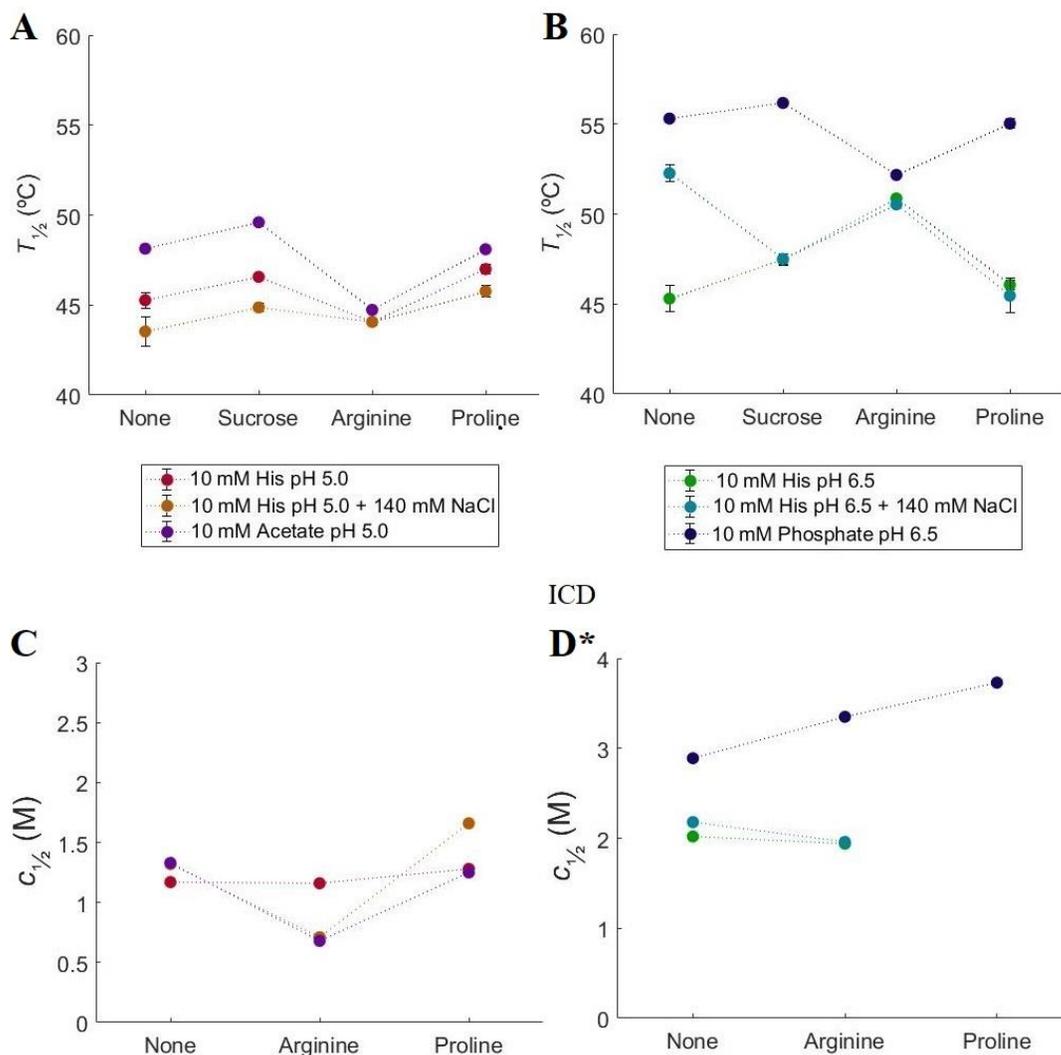


Figure 6.7: NanoDSF and ICD stability studies using different buffers and excipients. A: changes in $T_{1/2}$ at histidine (0 and 140 mM NaCl), and acetate pH 5.0; B: changes in $T_{1/2}$ at histidine (0 and 140 mM NaCl) and phosphate pH 6.5; C: changes in $c_{1/2}$ at histidine (0 and 140 mM NaCl) and acetate pH 5.0; D: changes in $c_{1/2}$ at histidine (0 and 140 mM NaCl) and phosphate pH 6.5. Purple: 10 mM acetate pH 5.0, red: 10 mM histidine pH 5.0, orange: 10 mM histidine pH 5.0 with 140 mM NaCl, blue: 10 mM phosphate pH 6.5, green: 10 mM histidine pH 6.5, cyan: 10 mM histidine pH 6.5 with 140 mM NaCl. *It was not possible to perform ICD studies at pH 5.0 with sucrose and pH 6.5 with sucrose and proline, due to the crystallization of solution in high concentrations of urea.

Table 6.1: Overview of the effect of excipients deduced from nanoDSF ($T_{1/2}$) and ICD ($c_{1/2}$) data analyses. Suc: sucrose, Arg: arginine-HCl, Pro: proline.

	10 mM acetate pH 5.0			10 mM histidine pH 5.0					
	Suc	Arg	Pro				+ 140 mM NaCl		
$T_{1/2}$	+	--	0	+	-	+	+	+	++
$c_{1/2}$	+	--	0	0	0	0	x	--	+
	10 mM phosphate pH 6.5			10 mM histidine pH 6.5					
	Suc	Arg	Pro				+ 140 mM NaCl		
$T_{1/2}$	+	--	-	++	+++	+	--	-	---
$c_{1/2}$	x	+	++	x	0	x	x	-	x
++++/----- $\Delta T_{1/2} > 10^\circ\text{C}$ $\Delta c_{1/2} > 1.5 \text{ M}$ +++/---- $5^\circ\text{C} < \Delta T_{1/2} < 10^\circ\text{C}$ $1 \text{ M} < \Delta c_{1/2} < 1.5 \text{ M}$ ++/-- $2^\circ\text{C} < \Delta T_{1/2} < 5^\circ\text{C}$ $0.5 \text{ M} < \Delta c_{1/2} < 1.0 \text{ M}$ +/- $0.5^\circ\text{C} < \Delta T_{1/2} < 2^\circ\text{C}$ $0.2 \text{ M} < \Delta c_{1/2} < 0.5 \text{ M}$ 0 $\Delta T_{1/2} < 0.5^\circ\text{C}$ $\Delta c_{1/2} < 0.2 \text{ M}$ Reference point for excipients: respective buffer without excipient. + stabilizes - destabilizes x – data not acquired									

Molecular dynamics simulations

In order to understand conformational changes of HSA-NEP under different physicochemical conditions, the interface between HSA and NEP in extended and compact conformations was studied using MD simulations. Subsequently, changes in the electrostatic surface potentials at the protein-protein interface were investigated to understand conformational preference with pH.

Extended

The inter-domain interface of the extended conformation is shown in Figure 6.8. The electrostatic surface at pH 5.0 shows less prominent positive and negative patches when compared to pH 8.5 (see Figure 6.8A and B). Furthermore, the free energy of interaction at the interface region is similar as they do not share a large interface in extended conformation. However, at pH 8.5, some residues in the linker region contribute more to the interaction energy (see Figure 6.8C and D). Due to the neutral patches around the interface, HSA and NEP tend to be closer in space at pH 5.0 (R_g of $45.97 \pm 0.22 \text{ \AA}$). Contrary to this, both proteins remain far from each other at pH 8.5 (R_g of $50.20 \pm 0.73 \text{ \AA}$), due to repulsive negative charge-charge interactions (see Figure 6.8D).

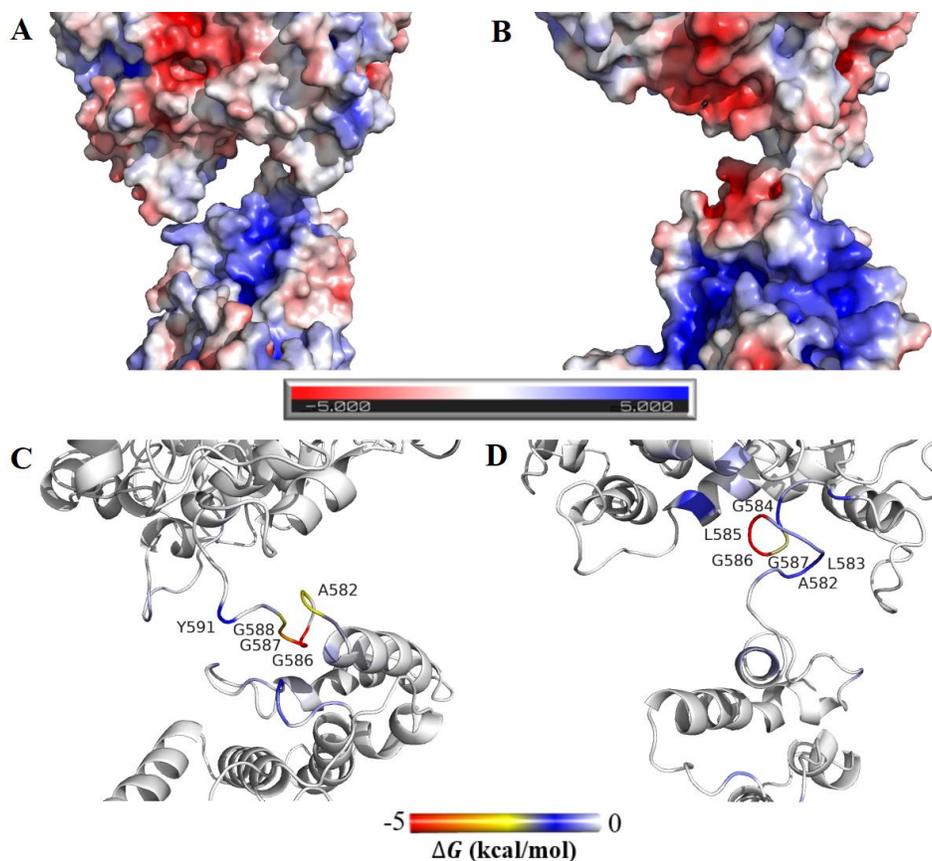


Figure 6.8: Surface coloring of the representative structure from MD trajectory clustering for extended conformation. Surface coloring based on the electrostatic potentials at A: pH 5.0, B: pH 8.5 (color-scale: red, white, and blue indicates negative, close to neutral, and positive potentials). Surface coloring based on the interaction energy at C: pH 5.0, and D: pH 8.5 (color-scale: red, blue, and white, indicates strong, medium, and weak interactions, respectively). In all the figures HSA is located in the bottom and NEP in the top.

Compact₁

The compact₁ inter-domain interface is shown in Figure 6.9. With increasing pH, residues at the interface contribute less to the free energy of interaction (see Figure 6.9C and D). Additionally, electrostatics around the interface changes with pH. At pH 5.0 the interface has an equal balance of positive and negative charges resulting in more compact conformation (R_g of 37.58 ± 0.23 Å), which contrasts the result at pH 8.5, where a strong positively charged patch causes positive charge-charge repulsion resulting in higher R_g of 39.67 ± 0.48 Å (see Figure 6.9A and B). Therefore, the balanced distribution of the positive and negative charges at pH 5.0 results in stronger interface interactions.

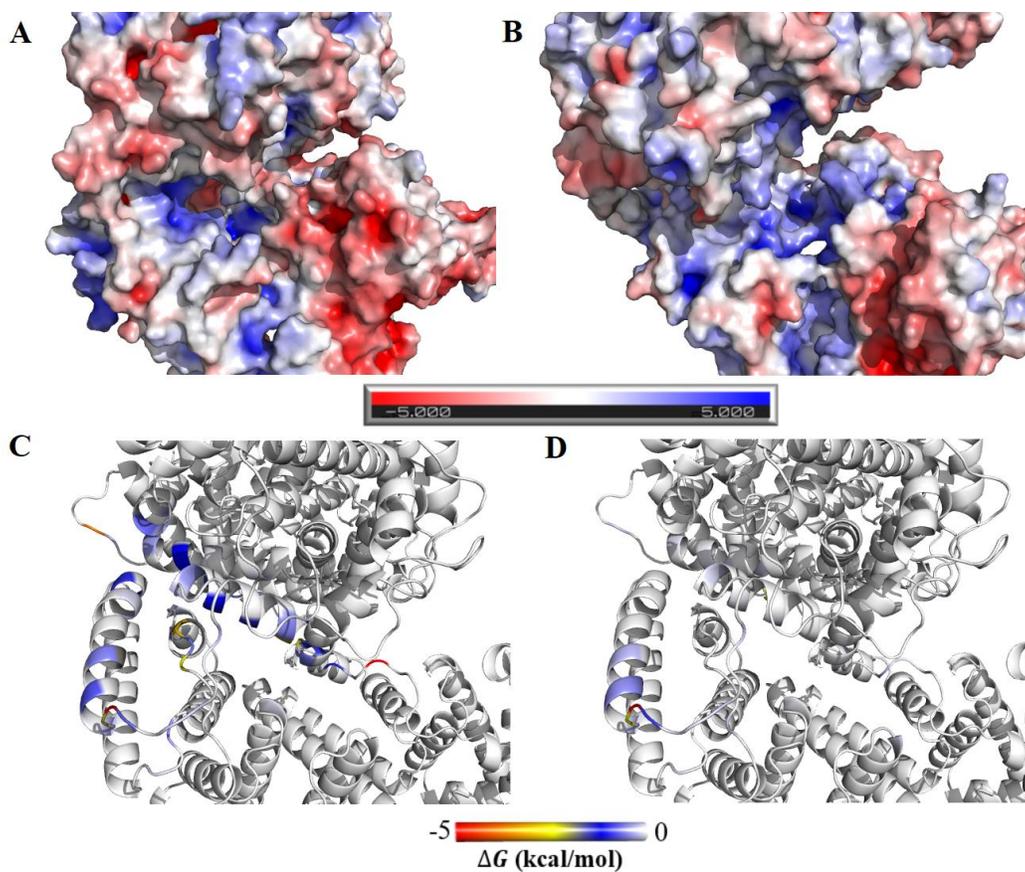


Figure 6.9: Surface coloring of the representative structure from MD trajectory clustering for the compact₁ conformation. Surface coloring based on the electrostatic potentials at A: pH 5.0, B: pH 8.5 (color-scale: red, white, and blue indicates negative, close to neutral, and positive potentials). Surface coloring based on the interaction energy at C: pH 5.0, and D: pH 8.5 (color-scale: red, blue, and white, indicates strong, medium, and weak interactions, respectively). In all the figures HSA is located in the bottom and NEP in the top.

Compact₂

The compact₂ inter-domain interface is shown in Figure 6.10. In the compact₂ conformation, the electrostatics change with pH around the interface region, which include residues E694, R691, D1044, E396, and E400 (see Figure 6.10A, B, and C). With increasing pH, the interface region becomes more hydrophilic, resulting in negative charge-charge repulsion. Moreover, at both pH 5.0 and 6.5, no significant differences in the interface electrostatics are seen. At pH 6.5 and 8.5, residues such as E1043, D553, and D549, have a higher contribution to the interaction energy, but no significant change in the interface interaction energy is seen (see Figure 6.10D, E, and F).

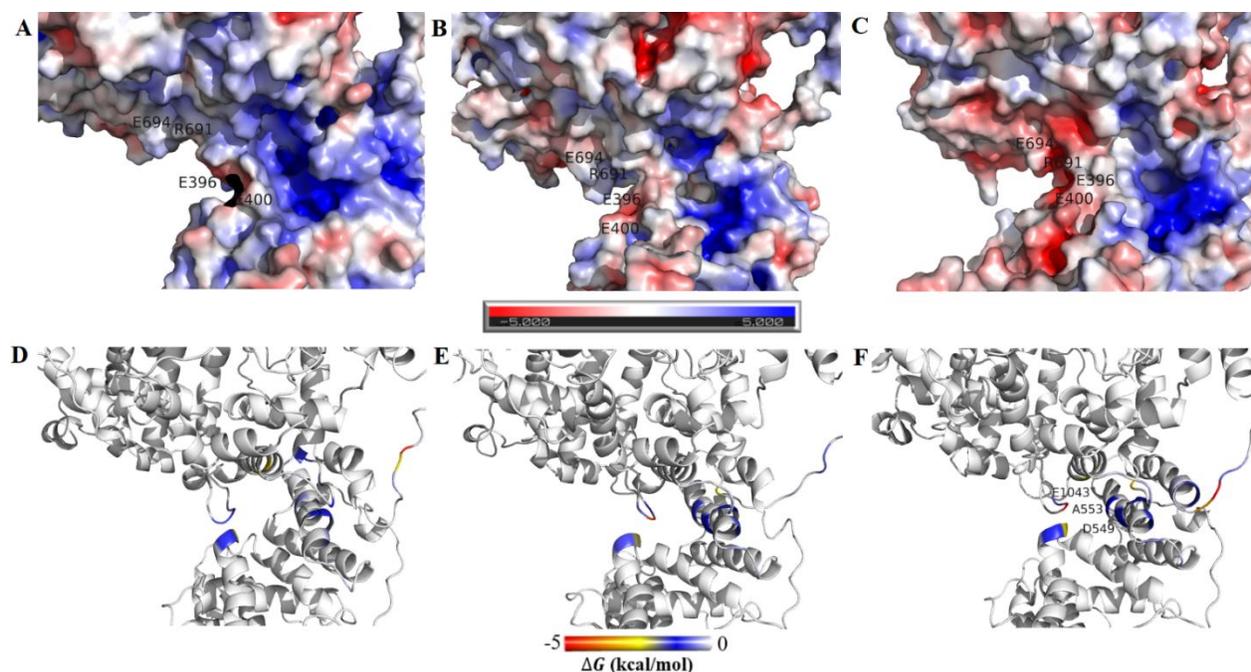


Figure 6.10: Surface coloring of the representative structure from MD trajectory clustering for compact₂ conformation. Surface coloring based on the electrostatic potentials at A: pH 5.0, B: pH 6.5, and C: pH 8.5 (color-scale: red, white, and blue indicates negative, close to neutral, and positive potentials). Surface coloring based on the interaction energy at D: pH 5.0, E: pH 6.5, and F: pH 8.5 (color-scale: red, blue, and white, indicates strong, medium, and weak interactions, respectively). In all the figures HSA is located in the bottom and NEP in the top.

Compact₃

The compact₃ inter-domain interface is shown in Figure 6.11. At pH 5.0, compact₃ conformation has a strong positive patch in the interface region, which makes HSA and NEP prone to repulsive behavior (see Figure 6.11A). With increasing pH, the interface becomes neutral, which makes compact₃ conformation more favorable at pH 8.5 (see Figure 6.11C). Additionally, interface analysis shows that the residues around the linker region (from 584 to 590) have higher contributions to the free energy of interaction at pH 8.5 (see Figure 6.11D), compared to pH 5.0 and 6.5, where this conformation is not seen. At pH 8.5, some of the interface residues, such as C492, G589, K746, and A606, have higher contributions to the overall interface energy in the presence of tris (see Figure 6.11D). Further, regions at the interface were identified which is prone to interacting more with tris (Figure 6.11F). In addition, the overall interface energy at pH 8.5 favors compact₃ in tris buffer (see Figure 6.11G).

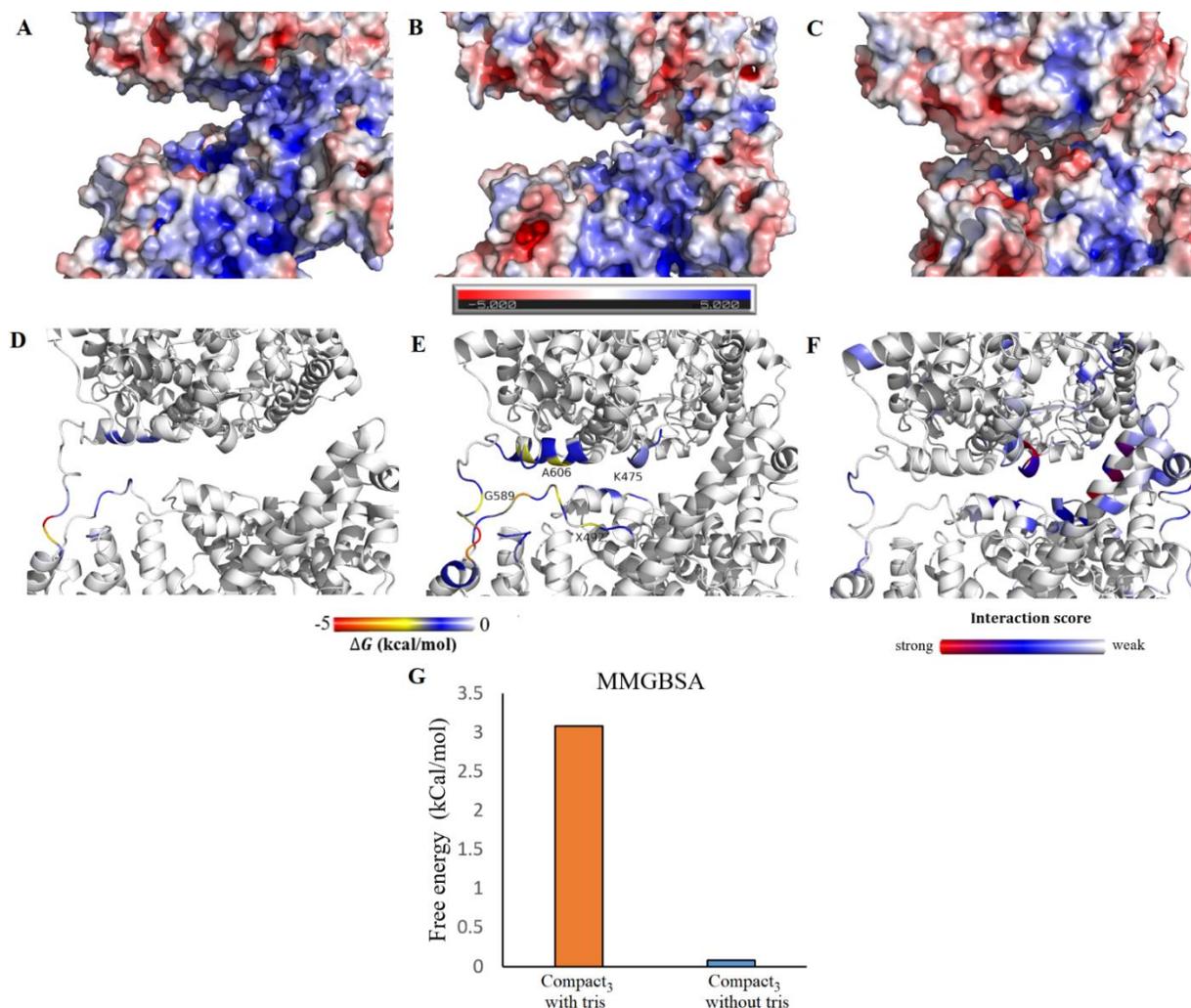


Figure 6.11: Surface coloring of the representative structure from MD trajectory clustering for compact₃ conformation. Surface coloring based on the electrostatic potentials at A: pH 5.0, B: pH 6.5, and C: pH 8.5 (color-scale: red, white, and blue indicates negative, close to neutral, and positive potentials). D: surface coloring based on the interaction energy at pH 8.5, and E: with tris at pH 8.5 (color-scale: red, blue, and white, indicates strong, medium, and weak interactions, respectively). F: Structure coloring based on tris interaction score (color-scale: white, blue, and red indicates no, weak, and strong interactions, respectively). G: MM-GBSA energy for compact₃ at pH 8.5 in the presence and absence of 10 mM tris. In all the figures HSA is located in the bottom and NEP in the top.

Denaturation process

Both urea and GuHCl were used as denaturants in the chemical denaturation studies, and both show a multi-state unfolding (see Figure 6.12A and B). However, they point to different unfolding mechanisms: in the presence of urea, the first two transitions are well separated with well-defined intermediate states, which is not the case with GuHCl. Unlike HSA-NEP, HSA alone has a simple two-state unfolding mechanism and requires higher concentrations of urea (~5M) and GuHCl (~2.3M) to unfold. It is seen in Figure 6.12B that HSA alone starts to unfold with addition of around 4 M urea, which corresponds to the beginning of the second transition of HSA-NEP. This suggests that in the presence of urea, the NEP domain unfolds first.

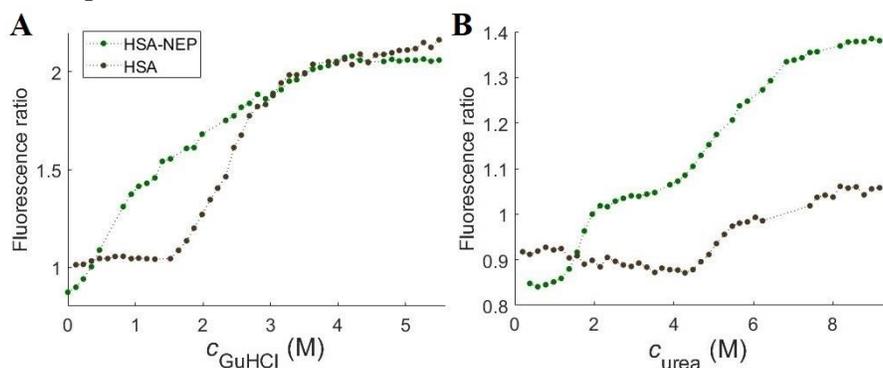


Figure 6.12: Chemical unfolding curves for HSA-NEP (green) and HSA (black) in histidine at pH 5.0 with A: GuHCl and B: urea.

In order to confirm the previous statement and follow conformational changes during unfolding, SAXS experiments were performed at different concentrations of urea in 10 mM histidine at pH 5.5, where the transitions are more well-defined (see Figure 6.13A and B). Without denaturant, HSA-NEP has a Kratky plot with a well-defined maximum and pair-distance distribution ($p(r)$) function with a double peak, which is characteristic for multidomain proteins. In the presence of 1 and 1.5 M of denaturant, the shape of the peak changes, pointing to small conformational changes. By increasing the concentration up to 3 M, the shape of the peak shifts to a single peak, which indicates significant conformational changes. The maximum intensity in the Kratky plot decreases with increasing concentration of urea, due to decreasing contrast. Finally, in the presence of 8 M urea, the Kratky plot has a plateau instead of a peak, meaning that HSA-NEP is fully unfolded.

The same experiment was performed with HSA alone (see Figure 6.13C). With addition of 1, 1.5, and 3 M urea, the shape of the peak in $p(r)$ function remains the same. By adding 5.5 M urea the peak becomes broader and the intensity increases at high q -values, which is characteristic for partially unfolded proteins. At the maximum concentration of urea, the Kratky plot of HSA (as well as HSA-NEP) has a plateau instead of a peak, which is characteristic of a fully unfolded protein.

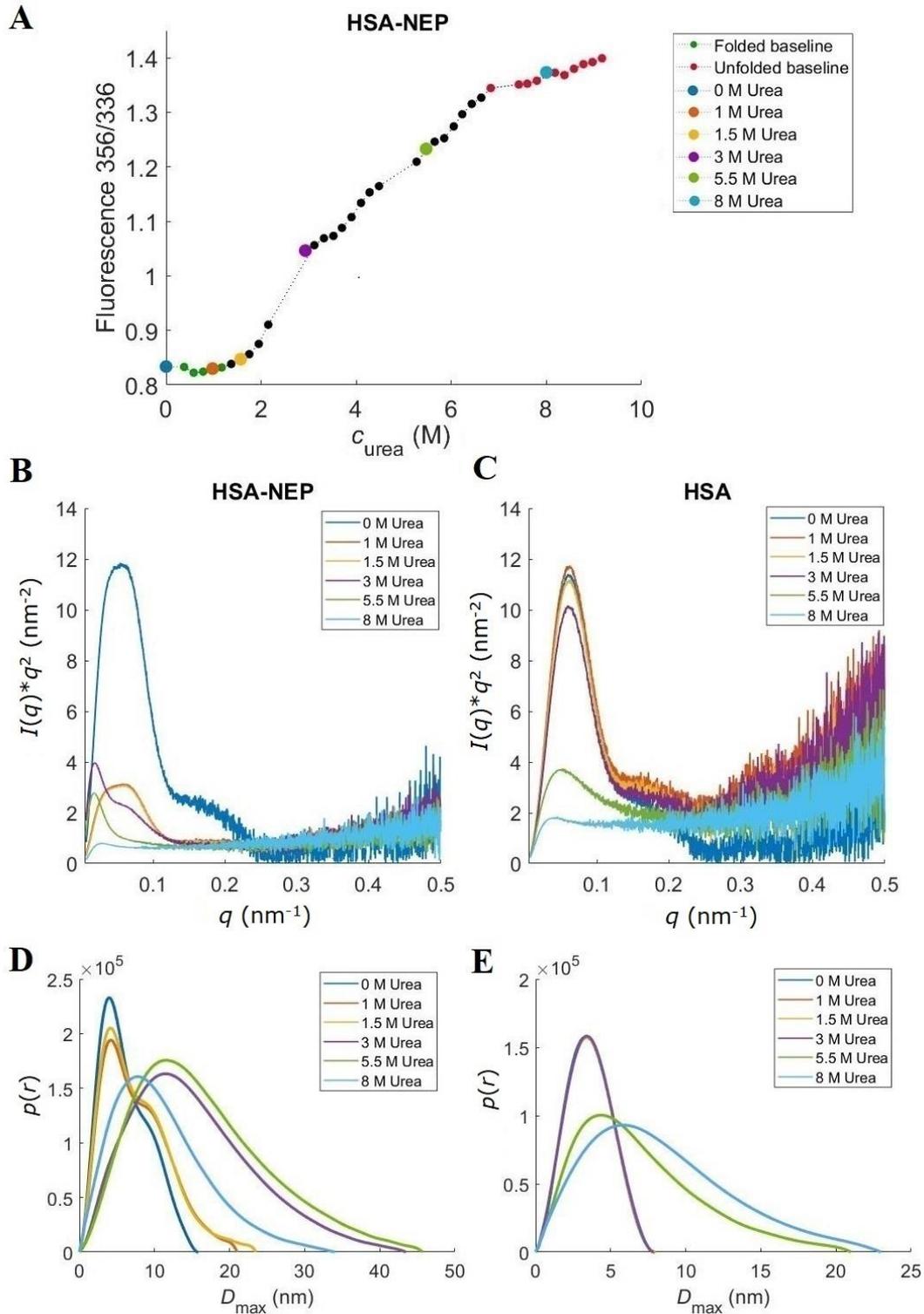


Figure 6.13: HSA-NEP chemical denaturation study. A: HSA-NEP (1 g/L) chemical unfolding curve. Kratky plots for B: HSA-NEP (5 g/L) at different concentrations of urea, and for C: HSA (5 g/L) at different concentrations of urea. $p(r)$ functions for D: HSA-NEP (5 g/L) at different concentrations of urea, and for E: HSA (5 g/L) at different concentrations of urea. Urea concentrations are given in the insets.

Discussion

From the nanoDSF and SAXS data analyses it is clear that there seem to be a correlation between $T_{1/2}$ and repulsive interactions in the system (see Figure 6.14A and B). A decrease in repulsion, seen as an increase in the structure factor, $S(0)$, in the SAXS data, is followed by a increase in $T_{1/2}$, which could be indicative of a destabilizing effect of repulsive interactions (see Figure 6.14A). The correlation seems to be dependent also on the choice of buffer, as seen in Figure 6.14B. This finding contradicts the findings of *Cordes et. al.*¹³⁵, where repulsive interactions in 10 mM acetate at pH 5.0 show a positive effect on the HSA-*hGH* fusion protein stability¹³⁵. Due to the similar pKa values for *hGH* and NEP (5.12 for *hGH*¹³⁷ and 5.47 for NEP), similar non-specific protein-protein interactions would be expected, however the specific interactions must be responsible for these differences.

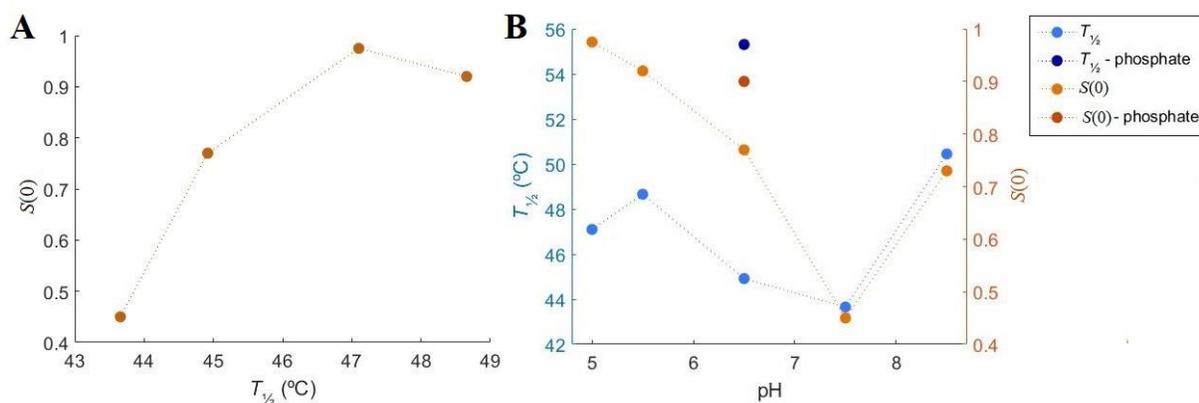


Figure 6.14: Correlation between thermal stability and repulsive interactions with pH. A: correlation between $S(0)$ and $T_{1/2}$, B: correlation between $S(0)$ and $T_{1/2}$ in histidine buffer (pH 5.0-7.5), phosphate buffer (pH 6.5), and tris buffer (pH 8.5). Thermal stability represented as $T_{1/2}$ (in blue); and repulsive interactions represented as $S(0)$ at $C_{\text{HSA-NEP}}$ around 10 g/L (in orange).

The increase in repulsion seen in histidine buffer from pH 5.5 to 7.5 can be explained by the pI values of the individual proteins. HSA has pI of 5.67 and NEP has pI of 5.47, which means that both proteins are negatively charged at pH > 6. Therefore, increasing pH leads to an increase in repulsive interactions, which is also in agreement with MD simulations (see Figure 6.8). The increase in repulsion between the artificially connected HSA and NEP proteins could induce internal stress, which, according to the SAXS data, is followed by an increased volume fraction of the extended conformation. One possible scenario is that in order to minimize the internal repulsive interactions in the protein, both domains keep as far away from each other as possible, which leads to an increase in R_g (see Figure 6.6A and B). This results in increased flexibility and that HSA-NEP is more easily unfolded. Additionally, from two possible conformations, HSA-NEP will prefer the extended conformation, where HSA and NEP are more separated in space (see Figure 6.6C).

Furthermore, it is seen that at pH 6.5, where the protein clearly shows repulsive interactions and low thermal stability, the addition of NaCl to HSA-NEP increases $T_{1/2}$ (see Figure 6.4A). At the same time, SAXS data indicates that NaCl screens the repulsive interactions (see Figure 6.5E).

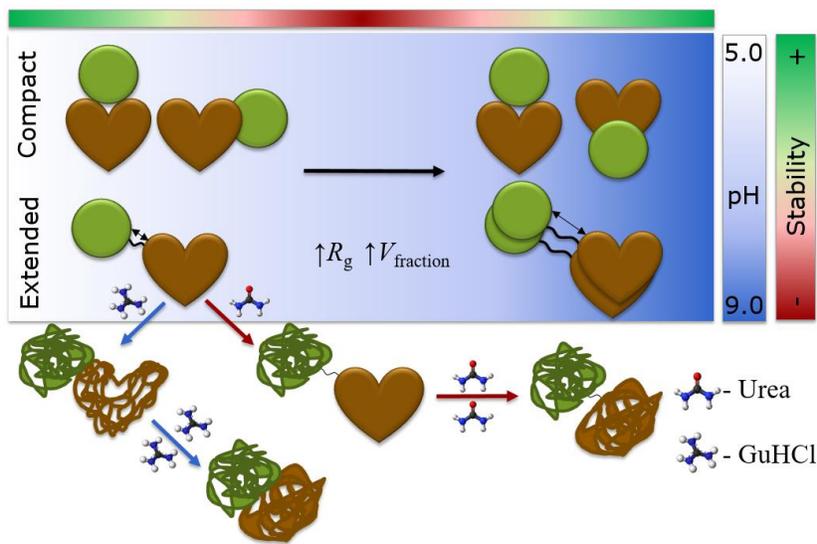


Figure 6.15: Schematic overview of HSA-NEP stability.

NanoDSF and ICD studies show that not only pH but also the buffer type influences HSA-NEP stability. Changing histidine buffer for phosphate buffer at pH 6.5 increases both $T_{1/2}$ and $c_{1/2}$ (Figure 6.7B and D), decreases repulsion (see Figure 6.5E and G), and changes the volume fractions of compact₁ and compact₂ (see Figure 6.6D). One of the reasons might be the significantly higher ionic strength in 10 mM phosphate buffer at pH 6.5 (0.013 M vs 0.005 M for 10 mM histidine buffer), which contributes to the screening of repulsive interactions. The changes in volume fractions for the compact conformations might happen due to specific interactions between the phosphate buffer and HSA-NEP. At pH 5.0, HSA and NEP do not have significant intramolecular interactions in histidine buffer (see Figure 6.5A). However, replacing it by acetate buffer leads to an increase in $T_{1/2}$ and $c_{1/2}$, which also might happen due to specific interaction between buffer and HSA-NEP.

Among all tested additives, arginine has a pronounced effect on the HSA-NEP stability. At pH 5.0, it generally destabilizes, for both chemical and thermal denaturation, with an exception in the thermal denaturation in histidine with 140 mM NaCl. In general, the effect of arginine seems to be comparable to NaCl effect for thermal denaturation. However, the chemical denaturation curves indicate that arginine may have another role in this system: helping unfolding at pH 5.0 and hampering unfolding at pH 6.5 in phosphate. Proline seems to have a clear stabilizing effect at pH 6.5 in phosphate buffer (chemical denaturation), and a clear destabilizing effect in histidine buffer +140 mM NaCl (thermal denaturation).

As already mentioned, HSA-NEP is present in different compact conformations: compact₁, compact₂, and compact₃ (see Figure 6.6E). For better understanding of conformational changes, the interface between HSA and NEP was studied by MD simulations. Molecular understanding

about the preference of the different conformation in varying pH was explored combining the electrostatics surface study and free energy of interaction at the HSA-NEP interface.

The SAXS data shows that up to pH 7.5 only the compact₁ and compact₂ conformations are present. According to the analysis of the MD simulations, compact₃ is not present below pH 7.5 due to highly unfavorable repulsive interactions in the interface region (see Figure 6.11C). Moreover, *EOM* analysis shows that up to pH 7.5, the volume fraction of compact₁ decreases, while the volume fractions for extended and compact₂ conformations increase (see Figure 6.6C). Analysis of the MD simulations shows that compact₁ becomes less favorable due to increasing positive charge-charge repulsion and decreasing interaction energy between HSA and NEP. A slight increase in the volume fraction of compact₂ can be explained by difference in interaction energy at the interface around residues E143, D553, and D549, which is higher for pH 6.5, and 8.5, and lower for pH 5.0.

Finally, SAXS data shows that at pH 8.5 HSA-NEP is present in compact₁, compact₃, and extended conformations. According to MD simulation results, the interface in compact₃ becomes neutral, and the interaction energy increases due to the contribution from the residues in the linker region. Additionally, using tris as a buffering system enhances the stability of compact₃, making it more a favorable conformation along with the extended conformations (see Figure 6.6C). From MD simulations at pH 8.5, compact₁ has less pronounced charge-charge repulsion when compared to compact₂, which means that compact₁ is more likely to be present at pH 8.5 than compact₂.

HSA-NEP unfolding

In combination, ICD and SAXS provide better understanding of the HSA-NEP unfolding process. HSA-NEP unfolding starts with NEP unfolding, which is followed by unfolding of HSA. Additionally, the GuHCl and urea denaturation curves point to different unfolding mechanisms (see Figure 6.15). In the presence of urea, the first two transitions are well separated by well-defined intermediate state, which suggests presence of a stable intermediate, where HSA is folded and NEP is unfolded. In the presence of GuHCl, HSA-NEP also has multi-state unfolding, but the transitions between different states are not well-defined. This suggests that with GuHCl, unfolding of NEP might affect the integrity of HSA, which leads to the absence of a well-defined intermediate state.

Moreover, multiple transitions in ICD curves (see Figure 6.12) can also be explained by the presence of different HSA-NEP conformations. Compact conformations have multiple interactions at the interface, which might stabilize NEP. However, in the extended conformation, the NEP domain is more exposed, resulting in a faster unfolding.

Conclusion

From this study, we conclude that it is difficult to predict behavior and stability of fusion proteins in general based on the knowledge of individual proteins. While the HSA-NEP protein shows increasing stability with decreasing repulsion, HSA alone shows the opposite¹³⁸. Furthermore, it is also not even possible to generalize the stability of albumin fusion proteins, as the MD

simulation results show the importance of the specific interactions indicating that for each system a separate analysis is needed.

Unlike most proteins, HSA-NEP is less stable with increasing repulsion, due to intramolecular repulsion. Moreover, interactions between HSA and NEP do not allow for full flexibility: HSA-NEP is present in an extended and multiple compact conformations. The compact conformations are mainly stabilized by salt bridges in the interface between HSA and NEP and are therefore more rigid, while the extended conformation is more flexible. Changes in pH induce changes at the interface, which shifts the equilibrium between different compact and extended conformations.

Methods

Dialysis and formulation

HSA-NEP was provided by AstraZeneca in 49 mg/ml solution and was dialyzed into 10 mM histidine pH 5.5, 7.0, and 10 mM tris pH 8.5 for pH and NaCl screening; and in 10 mM histidine pH 5.0 and 6.5, 10 mM acetate pH 5.0, and 10 mM phosphate pH 6.5 for buffer and excipients screening. After dialysis, HSA-NEP was diluted to 20 mg/ml with dialysis buffer, followed by 1:20 dilution with final formulation buffer within pH of ± 0.5 (see Figure 6.16). Concentration of HSA-NEP was measured on a Nanodrop™ 1000 (Thermo Fisher Scientific, Waltham, USA) using extinction coefficient provided by AstraZeneca (see Table S6.2).

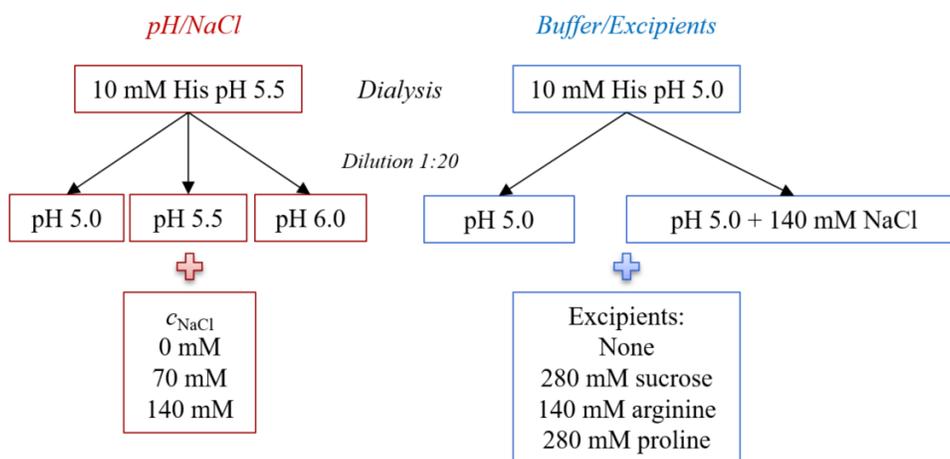


Figure 6.16: Schematic representation of dialysis and formulation process.

Isothermal chemical denaturation

All chemical denaturation studies were performed on Unchained Labs HUNK system - AVIA ICD 2304 (Unchained Labs, Pleasanton, USA). The excitation wavelength was 285 nm and emission intensities were recorded from 300 nm to 450 nm. The gain setting was set for 10, based on previously performed gain test. From the incubation test, 1134 min of additional incubation time were set. 48-point linear gradient was automatically generated for each condition. For the first screening urea and guanidine hydrochloride (GuHCl) were used as denaturants, while for the

second screening urea was selected. 10 M urea and 6 M GuHCl stock solutions were prepared in each tested condition. Protein stock solutions were prepared at 1 mg/ml and were posteriorly diluted 12.5 times to the final condition. Data analysis was performed using Formulator software v3.02 (Unchained Labs, Pleasanton, USA). For the native protein, the emission wavelength was selected from the data of the first point of the gradient, which corresponded to 336 nm. For the unfolded state, data of the last point of the gradient was considered and the maximum wavelength was 356 nm. The fluorescence ratio 356/336 was plotted against denaturant concentration to monitor the unfolding process. Secondary fits were performed for each pH combining different NaCl concentrations in order to minimize the error. Free energy of unfolding (ΔG_{unfold}), $c_{1/2}$, and m -values were calculated for both transitions.

Thermal denaturation

Thermal stability studies were performed with the Prometheus NT.48 (NanoTemper Technologies, Munich, Germany). NanoDSF Grade Standard Capillaries were manually loaded with 10 μl of protein at 1 mg/ml in the final conditions. All experiments were performed from 20 to 95°C with the linear thermal ramp with the heating rate of 1°C/min. Protein intrinsic fluorescence was measured and unfolding process was monitored by looking at the shift in the fluorescence spectra (350/330 nm). All measurements were done in triplicates and the data analysis was performed using PR.Control v1.12.2 software (NanoTemper Technologies, Munich, Germany).

Small Angle X-ray Scattering

Data collection was performed at the P12 beamline at the Petra III storage ring (DESY, Hamburg DE)¹⁰⁶ (see Table S6.2 in Appendix C for experimental details). Radius of gyration (R_g) and maximum dimension (D_{max}) were derived from the experimental data with the graphical data analysis program *PRIMUSQT*⁴³.

Ensemble Optimization Method (EOM) was used to analyze conformational polydispersity of HSA-NEP. *EOM* consists of two programs *RANCH* and *GAJOE* that were used separately. *RANCH* was used to generate a large pool of 10000 random conformations (genes) by using HSA (pdbid: 6EZQ¹³⁹) and NEP (pdbid: 6GID¹⁴⁰) from homology modelling. Flexible regions in C-terminal of HSA (LGLG) and N-terminal of NEP (YDDGICKS) were removed from the structures, increasing the length of the linker. *GAJOE* was used to select ensembles of conformations, such that the average structure fits to the experimental data. Experimental curves acquired at different HSA-NEP concentrations were merged in order to remove the noise. Additionally, first 80 points were removed due to the presence of repulsive interactions. The output files contain fit to experimental data (see Figure S6.3 in Appendix C), distribution of volume fractions, and information about R_g and D_{max} distribution (see Table S6.4 in Appendix C).

Molecular dynamics simulations

HSA (PDB ID: 6EZQ¹³⁹) and NEP (PDB ID: 6GID¹⁴⁰) crystal structures were used as template structures for homology modelling of SAXS based fusion protein models using Modeller9.20¹⁴¹ program. Minimization was performed on the generated models to account for structure optimization. Using PDB2PQR¹⁴² plugin in pymol¹⁴³, structures were prepared at different physicochemical conditions (pH 5, 6.5, and 8.5). Subsequently, these structures were taken for all-atom classical constant pH MD simulation of 50 ns in explicit solvent utilizing ff99SB¹²⁰ force-field for proteins. In total, the solvated system contained approximately 60000 water molecules. Each system was neutralized with either sodium or chloride depending on the overall charge of the protein. The complete protocol used to setup MD simulations is described in our previous work¹¹⁰. HSA-NEP interaction interface is defined as follows: consider HSA, residues belonging to NEP within 5 Å of HSA's protein surface, and vice versa is defined as interface residues for HSA-NEP interaction. The interface for each of the conformations is different, hence the interface was found separately for each conformation. The titratable residues such as Asp, Glu, His, and Lys in the interface were subjected to titration during constant pH MD simulations to account for change in protonation state upon protein structure dynamics. Using the MM-GBSA^{144,145} free energy method, the free energy of interaction at the interface was calculated. The last 10 ns of production run where systems had converged RMSD was taken for MM-GBSA calculations. Analyses were performed with CPPTRAJ¹²¹ in Amber 16, and VMD 1.9.3¹²². The electrostatic surface potential was calculated for the most representative structure found throughout the simulation using hierarchical clustering approach^{146,147}. Additionally, interface interaction energy based surface coloring scale was chosen to highlight residues that strongly contribute to the interface interaction free energy.

The compact₂ conformation was simulated in 10 mM tris to understand the effect of tris on the conformational stability. In total, 11 tris molecules were added to the solvated system containing approximately 60000 water molecules. Tris was obtained from Zinc Database¹¹². These molecules were prepared at pH 8.5 using the Ligprep tool in Schrödinger release 2016-3 (Schrödinger, LLC, New York, NY, USA)¹¹³. Parameter file for the tris molecule was prepared using the antechamber¹¹⁴ module in Amber 16 at pH 8.5 and applying the AM1-BCC¹¹⁵ charge method. Furthermore, an interaction score per residue ($P(I_{score})$)¹¹⁰ was calculated to estimate the binding capacity of tris to the protein surface. Centre of mass of tris was used for the determination of $P(I_{score})$.

Acknowledgements

We thank the European Union's Horizon 2020 research and innovation program for funding (grant agreement nr 675074).

Danish Agency for Science, Technology, and Innovation for funding the instrument center DanScatt.

EMBL P12 DESY for providing beam time for performing the SAXS experiments.

AstraZeneca for kindly providing us with HSA-NEP fusion protein.

Simulations were performed at the high performance computing (HPC) services at DTU and in-house CPU/GPU cluster facilities at DTU Chemistry.

Chapter 7

Monoclonal antibodies

Introduction

Immunoglobulin G (IgG) monoclonal antibodies (mAbs) are one of the largest classes of biopharmaceuticals. Due to their high binding specificity, mAbs are widely used in diagnostic tests and in protein-drug development: targeted transport of drugs, toxins, or radioactive compounds to tumors as a cancer therapy¹⁴⁸. Due to their high stability, mAbs are also used in fusion protein technology (see chapter 6), by attaching an fragment crystallizable region to less stable proteins⁵².

Immunoglobulins are Y shaped molecules, which is composed by four polypeptide chains: two light chains and two heavy chains that are connected by disulfide bonds (see Figure 7.1). These are interacting so that the IgGs can be divided in three domains: two fragments antigen-binding (Fab) and one fragment crystallizable region (Fc). The structure of immunoglobulin can also be divided into variable (V) and constant (C) regions. Constant regions (C_L and C_H) are very similar within the antibody class, while variable regions (V_L and V_H) have different aminoacid sequences, which determines the diversity of antibody specificity¹⁴⁹. Additionally, mAbs can have two different types of light chains, κ and λ , which differ in aminoacid composition in the C_L ¹⁵⁰.

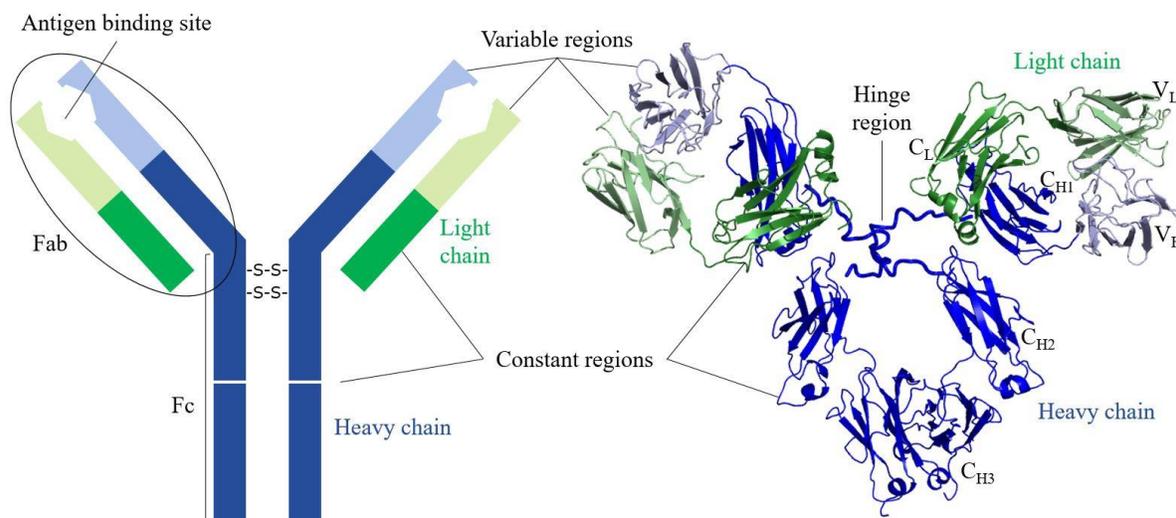


Figure 7.1: The structure of IgG. Light chains (V_L and C_L) in green; heavy chains (V_H , C_{H1} , C_{H2} , and C_{H3}) in blue. Constant regions are defined as C and variable regions are defined as V. Additionally, mAb can be divided in two regions: fragment antigen-binding (Fab) and fragment crystallizable region (Fc).

IgGs are divided into four different subclasses: IgG₁, IgG₂, IgG₃, and IgG₄, which mainly differ in the number of disulfide (S-S) bonds in the hinge region, and in relative abundance (see Table 7.1). Despite the fact that their sequence is more than 90% identical, each IgG subclass has a unique profile with respect to its function (*i.e.* antigen binding, half-life, placental transport, immune complex formation, and others)¹⁵¹.

Table 7.1: Characteristics of different IgG subclasses¹⁵¹.

	IgG ₁	IgG ₂	IgG ₃	IgG ₄
MW (kDa)	146	146	170	146
Amino acids in hinge region	15	12	62	12
S-S bonds inter-heavy chain	2	4	11	2
Relative abundance	60	32	4	4

Additionally, mAbs can be obtained from different sources and are classified as: murine, chimeric, mixed, humanized, and human (see Figure 7.2)⁶³. Muronomab was the first mAb approved by FDA in 1985 under the trade name Orthoclone OKT3®⁶². Muronomab is a fully murine antibody, which was administrated until 2011⁷² against renal, hepatic, cardiac, and combined kidney-pancreas transplant rejection⁶¹. Due to the high immunogenicity of murine mAbs, some parts of the sequence were replaced by human sequences resulting in new types of the antibodies: chimeric and humanized. In chimeric antibodies both the C_L and C_H domains are fully human, while V_L and V_H remain murine. Humanized antibodies have 95% human sequence and only some of the regions in variable domains remain murine¹⁵². The first fully human antibody, panitumumab (trade name Vectibix®), was produced in 2007 and is administrated against solid tumors¹⁵³.

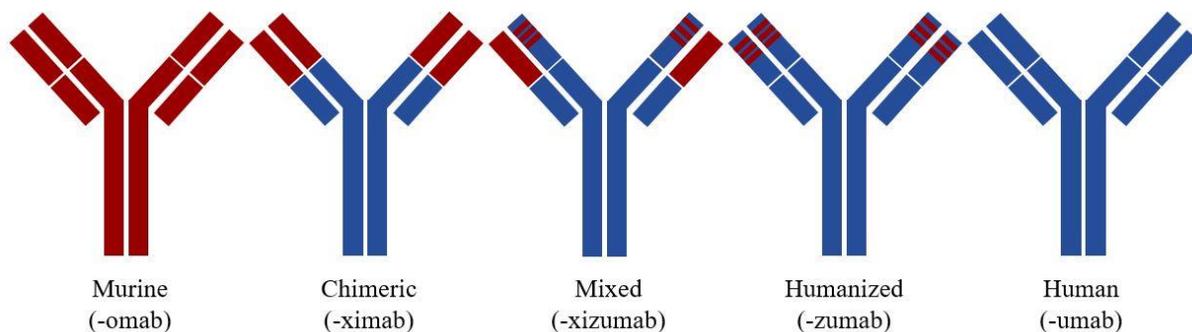


Figure 7.2: Classification of mAbs according to the content of human amino acid sequence (inspired from *Lutz et.al.*^{154,155}. Murine regions in red and human regions in blue.

On average, mAbs are more stable than other protein-drugs, but they are still prone to a variety of physical and chemical degradation pathways, which can compromise their efficacy and/or safety¹⁵⁶. Therefore, it is crucial to find adequate conditions (formulations) that will preserve the stability of mAbs during production, processing, and storage. Many different factors can affect protein-drug stability.

The aim of this study is to provide better understanding on the stability of different mAbs: IgG₁λ (mAb₁), IgG₁κ (mAb₂), and IgG₂κ (mAb₃) (see Table S7.1). The overall stability of all mAbs was investigated by thermal and chemical denaturation studies under different conditions, which were complemented by Small Angle X-ray Scattering (SAXS). Generally, the mAbs have very similar trends in stability, showing increasing stability with increasing pH. However, at low pH, the stability may be increased in the presence of repulsive interactions, which is in agreement with previous studies^{157,158}. In most cases, the tested buffers (histidine, acetate, and phosphate) and excipients (sucrose, arginine, and proline) show very similar effects in different antibodies, which is also in agreement with previously performed studies^{157,158}. However, there are few exceptions. Phosphate has a different effect on the stability of the different mAbs. Additionally, the combination of certain buffer types and excipients lead to different effects in stability. For instance, in histidine buffer at pH 5.0, arginine in combination with NaCl has positive effect on stability of all mAbs, while under other conditions it has strong negative effect. Moreover, in histidine at pH 6.5, proline in combination with NaCl has a strong stabilizing effect on mAb₃.

Results

The overall stability of the antibodies was analyzed by thermal and chemical denaturation using nano differential scanning fluorimetry (nanoDSF) and isothermal chemical denaturation (ICD). The initial screen was performed as a function of pH (5-9) and ionic strength (0, 70, and 140 mM NaCl), and was later complemented by stability studies with varying buffers and excipients.

Thermal and chemical unfolding curves are shown in the Figure 7.3. According to nanoDSF and ICD, both mAb₂ and mAb₃ have a three-state unfolding, although mAb₂ has a more defined intermediate state. Unlike other mAbs, mAb₁ has a two-state thermal unfolding. Moreover, according to the ICD results, mAb₁ is the least stable antibody, as its starts to unfold at very low concentrations of guanidine hydrochloride (GuHCl). Therefore, urea was selected as a denaturant for mAb₁.

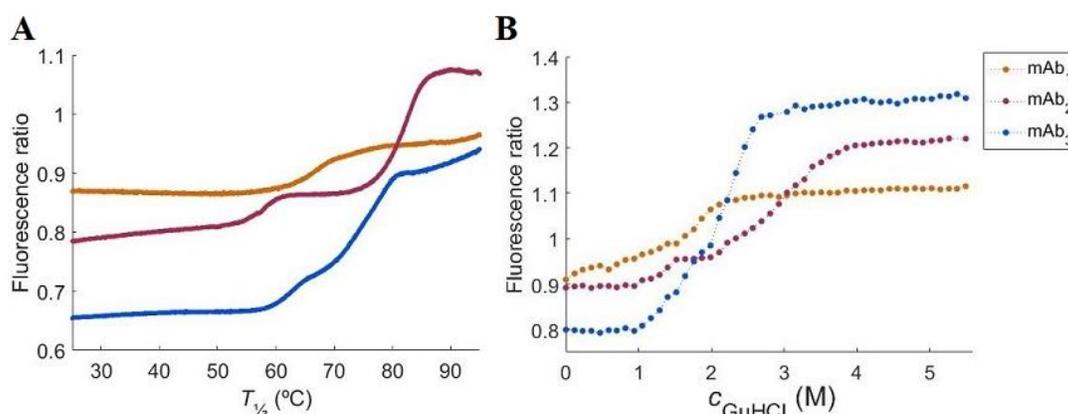


Figure 7.3: Thermal and chemical unfolding curves for mAb₁, mAb₂, and mAb₃ in 10 mM histidine pH 5.0. A: nanoDSF denaturation curves and B: ICD denaturation curves. Color code: mAb₁ in orange, mAb₂ in red, and mAb₃ in blue.

The chemical refolding study points to reversibility of the chemical unfolding (see Table S7.2 in Appendix D), meaning that it is possible to calculate variation in the Gibbs energy of unfolding ($\Delta G_{\text{unfolding}}$). However, the obtained values have large errors, due to the low number of points in the transition area (data not shown). Therefore, chemical denaturation studies were analyzed by only looking at concentration of denaturant required to unfold 50% of the protein ($c_{1/2}$). In order to obtain $\Delta G_{\text{unfolding}}$, additional experiments must be performed where more points are included in the transition area.

pH dependence

According to the thermal denaturation studies (see Figure 7.4A, C, and E), the temperature of unfolding ($T_{1/2}$) increases from pH 5.0 to 6.0-6.5, which points to an increase in stability for all tested mAbs. The chemical denaturation shows a similar picture for mAb₂ and mAb₃, where $c_{1/2}$ increases with increasing pH (see Figure 7.4D and F).

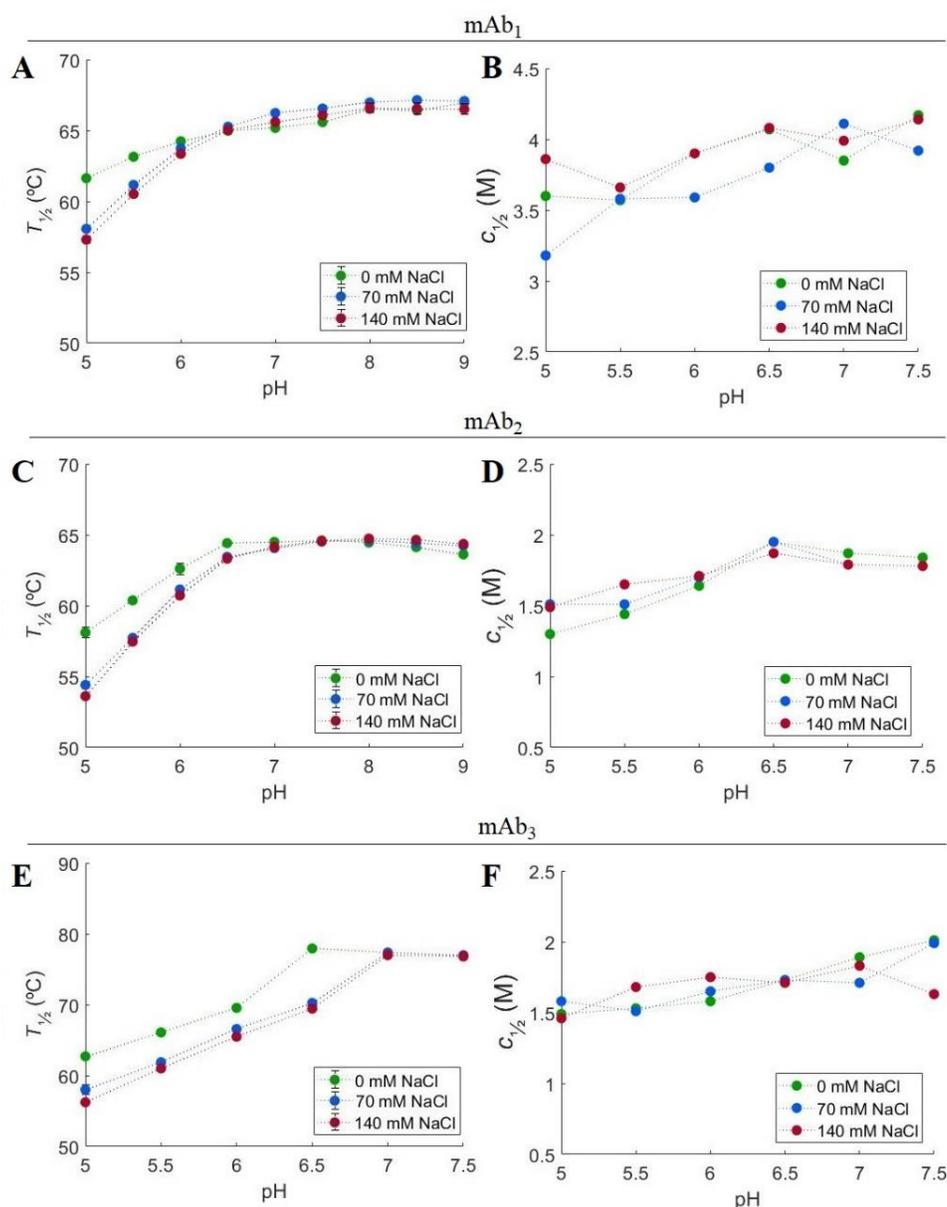


Figure 7.4: Thermal and chemical stability studies performed at different pH (5.0-9.0) and different NaCl concentrations (0, 70, and 140 mM). Changes in $T_{1/2}$ for A: mAb₁, C: mAb₂, and E: mAb₃; changes in $c_{1/2}$ for B: mAb₁, D: mAb₂, and F: mAb₃. Color code: 0 mM NaCl in green, 70 mM NaCl in blue, and 140 mM NaCl in red. mAb₁ denaturation was performed with urea, while mAb₂ and mAb₃ were denatured with GuHCl.

In order to study the interactions and conformational changes, SAXS concentration series data were collected at pH 5.0 and 6.5 for mAb₂ and mAb₃ and at pH 6.0 for all mAbs (see Figure S7.1, S7.2, and S7.4; and

Table S7.5, S7.6, and S7.7 in Appendix D). The interactions were analyzed by looking at the structure factor ($S(0)$): $S(0) > 1$ is characteristic for attractive interactions, $S(0) < 1$ is characteristic for repulsion, and $S(0) = 1$ points to an absence of interactions. At pH 6.0, $S(0)$ of mAb₁ is 1.01, which is typical for an ideal solution, however, an

increasing MW and D_{\max} in combination with a decreasing R_g from the Guinier plot, indicates both repulsive and attractive interactions (see

Table S7.5 in Appendix D). Unlike mAb_1 , both mAb_2 and mAb_3 have $S(0) < 1$, which implies repulsive interactions (see Figure 7.5A). At pH 6.5, while mAb_3 remains repulsive, $S(0)$ for mAb_2 increases up to 1.13, which is characteristic for attractive interactions. Additionally, an increase of D_{\max} , R_g , and MW points to the presence of aggregates (see Table S7.6 in Appendix D).

The pair-distance distribution ($p(r)$) functions of all mAbs have a double-peak, which is typical for multidomain proteins (see Figure 7.5B). The differences in shape of the peaks indicates conformational differences.

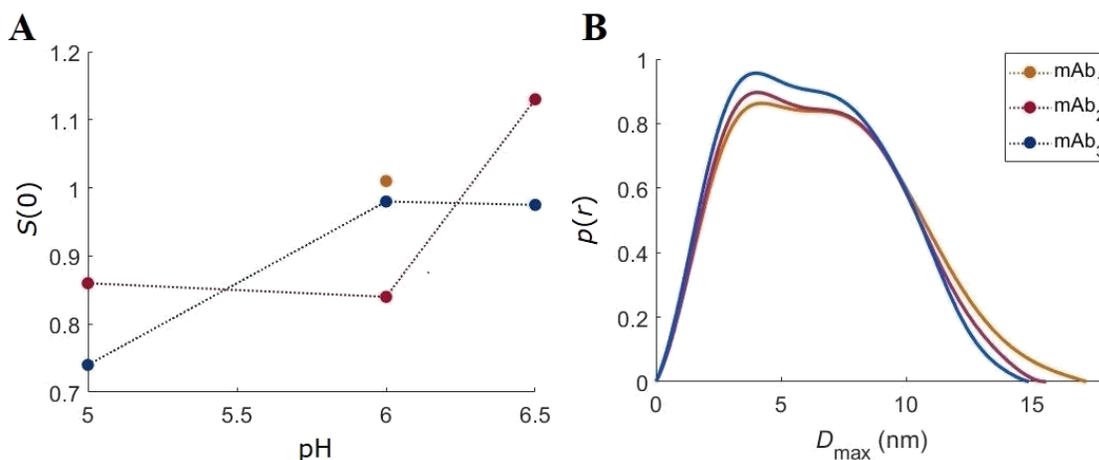


Figure 7.5: Analysis of SAXS data. A: $S(0)$ for mAb_1 , mAb_2 , and mAb_3 at around ≈ 10 g/L at different pH; B: $p(r)$ function for all mAbs at 1 g/L at 10 mM histidine pH 6.0. mAb_1 in orange, mAb_2 in red, and mAb_3 in blue.

The overall conformation at different pHs was investigated using *CORAL* (rigid body modelling). It was not possible to perform modelling on mAb_1 , such as *CORAL* requires amino acid sequence. Obtained models for mAb_2 and mAb_3 are shown in the Figure 7.6. Comparing the obtained models from pH 5.0 and 6.5, it is clear that mAb_2 has significant conformational changes: with increasing pH mAb_2 becomes more elongated and the Fab regions have a tendency to be closer in space (see Figure 7.6A). mAb_3 also shows conformational differences when the Fc domains are superposed (see Figure 7.6C), however, the overall conformation of mAb_3 does not change significantly (see Figure 7.6B), which means that pH does not induce significant conformational changes. Additionally, conformational changes with pH can be observed in the change of the peak shape in the $p(r)$ functions (see Figure 7.6D and E). Changes in the peak shape for mAb_2 are more pronounced than for mAb_3 , which means that the first of these goes through more conformational changes.

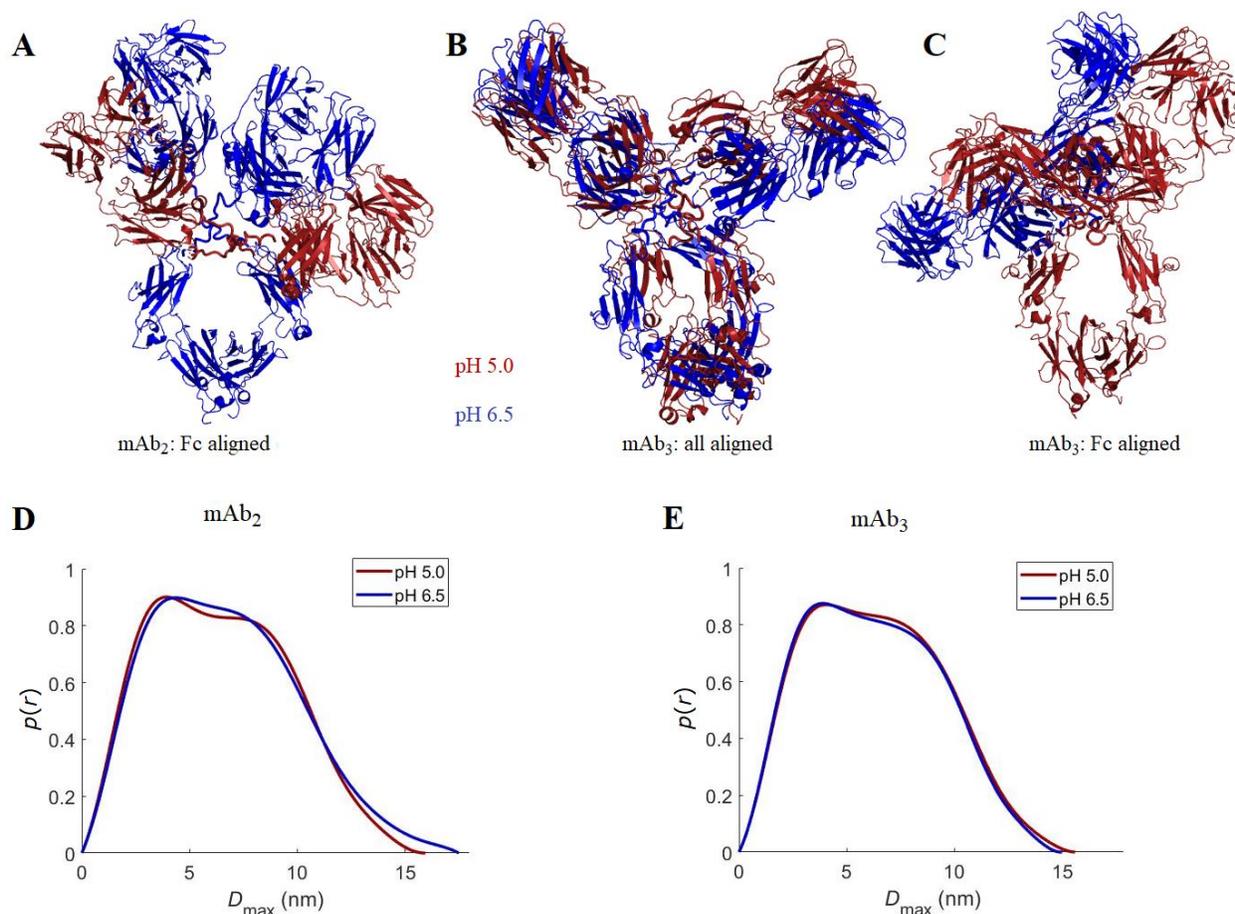


Figure 7.6: Conformational models for mAb₂ and mAb₃ obtained from the rigid body modelling (*CORAL*) at pH 5.0 and 6.5. Rigid body models for A: mAb₂ with Fc aligned, B: all mAb₃ aligned, and C: mAb₃ with Fc aligned. The $p(r)$ functions for D: mAb₂ and E: mAb₃. Color code: pH 5.0 in red and pH 6.5 in blue.

NaCl dependence

The presence of NaCl has a different effect in the thermal and chemical denaturation studies. Up to pH 6.0 for mAb₁ and up to pH 6.5 for mAb₂ and mAb₃, the addition of NaCl decreases $T_{1/2}$, which is an indication of a negative effect on the thermal stability (see Figure 7.4A, C, and E). Meanwhile, chemical denaturation studies do not show a clear trend for the effect of NaCl (see Figure 7.4B, D, and F).

The SAXS experiments performed on mAb₃ do not show significant changes in intramolecular interactions (see Figure S7.4B and E). However, in the presence of NaCl both the radius of gyration (R_g) and D_{\max} are higher (see Table S7.7 in Appendix D). This, along with the change in the peak shape, (see Figure 7.7) points to conformational changes.

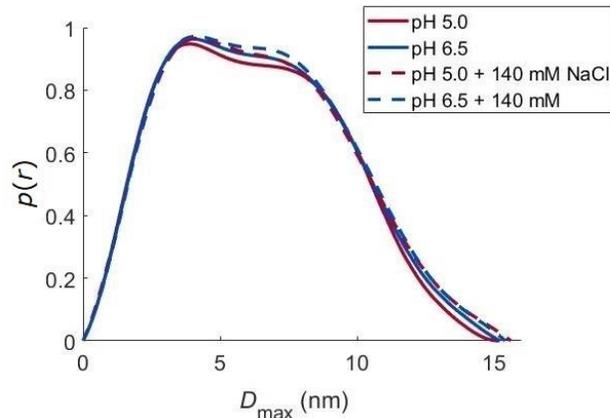


Figure 7.7: Effect of NaCl on the conformation of mAb₃: changes in $p(r)$ functions with addition of NaCl at pH 5.0 and 6.5. Color code: pH 5.0 in red and pH 6.5 in blue. Solid line for a sample without NaCl and dashed line for the sample with NaCl. Experiments performed at mAb₃ concentration around 1 g/L.

Buffer and excipient dependence

The effect of the different buffers and excipients was investigated at pH 5.0 and 6.5. Histidine buffer at pH 5.0 and 6.5 in combination with 0 and 140 mM NaCl was chosen, and additionally, acetate at pH 5.0 and phosphate at pH 6.5 were selected. Sucrose, arginine, and proline were selected as excipients (see Figure 7.8).

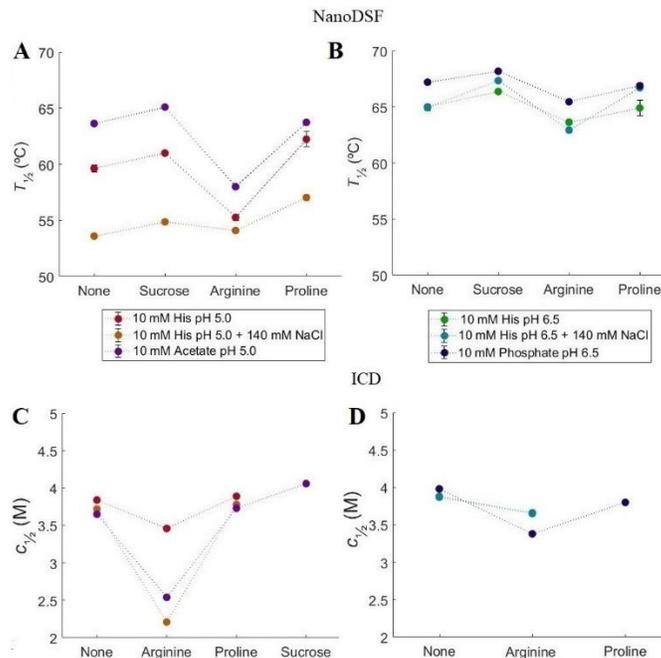


Figure 7.8: NanoDSF and ICD stability studies for mAb₁ using different buffers and excipients. A and B: thermal denaturation studies; C and D: chemical denaturation studies. Purple: 10 mM acetate pH 5.0, red: 10 mM histidine pH 5.0, orange: 10 mM histidine pH 5.0 with 140 mM NaCl, blue: 10 mM phosphate pH 6.5, green: 10 mM histidine pH 6.5, cyan: 10 mM histidine pH 6.5 with 140 mM NaCl.

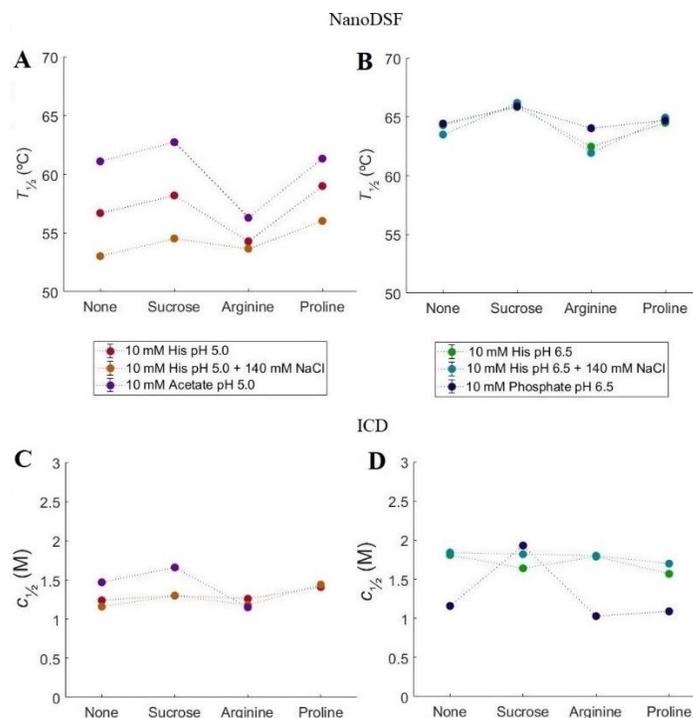


Figure 7.9: NanoDSF and ICD stability studies for mAb₂ using different buffers and excipients. A and B: thermal denaturation studies; C and D: chemical denaturation studies. Purple: 10 mM acetate pH 5.0, red: 10 mM histidine pH 5.0, orange: 10 mM histidine pH 5.0 with 140 mM NaCl, blue: 10 mM phosphate pH 6.5, green: 10 mM histidine pH 6.5, cyan: 10 mM histidine pH 6.5 with 140 mM NaCl.

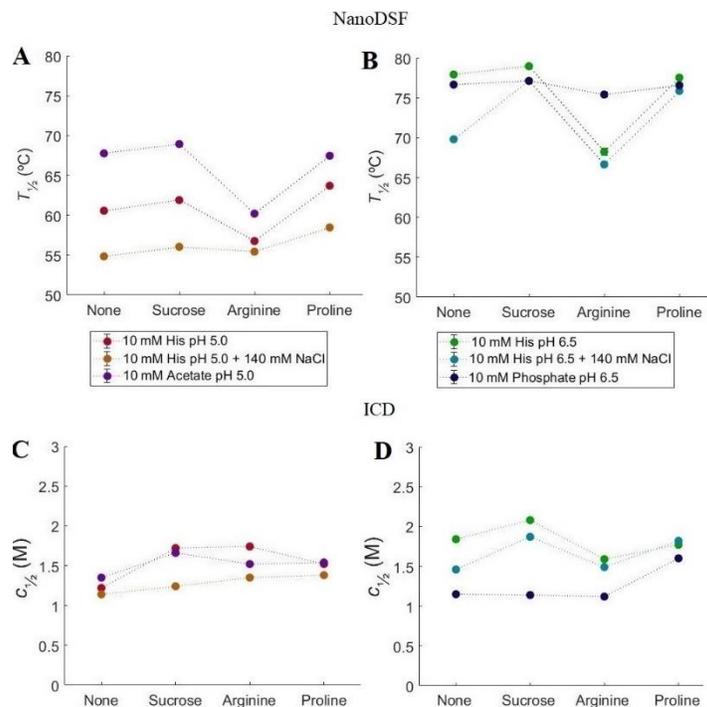


Figure 7.10: NanoDSF and ICD stability studies for mAb₃ using different buffers and excipients. A and B: thermal denaturation studies; C and D: chemical denaturation studies. Purple: 10 mM acetate pH 5.0, red: 10 mM histidine pH 5.0, orange: 10 mM histidine pH 5.0 with 140 mM NaCl, blue: 10 mM phosphate pH 6.5, green: 10 mM histidine pH 6.5, cyan: 10 mM histidine pH 6.5 with 140 mM NaCl.

NanoDSF shows a more pronounced buffer effect at pH 5.0 (see Figure 7.8A, Figure 7.9A, and Figure 7.10A). Comparing the results in the histidine and acetate buffers, all mAbs have higher $T_{1/2}$ (≈ 7 °C) in the acetate buffer, which points to an increase in stability. Addition of NaCl to the histidine buffer decreases $T_{1/2}$ by around 6°C, indicating a decrease in stability.

At pH 6.5, changing from histidine to phosphate buffer increases $T_{1/2}$ by 2.2°C for mAb₁, which suggests a stabilizing effect. In case of mAb₂ and mAb₃, $c_{1/2}$ decreases by around 1 M GuHCl, which indicates a decrease in stability (see Figure 7.8B, Figure 7.9B, and Figure 7.10B). Addition of NaCl to the histidine buffer at pH 6.5 does not change $T_{1/2}$ and $c_{1/2}$ significantly for mAb₁ and mAb₂. However, in case of mAb₃, $T_{1/2}$ decreases by 8°C and $c_{1/2}$ decreases by 0.4 M, which indicates a decrease in stability.

From all tested excipients, sucrose is the only excipient that increases $T_{1/2}$ and $c_{1/2}$ for all mAbs in all tested conditions (see Figure 7.8, Figure 7.9, and Figure 7.10). Proline shows different effects under different conditions. In the histidine buffer pH 5.0 proline increases $T_{1/2}$ by around 2-4°C for all mAbs and $c_{1/2}$ by around 0.2-0.3 M for mAb₂ and mAb₃. In the acetate buffer, proline does not have a significant effect for mAb₁ and mAb₃, but it decreases $c_{1/2}$ of mAb₂ by 1.5 M, which is an indication of decreasing stability. At pH 6.5, for most tested conditions proline does not contribute to significant changes in stability, with one exception for mAb₃. In histidine buffer, in combination with NaCl, proline increases $T_{1/2}$ by 6°C and $c_{1/2}$ by 0.4 M, which suggests an increase in stability.

Finally, the presence of arginine has a destabilizing effect in almost all tested conditions, with only one exception in histidine at pH 5.0 in the presence of NaCl. In combination with NaCl, arginine increases $T_{1/2}$ by 0.6°C for all mAbs, but decreases $c_{1/2}$ for mAb₁ by 1.5 M. According to the SAXS data collected for mAb₂, the presence of arginine changes the peak shape of the $p(r)$ function, which means changes in conformation and, also, increases D_{\max} (see Figure 7.11), which together with characteristic “tail” is an indication of the presence of aggregates.

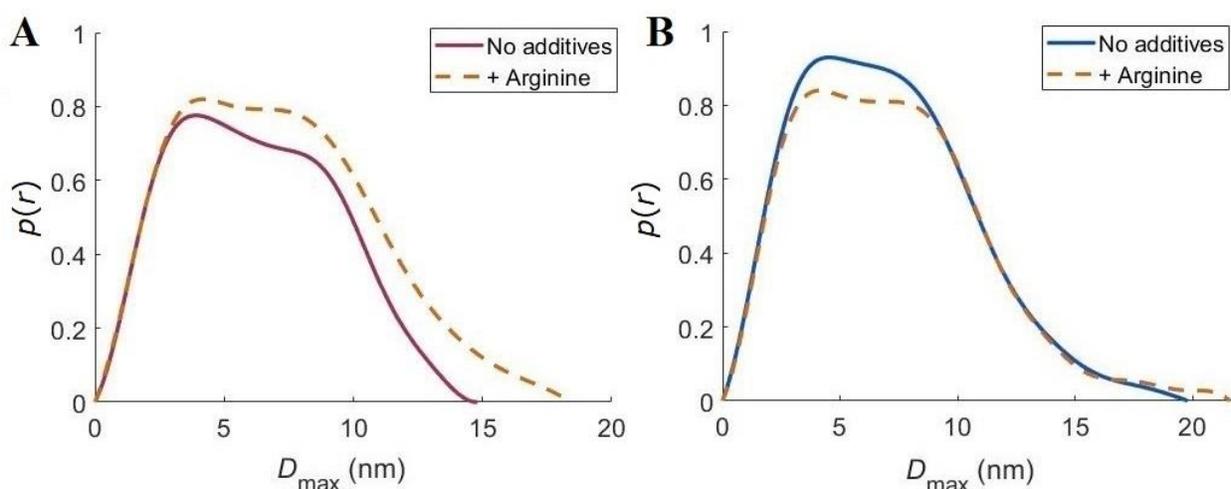


Figure 7.11: Effect of arginine on the conformation of mAb₂: changes in $p(r)$ functions at A: pH 5.0 and B: pH 6.5. The Solid line is a sample without arginine; dashed line is a sample with arginine. Experiments performed at a mAb₂ concentration around 8 g/L.

Chemical denaturation

As already mentioned, chemical unfolding is a multi-step process for all tested mAbs. In order to understand the unfolding mechanism SAXS studies were performed on mAb₁ in the presence of urea, and on mAb₂ in the presence of GuHCl (see Figure 7.12).

In both cases, the intensity of the peak in the Kratky plot is decreasing with increasing concentration of denaturant, which happens due to decreasing contrast between the protein and the solvent. Despite that, the Kratky plot of mAb₁ does not change significantly up to 2.5 M of urea (see Figure 7.12B). Only in the presence of 5 M urea the intensity at high q -value increases, which is characteristic for flexible/unfolded proteins. The $p(r)$ functions for different concentrations of denaturants were obtained from scaled SAXS curves, in order to make it easier to monitor changes in the peak. In the $p(r)$ function, up to 2.5 M of urea, intensity of the second peak is changing with urea concentration, pointing to small conformational changes (see Figure 7.12C). From 2.5 to 5 M of urea, the double peak changes into a single peak, pointing to significant conformational changes. Additionally, D_{\max} decreases with addition of urea up to 2 M, and increases from 2 to 5 M.

Similarly to mAb₁, the Kratky plot for mAb₂ does not change significantly up to 3 M GuHCl. Only in the presence of 4.5 M an increase can be seen in the intensity at high q -values, which is characteristic for flexible/unfolded proteins (see Figure 7.12E). However, even at high concentrations of denaturant, the peak is still present, which means that mAb₂ is not fully unfolded, and some structure is preserved. Just like mAb₁, the second peak in the $p(r)$ function for mAb₂ is changing with concentration of denaturant, and at the highest concentration the double peak changes into a single peak (see Figure 7.12F). With increasing concentration of denaturant D_{\max} increases from 15 to around 25 nm.

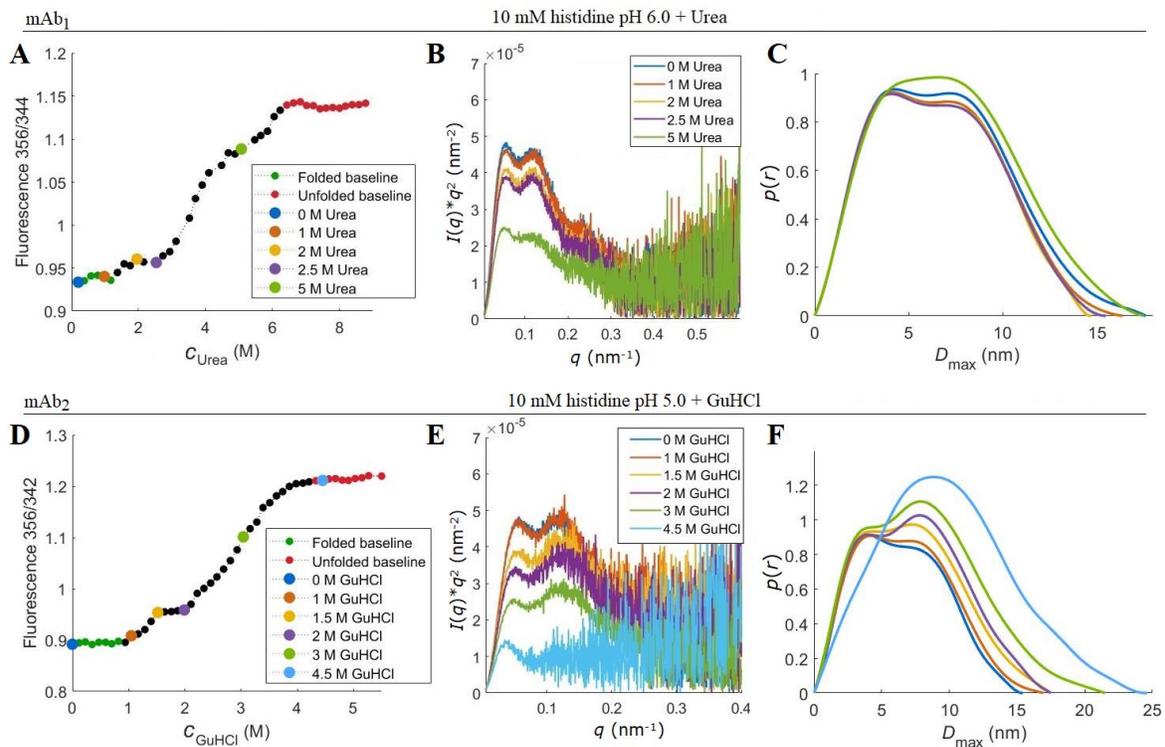


Figure 7.12: SAXS denaturation studies for mAb₁ with urea (A, B, and C) and for mAb₂ with GuHCl (D, E, and F). A and D: ICD unfolding curves; B and E: Kratky plots; C and F: $p(r)$ function created from the scaled data for better visualization of changes in the shape of the peak.

Discussion

pH and NaCl effect

The effect of different tested conditions is shown in the Figure 7.13. From the data analysis it is clear that all antibodies have a very similar pH and NaCl effect. Decreasing pH leads to a decrease in stability, which is especially pronounced from pH 6.0 for mAb₁ and from pH 6.5 for mAb₂ and mAb₃. According to the SAXS data (collected for mAb₂ and mAb₃), decreasing pH leads to an increase in repulsion (see Figure 7.5A), which seems to be related to a decrease in stability. However, addition of NaCl that screens charges and decreases repulsion, leads to decreasing stability, meaning that at low pH repulsive interactions actually improve stability of tested antibodies, and the negative effect is not related to protein-protein interactions. Possibly, decreasing pH induces chemical degradation by activating multiple degradation pathways (*i.e.* disulfide bond formation/exchange, deamination, fragmentation, isomerization, and others)¹⁵⁶. Negative effect of pH and NaCl on stability of mAbs is in agreement with previous studies performed by *Feng et.al*¹⁵⁷. Additionally, despite the fact that mAbs are more stable at pH 6.5, SAXS points to the presence of larger aggregates when compared to pH 5.0, which is in agreement with study performed by *Xinsheng et.al*¹⁵⁸.

Moreover, rigid body modelling shows that decreasing pH induces significant conformational changes in mAb₂, but not in mAb₃ (see Figure 7.6). One of the reasons might be the number of disulfide (S-S) bonds in the hinge region. mAb₃ is a IgG₂, which has four S-S bonds, while mAb₂ is IgG₁, which has only two S-S bonds. The presence of additional two S-S bonds reduces the flexibility of the Fab region, which makes overall structure more rigid.

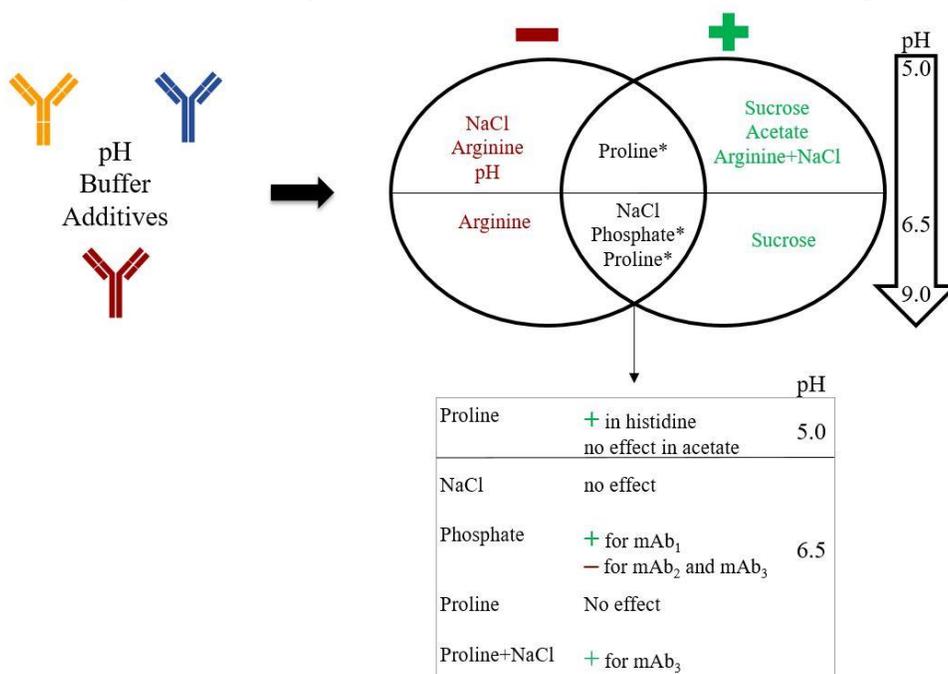


Figure 7.13: Overview of the effect of pH, NaCl, buffers, and excipients on stability of tested mAbs.

Buffer and excipients effect

In the current study, buffer effect was investigated by testing three different buffers: acetate at pH 5.0, phosphate at pH 6.5, and histidine in both pH 5.0 and 6.5. Effect of excipients was analyzed by testing sucrose, arginine, and proline in different buffers at pH 5.0 and 6.5 (see Figure 7.8, Figure 7.9, Figure 7.10, and Table 7.2). NanoDSF and ICD do not show exactly the same effect, due to the different denaturation process. In most of the cases, when $T_{1/2}$ increases/decreases, $c_{1/2}$ shows no significant effect and vice versa. Therefore, general effect of buffers and excipients was analyzed by looking at both $T_{1/2}$ and $c_{1/2}$.

Table 7.2: Overview of thermal and chemical denaturation studies using different buffers and excipients.

	Additives	nanoDSF			ICD		
		$T_{1/2}$			$c_{1/2}$		
		mAb ₁	mAb ₂	mAb ₃	mAb ₁	mAb ₂	mAb ₃
↑ pH		+++	+++	++++	0	++	++
	NaCl	--	--	---	0	0	0
Acetate pH 5.0		++	++	+++	0	+	0
	Sucrose	+	+	+	+	0	+
	Arginine	---	--	---	---	-	0
	Proline	0	0	0	0	---	0
Histidine pH 5.0	Sucrose	+	+	+	x	0	++
	Arginine	--	--	--	-	0	++
	Proline	++	++	++	0	0	+
+ NaCl	Sucrose	+	+	+	x	0	0
	Arginine	+	+	+	----	+	+
	Proline	++	++	++	0	+	+
Phosphate pH 6.5		++	0	-	0	--	--
	Sucrose	+	+	0	x	++	0
	Arginine	-	0	-	--	0	0
	Proline	0	0	0	0	0	+
Histidine pH 6.5	Sucrose	+	+	+	x	0	+
	Arginine	-	-	---	-	0	-
	Proline	0	0	0	x	-	0
+ NaCl	Sucrose	++	++	+++	x	0	+
	Arginine	--	-	--	-	0	0
	Proline	+	-	+++	x	0	+
Reference point for buffers: pH 5.0 – histidine pH 5.0; pH 6.5 – histidine pH 6.5 Reference point for excipients: respective buffer without excipient + stabilizes - destabilizes 0 no trend for pH/NaCl study x – data not acquired ++++/----- $\Delta T_{1/2} > 10^\circ\text{C}$ $\Delta c_{1/2} > 1.5 \text{ M}$ +++/---- $5^\circ\text{C} < \Delta T_{1/2} < 10^\circ\text{C}$ $1 \text{ M} < \Delta c_{1/2} < 1.5 \text{ M}$ ++/-- $2^\circ\text{C} < \Delta T_{1/2} < 5^\circ\text{C}$ $0.5 \text{ M} < \Delta c_{1/2} < 1.0 \text{ M}$ +/- $0.5^\circ\text{C} < \Delta T_{1/2} < 2^\circ\text{C}$ $0.2 \text{ M} < \Delta c_{1/2} < 0.5 \text{ M}$ 0 $\Delta T_{1/2} < 0.5^\circ\text{C}$ $\Delta c_{1/2} < 0.2 \text{ M}$							

Comparing buffers tested at pH 5.0, all mAbs have higher $T_{1/2}$ in acetate buffer, meaning positive effect on stability, which is in agreement with previous studies¹⁵⁷. At pH 6.5, phosphate has different effect on stability of different mAbs: for mAb₁ it has low positive effect (increases $T_{1/2}$), while for mAb₂ and mAb₃ it has negative effect (decreases $c_{1/2}$).

From all tested excipients, sucrose is the only excipient that has a positive effect for all mAbs at all conditions, most likely due to the preferential exclusion of the sucrose from the protein domain, which results in increase of free energy of the system^{159,160}. This is in agreement with previous studies performed by *Feng et.al.*¹⁵⁷ and *Xinsheng et.al.*¹⁵⁸.

Unlike sucrose, arginine has a destabilizing effect in almost all tested conditions, which is also in agreement with previous studies¹⁵⁷. Additionally, according to the SAXS data (see Figure 7.11), presence of arginine induces conformational changes in mAb₂ and increases MW . In previous chapters it was shown that arginine binds to both transferrin and albumin (see chapter 5 and 6). This could suggest that arginine also binds to mAbs, resulting in conformational changes inducing a decrease in stability and aggregation. In histidine buffer at pH 5.0 in combination with NaCl, arginine shows an opposite effect. When combined with NaCl, arginine slightly increases $T_{1/2}$ for all mAbs and $c_{1/2}$ for mAb₂ and mAb₃. This could indicate that the interactions between arginine/NaCl and the mAbs is altered when the other constituent also is present.

The presence of proline shows different effects in different conditions. In histidine pH 5.0 proline increases overall stability of mAbs, but by switching to the acetate buffer it has no effect for mAb₁ and mAb₃, and negative effect for mAb₂. At pH 6.5, proline does not have significant effect, with only exception in histidine in the presence of NaCl, where it increases stability of mAb₃. mAb₃ is a IgG₂, meaning that the main difference between it and the other tested mAbs is in the hinge region: where mAb₃ has four S-S bonds. Unlike other excipients, conclusions for proline is not fully in agreement with previous studies, where it is stated that proline has a minimal positive effect¹⁵⁷.

Chemical denaturation

Comparing chemical denaturation of mAb₁ and mAb₂ (both IgG₁), it is clear that different denaturant leads to a different unfolding process. As already mentioned, the SAXS data points to the presence of aggregates for mAb₁ in histidine pH 6.0. Addition of up to 2 M of urea results in decrease of D_{max} , which is an indication of decreasing amount of aggregates. From 2.5 M, D_{max} starts to increase and the shape of the peak shifts from double peak to a single peak, pointing to unfolding of mAb₁ (see Figure 7.12C).

Unlike urea, GuHCl does not show stabilizing effect for mAb₂. Compared to urea, GuHCl is a stronger denaturant, which induces significant conformational changes at the same concentrations (compare Figure 7.12C and F). With increasing concentration of GuHCl, D_{max} increases and the shape of the peak has significant changes, which is an indication of unfolding. The multi-step unfolding can be explained by different stability of C_{H2}, Fab, and C_{H3} domains. According to the previous studies, C_{H2} domain is the least stable domain, followed by Fab region that is the second least stable for IgG₁ and IgG₂^{161,162,163}. Therefore, chemical unfolding might start

with unfolding of C_{H2} , creating the first partially-unfolded intermediate (see Figure 7.14). As Fab and C_{H3} remain folded, the overall structure of the antibody is preserved, and SAXS does not show significant differences in $p(r)$ function up to 1.5 M of GuHCl (see Figure 7.12F). Increasing the concentration of GuHCl up to 3 M results in significant conformational changes, due to unfolding of the Fab domains. The shape of the peak of the $p(r)$ function changes significantly, however it remains as a double peak, characteristic for multidomain proteins. This is because C_{H3} domain remains folded, while the rest of the mAb unfolds. Finally, increasing concentration up to 4.5 M of GuHCl results in an increase in D_{max} and significant changes in the shape of the peak, which shifts from double peak to a single peak. At this stage, mAb₂ is fully unfolded, but due to the presence of S-S bonds, the Kratky plot has a small peak instead of plateau that is characteristic for fully unfolded proteins.

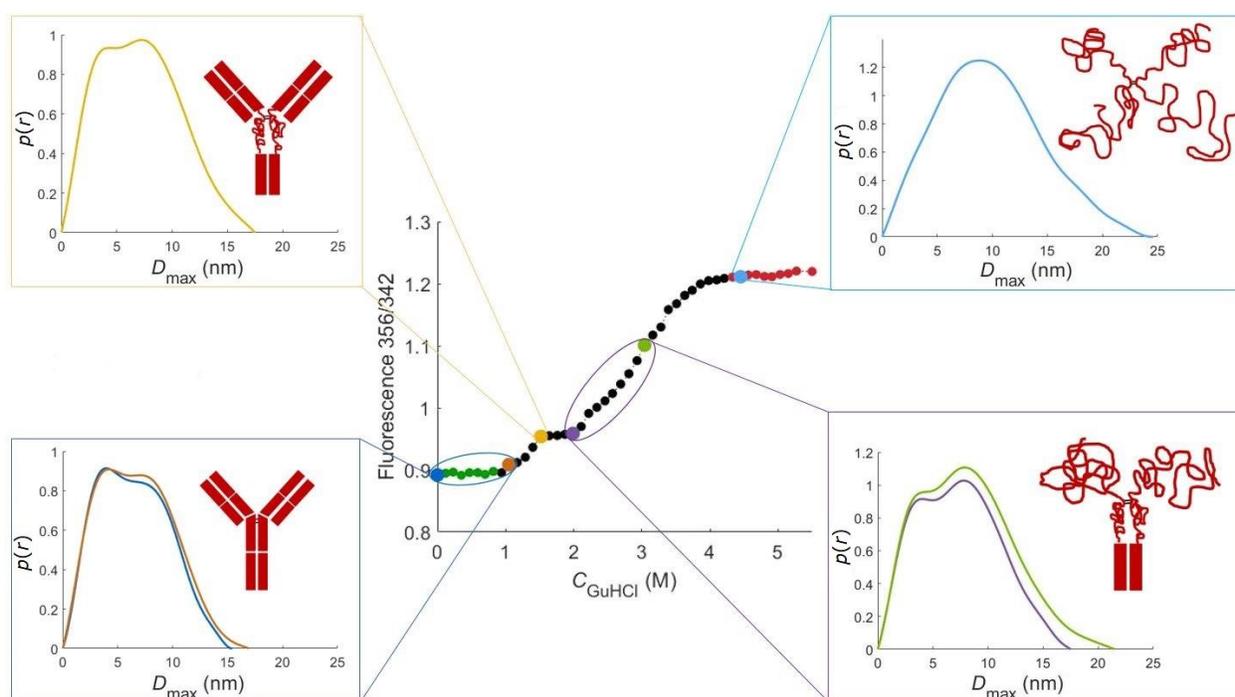


Figure 7.14: Chemical unfolding of mAb₂.

Additionally, ICD studies show that mAb₂ and mAb₃ have different unfolding curves, even though in both cases GuHCl was used as a denaturant. It can be explained by different thermal stability of Fab and C_{H3} . Differential scanning fluorimetry studies performed by *Garber et al.*¹⁶³ show that, in IgG₁, Fab unfolds around 77°C, while C_{H3} unfolds around 84°C, pointing to a significant difference in thermal stability. This difference in stability results in presence of well-defined intermediate state, where both C_{H2} and Fab are unfolded, but C_{H3} remains folded (see Figure 7.15). Unlike in IgG₁, Fab and C_{H3} domains in IgG₂ do not have a significant difference in thermal stability and unfold, respectively, at around 78°C and 80°C. Hence, C_{H3} starts to unfold immediately after Fab and no clear intermediate state is observed (see Figure 7.15).

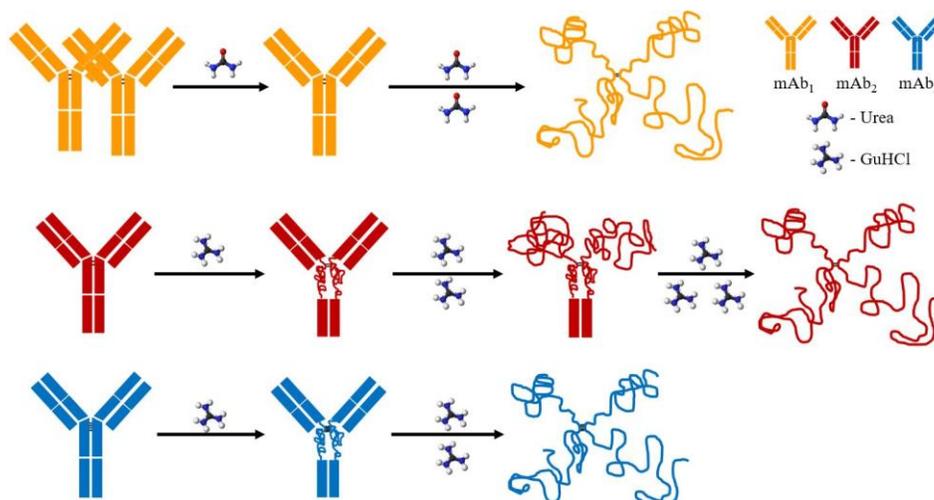


Figure 7.15: Proposed mechanism of chemical unfolding for mAb₁ (in yellow), mAb₂ (in red), and mAb₃ (in blue).

Conclusion

In conclusion, despite the similar sequences and structures, the investigated mAbs show differences in stability. In most of the cases, tested conditions show similar effects on all three mAbs (*i.e.* pH, NaCl, acetate, sucrose, and arginine). This, however, is not the case for phosphate and proline with NaCl at pH 6.5. Phosphate stabilizes mAb₁, but destabilizes the other two mAbs, while proline with NaCl has a strong positive effect on the stability of mAb₃, but no effect on the other mAbs. Additionally, the combination of different excipients can result in an opposite effect: that both arginine and NaCl has a negative effect, but combined they result in an increase in thermal stability for all mAbs.

Thermal and chemical denaturation studies do not always show the same result, most likely due to the different denaturation processes when we are performing these experiments.

Finally, when combining SAXS and ICD it was possible to propose mechanisms of unfolding for all three mAbs. It was shown that the different mAbs have different unfolding mechanisms. Moreover, when using different denaturants, these also affect the unfolding process. This topic can be further investigated by performing ICD-SAXS studies on all mAbs using both urea and GuHCl.

Some trends, *e.g.* stabilizing effect of acetate and different effect of certain excipients/combination of excipients (*i.e.* phosphate buffer, proline, and proline with NaCl) are difficult to explain. In order to understand those trends, it is necessary to perform *in-silico* modelling or other experiments that probes whether these have distinct binding to different mAbs, and where they bind.

Methods

Dialysis and formulation

AstraZeneca provided monoclonal antibodies in different concentrations: mAb₁ in 21 g/L, mAb₂ in 48.9 g/L, and mAb₃ in 54.3 g/L (see Table S7.1 in Appendix D). The stock solutions were dialyzed into 10 mM histidine pH 5.5, 7.0 and 10 mM tris pH 8.5 for pH and NaCl screening. For the buffer and excipients dependence study, all mAbs were dialyzed into 10 mM histidine pH 5.0 and 6.5, 10 mM acetate pH 5.0 and 10 mM phosphate pH 6.5. Final solutions were obtained by diluting dialyzed proteins into the right buffer (with pH±0.5 and/or including additives) (see Figure 7.16). Protein concentration was measured on a Nanodrop™ 1000 (Thermo Fisher Scientific, Waltham, USA) using extinction coefficient provided by AstraZeneca (see Table S7.1 in Appendix D).

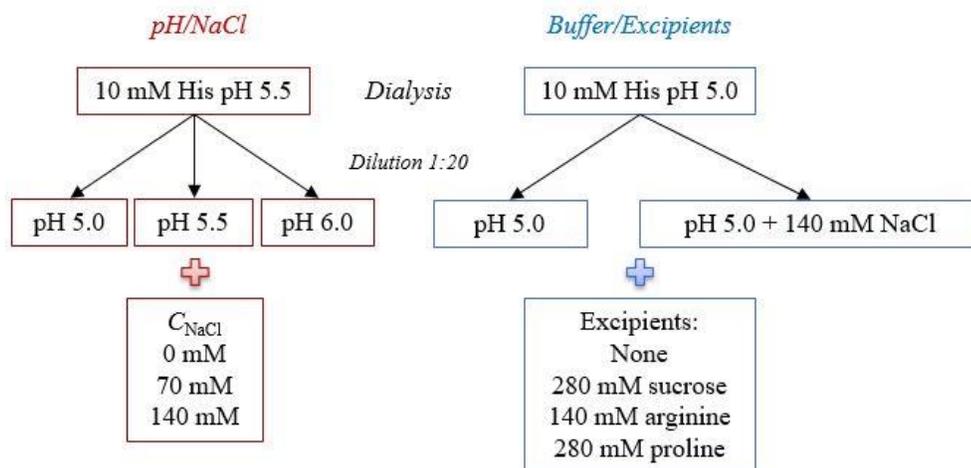


Figure 7.16: Schematic representation of dialysis and formulation process.

Thermal stability studies

Thermal denaturation studies were carried out in the Prometheus NT.48 (NanoTemper Technologies, Munich, Germany) using NanoDSF Grade Standard Capillaries, which were manually loaded with 10 µl of protein at 1 g/L in the final conditions. All experiments were performed at the heating rate of 1°C/min from 20 to 95°C. Unfolding process was monitored by the shift in the intrinsic fluorescence spectra (350/330 nm). All measurements were done in triplicates and the data was analyzed using PR.Control v1.12.2 software (NanoTemper Technologies, Munich, Germany).

Isothermal Chemical Denaturation

Chemical denaturation studies were carried out in the Unchained Labs HUNK system - AVIA ICD 2304 (Unchained Labs, Pleasanton, USA). The data was collected using excitation wavelength of 285 nm, and emission intensities were recorded from 300 nm to 450 nm. The gain setting was set for 10 and the incubation time was set for additional 1134 min, based on previously performed

gain and incubation tests. 48-point linear gradient of denaturant was automatically generated for each condition. For the pH and NaCl screening both urea and guanidine hydrochloride (GuHCl) were used as denaturants. For the buffer and excipients dependence urea was selected for mAb₁ and GuHCl was selected for mAb₂ and mAb₃. Urea and GuHCl denaturation buffers were prepared by adding, respectively, 10 M of urea and 6 M of GuHCl in each tested condition. 1 g/L protein stocks were subsequently diluted 12.5 times into a final condition. The data collection and analysis were performed by Formulator software v3.02 (Unchained Labs, Pleasanton, USA). Unfolding process was monitored by looking at the shift in the maximum fluorescence with increasing concentration of denaturant: 356/344 nm for mAb₁, 356/342 nm for mAb₂, and 358/338 for mAb₃. In order to minimize the error, a secondary fit was performed for each pH value combining different NaCl concentrations. Free energy of unfolding (ΔG_{unfold}), $c_{1/2}$, and m -values were calculated for all transitions.

Small Angle X-ray Scattering

Data collection was performed at the P12 beamline at the Petra III storage ring (DESY, Hamburg DE)¹⁰⁶ (see Table S7.4 in Appendix D). Radius of gyration (R_g) and maximum dimension (D_{max}) were derived from the experimental data with the graphical data analysis program *PRIMUSQT*⁴⁶. Rigid body modelling was performed using *CORAL*⁴³ against previously merged scattering curves.

Chapter 8

Overall discussion

This study shows that combination of multiple methods leads to a better understanding of protein stability. Studies of the conformational stability performed using nanoDSF and ICD give complementary information about the overall conformational stability of the proteins, whereas SAXS and MD provide molecular insides, which are important to understand the rationale behind the changes in stability. Moreover, the multidomain proteins (transferrin (rTrF), albumin-neprilysin fusion protein (HSA-NEP), and monoclonal antibodies (mAb₁, mAb₂, and mAb₃)), despite of their significant structural and functional differences, have some common stability trends.

A summary of nanoDSF and ICD results for all proteins are shown in Table 8.1. It was not possible to find the same stability trends for all the proteins. NanoDSF shows clear pH and NaCl effects: rTrF and mAbs become more stable with increasing pH, while the NaCl effect depends on pH of the sample. Although these trends are similar, the molecular reasoning is different. In case of rTrF, increasing pH shifts the equilibrium towards the most stable (closed) conformation, while mAbs become less prone to chemical degradation at higher pH¹⁵⁶. Presence of NaCl does not have an effect at high pH. For rTrF, NaCl at low pH enhances opening of the N-lobe, shifting the equilibrium towards the less stable (partially open) conformation (see Figure 8.1A). In case of mAbs, which are positively charged at low pH (pI \approx 8), addition of NaCl leads to decreasing intermolecular repulsion, which results in increased tendency to aggregate and decreasing stability (see Figure 8.1B).

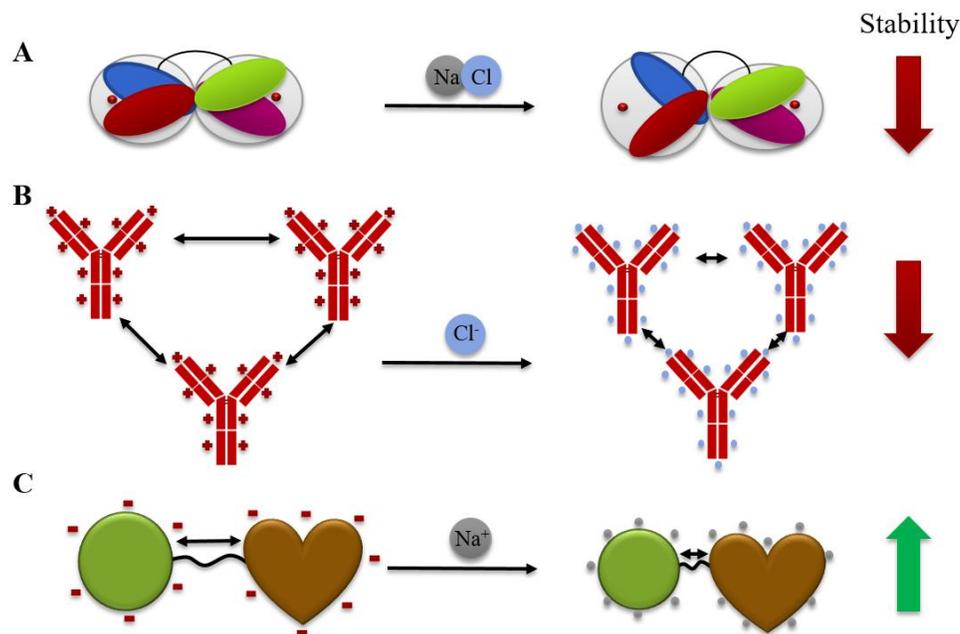


Figure 8.1: Effect of NaCl on rTrF, mAbs, and HSA-NEP. A: addition of NaCl shifts the equilibrium from closed to partially open conformation, which results in decreasing stability. B: chloride ions (Cl⁻) screen positive charges on mAbs surface, which leads to decrease in intermolecular repulsion and increase in stability. C: sodium ions (Na⁺) screen negative charged on HSA-NEP, which results in a decrease of intramolecular repulsion and increase in stability.

HSA-NEP fusion protein has different stability trends: it is more stable at low and high pH values (at pH 5, and from 8 to 9) and in the presence of NaCl. Unlike rTrF and mAbs, HSA-NEP is fused using recombinant DNA technology and originally HSA and NEP exist as separate proteins. Both HSA and NEP have pI around 5, meaning that with increasing pH values both become negatively charged and therefore the protein will have destabilizing intramolecular repulsion. Addition of NaCl screens negative charges, which decreases repulsion and leads to an increase in stability (see Figure 8.1C). Surprisingly, from pH 8 specific interaction at the interface between HSA and NEP create new more stable conformation, which increases the overall stability of the fusion protein.

According to the nanoDSF results, most of the studied excipients have a similar effect on the thermal stability of all proteins (see Figure 8.2 and Table 8.1). Generally, acetate, sucrose, and proline improve stability, while arginine seems to have a negative effect. However, combination of different excipients can have an opposite effect: *e.g.* combination of arginine and NaCl has positive effect on all the proteins, whereas proline with phosphate decrease the stability of rTrF and HSA-NEP. For a better understanding, combination of different excipients must be further investigated with molecular dynamics.

Table 8.1: Overview of thermal and chemical denaturation studies for rTrF, HSA-NEP (H-N), and mAbs using different buffers and excipients.

	Additives	nanoDSF					ICD				
		$T_{1/2}$					$C_{1/2}$				
		rTrF	H-N	mAb ₁	mAb ₂	mAb ₃	rTrF	H-N	mAb ₁	mAb ₂	mAb ₃
Acetate pH 5.0		++	++	++	++	+++	0	+	0	+	0
	Sucrose	+	+	+	+	+	+	+	+	0	+
	Arginine	----	--	---	--	---	+	--	---	-	0
	Proline	0	0	0	0	0	+	0	0	---	0
Histidine pH 5.0	Sucrose	+	+	+	+	+	++	0	x	0	++
	Arginine	----	-	--	--	--	++	0	-	0	++
	Proline	++	+	++	++	++	+	0	0	0	+
+ NaCl		----	-	---	--	---	0	0	0	0	0
	Sucrose	+	+	+	+	+	+	X	x	0	0
	Arginine	+	+	+	+	+	+	--	----	+	+
	Proline	++	++	++	++	++	0	+	0	+	+
Phosphate pH 6.5		----	++++	++	0	-	-	++	0	--	--
	Sucrose	++	+	+	+	0	0	X	x	++	0
	Arginine	-	--	-	0	-	0	+	--	0	0
	Proline	-	-	0	0	0	+	++	0	0	+
Histidine pH 6.5	Sucrose	+	++	+	+	+	+	X	x	0	+
	Arginine	---	+++	-	-	---	-	0	-	0	-
	Proline	-	+	0	0	0	0	X	x	-	0
+ NaCl		--	+++	0	-	---	++	0	0	0	-
	Sucrose	+++	--	++	++	+++	-	x	x	0	+
	Arginine	---	-	--	-	--	0	-	-	0	0
	Proline	++	---	+	-	+++	-	x	x	0	+

Reference point for buffers: pH 5.0 – histidine pH 5.0; pH 6.5 – histidine pH 6.5
Reference point for excipients: respective buffer without excipient
+ stabilizes
- destabilizes
0 no trend for pH/NaCl study
x – data not acquired

++++/----	$\Delta T_{1/2} > 10^\circ\text{C}$	$\Delta C_{1/2} > 1.5 \text{ M}$
+++/-	$5^\circ\text{C} < \Delta T_{1/2} < 10^\circ\text{C}$	$1 \text{ M} < \Delta C_{1/2} < 1.5 \text{ M}$
++/--	$2^\circ\text{C} < \Delta T_{1/2} < 5^\circ\text{C}$	$0.5 \text{ M} < \Delta C_{1/2} < 1.0 \text{ M}$
+/-	$0.5^\circ\text{C} < \Delta T_{1/2} < 2^\circ\text{C}$	$0.2 \text{ M} < \Delta C_{1/2} < 0.5 \text{ M}$
0	$\Delta T_{1/2} < 0.5^\circ\text{C}$	$\Delta C_{1/2} < 0.2 \text{ M}$

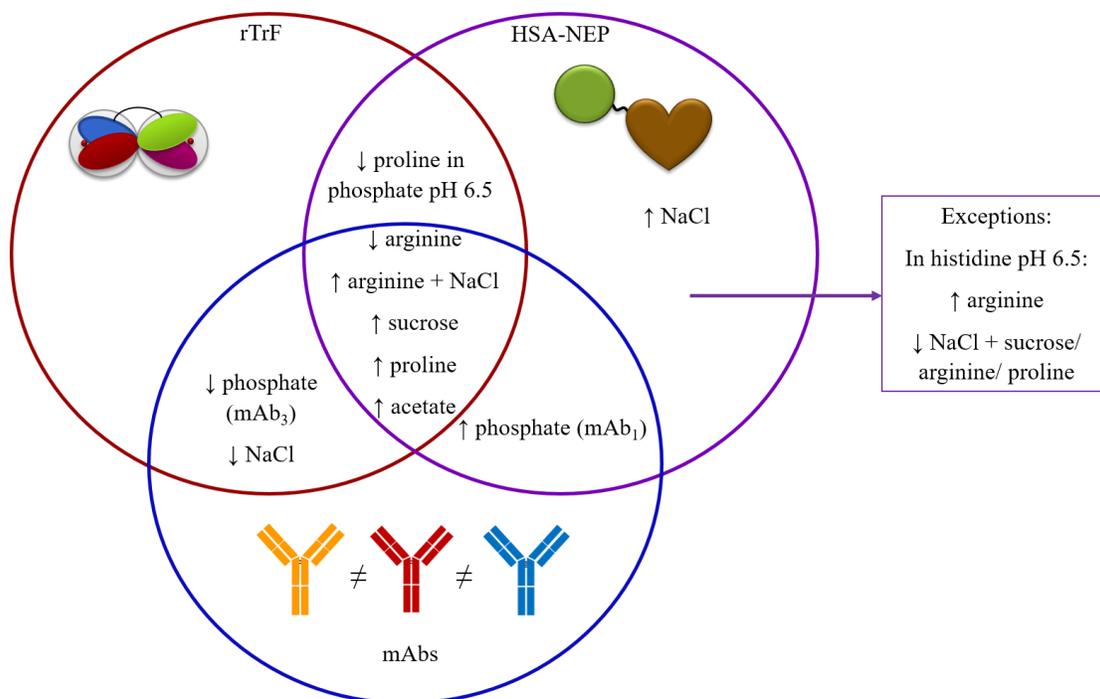


Figure 8.2: Overview of the excipients and buffer effect on stability of rTrF, mAbs, and HSA-NEP. Arrow up/down symbolizes positive/negative effect on the protein thermal stability.

ICD does not show a clear trend when NaCl is added, while pH and excipients effect is not always in agreement with the results from nanoDSF studies (see Table 8.1). The possible reason is the presence of the high concentration of denaturant, which might create additional interactions, leading to a different unfolding process. An example is the denaturant guanidine hydrochloride, which is salt that affects the ionic strength of the protein as it is added. Moreover, denaturation agent might interact with excipients, meaning that higher concentrations of excipients are required to be able to see a clear effect on the protein stability.

Clearly, the combination of ICD and SAXS leads to a better understanding of the protein unfolding process. SAXS data collected for the protein at different concentrations of denaturant provides information about protein dimension, average molecular weight, overall shape, and flexibility. Additionally, ICD can provide valuable thermodynamic parameters: free energy of unfolding (ΔG_{unfold}) and m -values (related to the changes in solvent accessible area). However, due to the low number of data points in the transition region from folded to unfolded state, ΔG_{unfold} and m -values have high uncertainty. Therefore, in order to obtain the values with the acceptable error, it is important to optimize each the concentration range and perform the same experiment in duplicates or in triplicates.

Moreover, initial stability studies for pH and NaCl correlate to long-term stability studies¹⁶⁴, meaning that both nanoDSF and ICD are good tools for initial screening in biopharmaceutical development. A similar study will be performed to look at correlation for buffer species and excipients.

Chapter 9

Conclusion

High throughput stability studies in combination with structural studies lead to a better understanding of protein stability. The current study shows that despite of the common trends in stability between studied multidomain proteins, the molecular background is different and is highly dependent on the nature, structure, and function of the protein.

Initial conformational stability studies show similar pH and NaCl effect for rTrF and mAbs. However, rTrF stability is highly related to its function, where iron release requires conformational changes resulting in the less stable partially open conformation, while mAbs stability decreases due to chemical degradation. Additionally, different mAbs have diverse stability trends in the presence of certain excipients (*e.g.* phosphate), highlighting the importance of specific interactions with the buffer components.

Contrary to rTrF and mAbs, the HSA-NEP fusion protein has more distinct stability trends. Unlike mAbs and most of the single domain proteins, increasing repulsive interactions have a negative effect. Additionally, HSA-NEP is present in multiple conformations and the equilibrium between them changes with pH. From pH 8.0, despite of the intramolecular repulsion, changes in the interaction in the interface between HSA and NEP result in new conformation, which increases overall stability of the fusion protein.

In conclusion, combination of high throughput methods with small angle X-ray scattering and molecular dynamics simulations leads to an understanding of protein stability on the molecular level. Therefore, using the same approach to study a statistically relevant number of buffer/protein combinations can provide enough information to propose a general stability model.

References

- (1) Shah, D. K. Pharmacokinetic and Pharmacodynamic Considerations for the next Generation Protein Therapeutics. *J. Pharmacokinet. Pharmacodyn.* **2015**, *42* (5), 553–571. <https://doi.org/10.1007/s10928-015-9447-8>.
- (2) Mullard, A. 2018 FDA Drug Approvals. *Nat. Rev. Drug Discov.* **2019**, *18* (2), 85–89. <https://doi.org/10.1038/d41573-019-00014-x>.
- (3) Carton, J. M.; Strohl, W. R. Protein Therapeutics (Introduction to Biopharmaceuticals). In *Introduction to Biological and Small Molecule Drug Research and Development*; Elsevier, 2013; pp 127–159. <https://doi.org/10.1016/B978-0-12-397176-0.00004-2>.
- (4) Wang, W. Instability, Stabilization, and Formulation of Liquid Protein Pharmaceuticals. *Int. J. Pharm.* **1999**, *185* (2), 129–188. [https://doi.org/10.1016/S0378-5173\(99\)00152-0](https://doi.org/10.1016/S0378-5173(99)00152-0).
- (5) PIPPI (Protein-excipient Interactions and Protein-Protein Interactions in formulation) <http://www.pippi.kemi.dtu.dk/> (accessed Jul 26, 2019).
- (6) Rader, R. A. (Re)Defining Biopharmaceutical. *Nat. Biotechnol.* **2008**, *26* (7), 743–751. <https://doi.org/10.1038/nbt0708-743>.
- (7) Carton, J. M.; Strohl, W. R. Protein Therapeutics (Introduction to Biopharmaceuticals). In *Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies*; 2013; pp 127–159. <https://doi.org/10.1016/B978-0-12-397176-0.00004-2>.
- (8) Halpern, W.; Hutto, D. Biopharmaceuticals. In *Haschek and Rousseaux's Handbook of Toxicologic Pathology*; Haschek, W., Rousseaux, C., Walling, M., Eds.; Academic Press 2013, 2013. <https://doi.org/10.1016/B978-0-12-415759-0.00025-X>.
- (9) U.S. Food and Drug Administration <https://www.fda.gov/> (accessed May 28, 2019).
- (10) Usmani, S. S.; Bedi, G.; Samuel, J. S.; Singh, S.; Kalra, S.; Kumar, P.; Ahuja, A. A.; Sharma, M.; Gautam, A.; Raghava, G. P. S. THPdb: Database of FDA-Approved Peptide and Protein Therapeutics. *PLoS One* **2017**, *12* (7), e0181748. <https://doi.org/10.1371/journal.pone.0181748>.

- (11) James, L. C.; Tawfik, D. S. Conformational Diversity and Protein Evolution – a 60-Year-Old Hypothesis Revisited. *Trends Biochem. Sci.* **2003**, *28* (7), 361–368. [https://doi.org/10.1016/S0968-0004\(03\)00135-X](https://doi.org/10.1016/S0968-0004(03)00135-X).
- (12) Mitragotri, S.; Burke, P. A.; Langer, R. Overcoming the Challenges in Administering Biopharmaceuticals: Formulation and Delivery Strategies. *Nat. Rev. Drug Discov.* **2014**, *13* (9), 655–672. <https://doi.org/10.1038/nrd4363>.
- (13) Zhang, J. Protein-Protein Interactions in Salt Solutions. In *Protein-Protein Interactions - Computational and Experimental Tools*; 2012. <https://doi.org/10.5772/38056>.
- (14) 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. *Adv. Drug Deliv. Rev.* **2011**, *63* (13), 1118–1159. <https://doi.org/10.1016/j.addr.2011.07.006>.
- (15) Xie, G.; Timasheff, S. N. The Thermodynamic Mechanism of Protein Stabilization by Trehalose. *Biophys. Chem.* **1997**, *64* (1–3), 25–43. [https://doi.org/10.1016/S0301-4622\(96\)02222-3](https://doi.org/10.1016/S0301-4622(96)02222-3).
- (16) Xie, G.; Timasheff, S. N. Mechanism of the Stabilization of Ribonuclease a by Sorbitol: Preferential Hydration Is Greater for the Denatured than for the Native Protein. *Protein Sci.* **2008**, *6* (1), 211–221. <https://doi.org/10.1002/pro.5560060123>.
- (17) Timasheff, S. N. Control of Protein Stability and Reactions by Weakly Interacting Cosolvents: The Simplicity of the Complicated; 1998; pp 355–432. [https://doi.org/10.1016/S0065-3233\(08\)60656-7](https://doi.org/10.1016/S0065-3233(08)60656-7).
- (18) Ohtake, S.; Kita, Y.; Arakawa, T. Interactions of Formulation Excipients with Proteins in Solution and in the Dried State. *Adv. Drug Deliv. Rev.* **2011**, *63* (13), 1053–1073. <https://doi.org/10.1016/j.addr.2011.06.011>.
- (19) Giancola, C.; De Sena, C.; Fessas, D.; Graziano, G.; Barone, G. DSC Studies on Bovine Serum Albumin Denaturation Effects of Ionic Strength and SDS Concentration. *Int. J. Biol. Macromol.* **1997**, *20* (3), 193–204. [https://doi.org/10.1016/S0141-8130\(97\)01159-8](https://doi.org/10.1016/S0141-8130(97)01159-8).
- (20) Kohn, W. D.; Kay, C. M.; Hodges, R. S. Salt Effects on Protein Stability: Two-Stranded α -Helical Coiled-Coils Containing Inter- or Intrahelical Ion Pairs. *J. Mol. Biol.* **1997**, *267* (4), 1039–1052. <https://doi.org/10.1006/jmbi.1997.0930>.
- (21) Zhang, Y.; Cremer, P. S. Chemistry of Hofmeister Anions and Osmolytes. *Annu. Rev. Phys. Chem.* **2010**, *61* (1), 63–83. <https://doi.org/10.1146/annurev.physchem.59.032607.093635>.
- (22) Zhang, Y.; Cremer, P. Interactions between Macromolecules and Ions: The Hofmeister Series. *Curr. Opin. Chem. Biol.* **2006**, *10* (6), 658–663. <https://doi.org/10.1016/j.cbpa.2006.09.020>.
- (23) Kristjánsson, M. M.; Kinsella, J. E. Protein and Enzyme Stability: Structural, Thermodynamic, and Experimental Aspects. In *Advances in Food and Nutrition Research*; 1991; pp 237–316. [https://doi.org/10.1016/S1043-4526\(08\)60066-2](https://doi.org/10.1016/S1043-4526(08)60066-2).

- (24) Arakawa, T.; Tsumoto, K.; Kita, Y.; Chang, B.; Ejima, D. Biotechnology Applications of Amino Acids in Protein Purification and Formulations. *Amino Acids* **2007**, *33* (4), 587–605. <https://doi.org/10.1007/s00726-007-0506-3>.
- (25) Jensen, W. A.; Armstrong, J. M.; De Giorgio, J.; Hearn, M. T. W. Stability Studies on Pig Heart Mitochondrial Malate Dehydrogenase: The Effect of Salts and Amino Acids. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1996**, *1296* (1), 23–34. [https://doi.org/10.1016/0167-4838\(96\)00049-0](https://doi.org/10.1016/0167-4838(96)00049-0).
- (26) I. Razinkov, V.; J. Treuheit, M.; W. Becker, G. Methods of High Throughput Biophysical Characterization in Biopharmaceutical Development. *Curr. Drug Discov. Technol.* **2013**, *10* (1), 59–70. <https://doi.org/10.2174/1570163811310010008>.
- (27) Senisterra, G.; Chau, I.; Vedadi, M. Thermal Denaturation Assays in Chemical Biology. *Assay Drug Dev. Technol.* **2012**, *10* (2), 128–136. <https://doi.org/10.1089/adt.2011.0390>.
- (28) Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2* (9), 2212–2221. <https://doi.org/10.1038/nprot.2007.321>.
- (29) Magnusson, A. O.; Szekrenyi, A.; Joosten, H.; Finnigan, J.; Charnock, S.; Fessner, W. NanoDSF as Screening Tool for Enzyme Libraries and Biotechnology Development. *FEBS J.* **2019**, *286* (1), 184–204. <https://doi.org/10.1111/febs.14696>.
- (30) Breitsprecher, D.; Glücklich, N.; Hawe, A.; Menzen, T. Thermal Unfolding of Antibodies Comparison of NanoDSF and MDSC for Thermal Stability Assessment during Biopharmaceutical Formulation Development. *Nanotemper Appl. Note* **2016**.
- (31) Freire, E.; Schön, A.; Hutchins, B. M.; Brown, R. K. Chemical Denaturation as a Tool in the Formulation Optimization of Biologics. *Drug Discovery Today.* 2013. <https://doi.org/10.1016/j.drudis.2013.06.005>.
- (32) Temel, D. B.; Landsman, P.; Brader, M. L. Orthogonal Methods for Characterizing the Unfolding of Therapeutic Monoclonal Antibodies: Differential Scanning Calorimetry, Isothermal Chemical Denaturation, and Intrinsic Fluorescence with Concomitant Static Light Scattering. *Methods Enzymol.* **2016**, *567*, 359–389. <https://doi.org/10.1016/bs.mie.2015.08.029>.
- (33) Schellman, J. A. Temperature, Stability, and the Hydrophobic Interaction. *Biophys. J.* **1997**, *73* (6), 2960–2964. [https://doi.org/10.1016/S0006-3495\(97\)78324-3](https://doi.org/10.1016/S0006-3495(97)78324-3).
- (34) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J. Denaturant m Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Sci.* **1995**, *4* (10), 2138–2148. <https://doi.org/10.1002/pro.5560041020>.
- (35) Clarkson, B. R.; Schön, A.; Freire, E. Conformational Stability and Self-Association Equilibrium in Biologics. *Drug Discovery Today.* 2016. <https://doi.org/10.1016/j.drudis.2015.11.007>.

- (36) Chaudhuri, B. N.; Muñoz, I. G.; Qian, S.; Urban, V. S. *Biological Small Angle Scattering: Techniques, Strategies and Tips*; Chaudhuri, B., Muñoz, I. G., Qian, S., Urban, V. S., Eds.; Advances in Experimental Medicine and Biology; Springer Singapore: Singapore, 2017; Vol. 1009. <https://doi.org/10.1007/978-981-10-6038-0>.
- (37) Svergun, D. I.; Koch, M. H. J.; Timmins, P. A.; May, R. P. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*; Oxford University Press, 2013. <https://doi.org/10.1093/acprof:oso/9780199639533.001.0001>.
- (38) Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J.* **1999**, *76* (6), 2879–2886. [https://doi.org/10.1016/S0006-3495\(99\)77443-6](https://doi.org/10.1016/S0006-3495(99)77443-6).
- (39) Franke, D.; Svergun, D. I. DAMMIF, a Program for Rapid Ab-Initio Shape Determination in Small-Angle Scattering. *J. Appl. Crystallogr.* **2009**, *42* (2), 342–346. <https://doi.org/10.1107/S0021889809000338>.
- (40) Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophys. J.* **2001**, *80* (6), 2946–2953. [https://doi.org/10.1016/S0006-3495\(01\)76260-1](https://doi.org/10.1016/S0006-3495(01)76260-1).
- (41) Svergun, D.; Barberato, C.; Koch, M. H. CRY SOL - A Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28* (6), 768–773. <https://doi.org/10.1107/S0021889895007047>.
- (42) Petoukhov, M. V.; Svergun, D. I. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys. J.* **2005**, *89* (2), 1237–1250. <https://doi.org/10.1529/biophysj.105.064154>.
- (43) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2012**, *45* (2), 342–350. <https://doi.org/10.1107/S0021889812007662>.
- (44) Cordeiro, T. N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. In *Advances in Experimental Medicine and Biology*; 2017; Vol. 1009, pp 107–129. https://doi.org/10.1007/978-981-10-6038-0_7.
- (45) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced Ensemble Modelling of Flexible Macromolecules Using X-Ray Solution Scattering. *IUCrJ* **2015**, *2*, 207–217. <https://doi.org/10.1107/S205225251500202X>.
- (46) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. PRIMUS: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2003**, *36* (5), 1277–1282. <https://doi.org/10.1107/S0021889803012779>.

- (47) Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D. MicroScale Thermophoresis: Interaction Analysis and Beyond. *J. Mol. Struct.* **2014**, *1077*, 101–113. <https://doi.org/10.1016/j.molstruc.2014.03.009>.
- (48) Zillner, K.; Jerabek-Willemsen, M.; Duhr, S.; Braun, D.; Längst, G.; Baaske, P. Microscale Thermophoresis as a Sensitive Method to Quantify Protein: Nucleic Acid Interactions in Solution. In *Methods in Molecular Biology*; 2012; pp 241–252. https://doi.org/10.1007/978-1-61779-424-7_18.
- (49) Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S. Protein-Binding Assays in Biological Liquids Using Microscale Thermophoresis. *Nat. Commun.* **2010**, *1* (1), 100. <https://doi.org/10.1038/ncomms1093>.
- (50) Jerabek-Willemsen, M.; Wienken, C. J.; Braun, D.; Baaske, P.; Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *Assay Drug Dev. Technol.* **2011**, *9* (4), 342–353. <https://doi.org/10.1089/adt.2011.0380>.
- (51) Lagassé, H. A. D.; Alexaki, A.; Simhadri, V. L.; Katagiri, N. H.; Jankowski, W.; Sauna, Z. E.; Kimchi-Sarfaty, C. Recent Advances in (Therapeutic Protein) Drug Development. *F1000Research* **2017**, *6*, 113. <https://doi.org/10.12688/f1000research.9970.1>.
- (52) Schmidt, S. R. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges*; 2013. <https://doi.org/10.1002/9781118354599>.
- (53) Andreeva, A.; Howorth, D.; Chandonia, J.-M.; Brenner, S. E.; Hubbard, T. J. P.; Chothia, C.; Murzin, A. G. Data Growth and Its Impact on the SCOP Database: New Developments. *Nucleic Acids Res.* **2007**, *36* (Database), D419–D425. <https://doi.org/10.1093/nar/gkm993>.
- (54) Fox, N. K.; Brenner, S. E.; Chandonia, J.-M. SCOPe: Structural Classification of Proteins-Extended, Integrating SCOP and ASTRAL Data and Classification of New Structures. *Nucleic Acids Res.* **2014**, *42* (Database issue), D304–9. <https://doi.org/10.1093/nar/gkt1240>.
- (55) Orengo, C.; Michie, A.; Jones, S.; Jones, D.; Swindells, M.; Thornton, J. CATH – a Hierarchic Classification of Protein Domain Structures. *Structure* **1997**, *5* (8), 1093–1109. [https://doi.org/10.1016/S0969-2126\(97\)00260-8](https://doi.org/10.1016/S0969-2126(97)00260-8).
- (56) Dengler, U.; Siddiqui, A. S.; Barton, G. J. Protein Structural Domains: Analysis of the 3dee Domains Database. *Proteins Struct. Funct. Genet.* **2001**, *42* (3), 332–344. [https://doi.org/10.1002/1097-0134\(20010215\)42:3<332::AID-PROT40>3.0.CO;2-S](https://doi.org/10.1002/1097-0134(20010215)42:3<332::AID-PROT40>3.0.CO;2-S).
- (57) Holm, L.; Sander, C. The FSSP Database of Structurally Aligned Protein Fold Families. *Nucleic Acids Res.* **1994**, *22* (17), 3600–3609. <https://doi.org/061/2>.
- (58) Gibrat, J. F.; Madej, T.; Bryant, S. H. Surprising Similarities in Structure Comparison. *Current Opinion in Structural Biology.* 1996, pp 377–385. [https://doi.org/10.1016/S0959-440X\(96\)80058-3](https://doi.org/10.1016/S0959-440X(96)80058-3).
- (59) Russell, R. B. Classification of Protein Folds. *Mol. Biotechnol.* **2007**, *36* (3), 238–247. <https://doi.org/10.1007/s12033-007-0032-2>.

- (60) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* **1995**, *247* (4), 536–540. [https://doi.org/10.1016/S0022-2836\(05\)80134-2](https://doi.org/10.1016/S0022-2836(05)80134-2).
- (61) Todd, P. A.; Brogden, R. N. Muromonab CD3. *Drugs* **1989**, *37* (6), 871–899. <https://doi.org/10.2165/00003495-198937060-00004>.
- (62) Smith, S. Ten Years of Orthoclone OKT3 (Muromonab-CD3): A Review. *J. Transpl. Coord.* **1996**, *6* (3), 109–121. <https://doi.org/10.7182/prtr.1.6.3.814513u185493182>.
- (63) Jones, T. D.; Carter, P. J.; Plückthun, A.; Vásquez, M.; Holgate, R. G. E.; Hötzel, I.; Popplewell, A. G.; Parren, P. W. H. I.; Enzelberger, M.; Rademaker, H. J.; et al. The INNs and Outs of Antibody Nonproprietary Names. *MAbs* **2016**, *8* (1), 1–9. <https://doi.org/10.1080/19420862.2015.1114320>.
- (64) Kaminski, M. S.; Tuck, M.; Estes, J.; Kolstad, A.; Ross, C. W.; Zasadny, K.; Regan, D.; Kison, P.; Fisher, S.; Kroll, S.; et al. 131 I-Tositumomab Therapy as Initial Treatment for Follicular Lymphoma. *N. Engl. J. Med.* **2005**, *352* (5), 441–449. <https://doi.org/10.1056/NEJMoa041511>.
- (65) Grillo-López, A. J. Zevalin: The First Radioimmunotherapy Approved for the Treatment of Lymphoma. *Expert Rev. Anticancer Ther.* **2002**, *2* (5), 485–493. <https://doi.org/10.1586/14737140.2.5.485>.
- (66) Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **1999**, *27* (1), 29–34. <https://doi.org/10.1093/nar/27.1.29>.
- (67) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34* (Database issue), D668–72. <https://doi.org/10.1093/nar/gkj067>.
- (68) Warr, W. A.; Thalheim, T.; Vollmer, A.; Ebert, R. U.; Kühne, R.; Schüürmann, G.; Special Libraries Association, Chemistry Division and American Chemical Society, D. of C. I.; SciFinder; Ridley, D. D. D.; Peters, M. C.; et al. SciFinder® Overview. *Sci. Technol. Libr.* **2011**, *29* (4), P4. <https://doi.org/10.1023/A:1010529114941>.
- (69) Friedman, Y. DrugPatentWatch www.drugpatentwatch.com.
- (70) Madden, T. The BLAST Sequence Analysis Tool. *NCBI Handb. [Internet]* **2002**, 1–15.
- (71) Tools, D.; Mirrors, S.; Contact, A. ExpASy Proteomics Server. *Search* **2010**, 1–2.
- (72) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>.

- (73) Porter, J. J.; Wildsmith, J.; Melm, C. D.; Schuchard, M. D.; Ray, K. M.; Chen, D. E.; Scott, G. B. I. Absolute Quantification of the Lower Abundance Proteome Through Immunoaffinity Depletion of the Twenty Most Abundant Proteins in Human Serum. *Sigma-Adrich Tech. Note* **2006**.
- (74) de Jong, G.; van Dijk, J. P.; van Eijk, H. G. The Biology of Transferrin. *Clinica Chimica Acta*. 1990, pp 1–46. [https://doi.org/10.1016/0009-8981\(90\)90278-Z](https://doi.org/10.1016/0009-8981(90)90278-Z).
- (75) Wally, J.; Halbrooks, P. J.; Vonrhein, C.; Rould, M. A.; Everse, S. J.; Mason, A. B.; Buchanan, S. K. The Crystal Structure of Iron-Free Human Serum Transferrin Provides Insight into Inter-Lobe Communication and Receptor Binding. *J. Biol. Chem.* **2006**, *281* (34), 24934–24944. <https://doi.org/10.1074/jbc.M604592200>.
- (76) Mackenzie, E. L.; Iwasaki, K.; Tsuji, Y. Intracellular Iron Transport and Storage: From Molecular Mechanisms to Health Implications. *Antioxid. Redox Signal.* **2008**, *10* (6), 997–1030. <https://doi.org/10.1089/ars.2007.1893>.
- (77) Gomme, P. T.; McCann, K. B.; Bertolini, J. Transferrin: Structure, Function and Potential Therapeutic Actions. *Drug Discov. Today* **2005**, *10* (4), 267–273. [https://doi.org/10.1016/S1359-6446\(04\)03333-1](https://doi.org/10.1016/S1359-6446(04)03333-1).
- (78) Hayashi, A.; Wada, Y.; Suzuki, T.; Shimizu, A. Studies on Familial Hypotransferrinemia: Unique Clinical Course and Molecular Pathology. *Am. J. Hum. Genet.* **1993**, *53* (1), 201–213.
- (79) Kaminski, K. A.; Bonda, T. A.; Korecki, J.; Musial, W. J. Oxidative Stress and Neutrophil Activation—the Two Keystones of Ischemia/Reperfusion Injury. *Int. J. Cardiol.* **2002**, *86* (1), 41–59. [https://doi.org/10.1016/S0167-5273\(02\)00189-4](https://doi.org/10.1016/S0167-5273(02)00189-4).
- (80) de Vries, B.; Walter, S. J.; von Bonsdorff, L.; Wolfs, T. G. A. M.; van Heurn, L. W. E.; Parkkinen, J.; Buurman, W. A. Reduction of Circulating Redox-Active Iron by Apotransferrin Protects against Renal Ischemia-Reperfusion Injury. *Transplantation* **2004**, *77* (5), 669–675. <https://doi.org/10.1097/01.TP.0000115002.28575.E7>.
- (81) Campenhout, A. Van; Van Campenhout, C. M.; Lagrou, A. R.; Manuel-y-Keenoy, B. Transferrin Modifications and Lipid Peroxidation: Implications in Diabetes Mellitus. *Free Radic. Res.* **2003**, *37* (10), 1069–1077. <https://doi.org/10.1080/10715760310001600390>.
- (82) Parkkinen, J.; Bonsdorff, L. Von; Ebeling, F.; Sahlstedt, L. Function and Therapeutic Development of Apotransferrin. *Vox Sang.* **2002**, *83* (1), 321–326. <https://doi.org/10.1111/j.1423-0410.2002.tb05327.x>.
- (83) Saavedra, S.; Sanz, G.; Jarque, I.; Moscardó, F.; Jiménez, C.; Lorenzo, I.; Martín, G.; Martínez, J.; De la Rubia, J.; Andreu, R.; et al. Early Infections in Adult Patients Undergoing Unrelated Donor Cord Blood Transplantation. *Bone Marrow Transplant.* **2002**, *30* (12), 937–943. <https://doi.org/10.1038/sj.bmt.1703764>.

- (84) Okamoto, T.; Tani, R.; Yabumoto, M.; Sakamoto, A.; Takada, K.; Sato, G. H.; Denry Sato, J. Effects of Insulin and Transferrin on the Generation of Lymphokine-Activated Killer Cells in Serum-Free Medium. *J. Immunol. Methods* **1996**, *195* (1–2), 7–14. [https://doi.org/10.1016/0022-1759\(96\)00081-6](https://doi.org/10.1016/0022-1759(96)00081-6).
- (85) Qian, Z. M. Targeted Drug Delivery via the Transferrin Receptor-Mediated Endocytosis Pathway. *Pharmacol. Rev.* **2002**, *54* (4), 561–587. <https://doi.org/10.1124/pr.54.4.561>.
- (86) Paterson, J.; Webster, C. I. Exploiting Transferrin Receptor for Delivering Drugs across the Blood-Brain Barrier. *Drug Discov. Today Technol.* **2016**, *20*, 49–52. <https://doi.org/10.1016/j.ddtec.2016.07.009>.
- (87) Shen, W.; Liu, Y.; Wang, H. Insulin-Transferrin Fusion Protein and Its Prodrug, Proinsulin-Transferrin, for Overcoming Insulin Resistance. *World Pat. WO/2019/014552* **2019**.
- (88) Yoon, J.-H.; Chang, B.-H.; Kim, S. N.; Jeong, I.-H.; Park, K. S.; Nam, K.-Y. Pharmaceutical Composition Comprising Mutant Human Growth Hormone Protein or Transferrin Fusion Protein Thereof as Effective Ingredient. *World Pat. WO/2018/004294* **2018**.
- (89) Wang, B.; Turner, Andrew, J. Anchored Transferrin Fusion Protein Libraries. *World Pat. WO/2006/138700* **2006**.
- (90) Fishman, J. B.; Rubin, J. B.; Handrahan, J. V.; Connor, J. R.; Fine, R. E. Receptor-Mediated Transcytosis of Transferrin across the Blood-Brain Barrier. *J. Neurosci. Res.* **1987**, *18* (2), 299–304. <https://doi.org/10.1002/jnr.490180206>.
- (91) Noinaj, N.; Easley, N. C.; Oke, M.; Mizuno, N.; Gumbart, J.; Boura, E.; Steere, A. N.; Zak, O.; Aisen, P.; Tajkhorshid, E.; et al. Structural Basis for Iron Piracy by Pathogenic Neisseria. *Nature* **2012**. <https://doi.org/10.1038/nature10823>.
- (92) Yang, N.; Zhang, H.; Wang, M.; Hao, Q.; Sun, H. Iron and Bismuth Bound Human Serum Transferrin Reveals a Partially-Opened Conformation in the N-Lobe. *Sci. Rep.* **2012**. <https://doi.org/10.1038/srep00999>.
- (93) Byrne, S. L.; Chasteen, N. D.; Steere, A. N.; Mason, A. B. The Unique Kinetics of Iron Release from Transferrin: The Role of Receptor, Lobe–Lobe Interactions, and Salt at Endosomal PH. *J. Mol. Biol.* **2010**, *396* (1), 130–140. <https://doi.org/10.1016/j.jmb.2009.11.023>.
- (94) Martel, P.; Kim, S. M.; Powell, B. M. Physical Characteristics of Human Transferrin from Small Angle Neutron Scattering. *Biophys. J.* **1980**, *31* (3), 371–380. [https://doi.org/10.1016/S0006-3495\(80\)85065-X](https://doi.org/10.1016/S0006-3495(80)85065-X).
- (95) Castellano, A. C.; Barteri, M.; Bianconi, A.; Borghi, E.; Cassiano, L.; Castagnola, M.; Della Longa, S. X-Ray Small Angle Scattering of the Human Transferrin Protein Aggregates. A Fractal Study. *Biophys. J.* **1993**, *64* (2), 520–524. [https://doi.org/10.1016/S0006-3495\(93\)81394-8](https://doi.org/10.1016/S0006-3495(93)81394-8).

- (96) Kilar, F.; Simon, I. The Effect of Iron Binding on the Conformation of Transferrin. A Small Angle x-Ray Scattering Study. *Biophys. J.* **1985**, *48* (5), 799–802. [https://doi.org/10.1016/S0006-3495\(85\)83838-8](https://doi.org/10.1016/S0006-3495(85)83838-8).
- (97) Grossmann, J. G.; Neu, M.; Pantos, E.; Schwab, F. J.; Evans, R. W.; Townes-Andrews, E.; Lindley, P. F.; Appel, H.; Thies, W.-G.; Hasnain, S. S. X-Ray Solution Scattering Reveals Conformational Changes upon Iron Uptake in Lactoferrin, Serum and Ovo-Transferrins. *J. Mol. Biol.* **1992**, *225* (3), 811–819. [https://doi.org/10.1016/0022-2836\(92\)90402-6](https://doi.org/10.1016/0022-2836(92)90402-6).
- (98) Mecklenburg, S. L.; Donohoe, R. J.; Olah, G. A. Tertiary Structural Changes and Iron Release from Human Serum Transferrin. *J. Mol. Biol.* **1997**, *270* (5), 739–750. <https://doi.org/10.1006/jmbi.1997.1126>.
- (99) HE, Q.-Y.; MASON, A. B.; NGUYEN, V.; MacGILLIVRAY, R. T. A.; WOODWORTH, R. C. The Chloride Effect Is Related to Anion Binding in Determining the Rate of Iron Release from the Human Transferrin N-Lobe. *Biochem. J.* **2000**, *350* (3), 909–915. <https://doi.org/10.1042/bj3500909>.
- (100) Eckenroth, B. E.; Steere, A. N.; Chasteen, N. D.; Everse, S. J.; Mason, A. B. How the Binding of Human Transferrin Primes the Transferrin Receptor Potentiating Iron Release at Endosomal PH. *Proc. Natl. Acad. Sci.* **2011**, *108* (32), 13089–13094. <https://doi.org/10.1073/pnas.1105786108>.
- (101) Abdizadeh, H.; Atilgan, A. R.; Atilgan, C. Mechanisms by Which Salt Concentration Moderates the Dynamics of Human Serum Transferrin. *J. Phys. Chem. B* **2017**, *121* (18), 4778–4789. <https://doi.org/10.1021/acs.jpcc.7b02380>.
- (102) Mason, A. B.; Halbrooks, P. J.; James, N. G.; Connolly, S. A.; Larouche, J. R.; Smith, V. C.; MacGillivray, R. T. A.; Chasteen, N. D. Mutational Analysis of C-Lobe Ligands of Human Serum Transferrin: Insights into the Mechanism of Iron Release †. *Biochemistry* **2005**, *44* (22), 8013–8021. <https://doi.org/10.1021/bi050015f>.
- (103) Bali, P. K.; Zak, O.; Aisen, P. A New Role for the Transferrin Receptor in the Release of Iron from Transferrin. *Biochemistry* **1991**, *30* (2), 324–328. <https://doi.org/10.1021/bi00216a003>.
- (104) Sipe, D. M.; Murphy, R. F. Binding to Cellular Receptors Results in Increased Iron Release from Transferrin at Mildly Acidic PH. *J. Biol. Chem.* **1991**, *266* (13), 8002–6007.
- (105) Hu, Y. B.; Dammer, E. B.; Ren, R. J.; Wang, G. The Endosomal-Lysosomal System: From Acidification and Cargo Sorting to Neurodegeneration. *Translational Neurodegeneration*. 2015. <https://doi.org/10.1186/s40035-015-0041-1>.
- (106) Blanchet, C. E.; Spilotros, A.; Schwemmer, F.; Graewert, M. A.; Kikhney, A.; Jeffries, C. M.; Franke, D.; Mark, D.; Zengerle, R.; Cipriani, F.; et al. Versatile Sample Environments and Automation for Biological Solution X-Ray Scattering Experiments at the P12 Beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **2015**, *48* (2), 431–443. <https://doi.org/10.1107/S160057671500254X>.

- (107) Round, A. R.; Franke, D.; Moritz, S.; Huchler, R.; Fritsche, M.; Malthan, D.; Klaering, R.; Svergun, D. I.; Roessle, M. Automated Sample-Changing Robot for Solution Scattering Experiments at the EMBL Hamburg SAXS Station X33. *J. Appl. Crystallogr.* **2008**, *41* (5), 913–917. <https://doi.org/10.1107/S0021889808021018>.
- (108) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank, 1999–. In *International Tables for Crystallography*; International Union of Crystallography: Chester, England, 2006; pp 675–684. <https://doi.org/10.1107/97809553602060000722>.
- (109) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H⁺⁺: A Server for Estimating PKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33* (Web Server), W368–W371. <https://doi.org/10.1093/nar/gki464>.
- (110) Indrakumar, S.; Zalar, M.; Pohl, C.; Nørgaard, A.; Streicher, W.; Harris, P.; Golovanov, A. P.; Peters, G. H. J. Conformational Stability Study of a Therapeutic Peptide Plectasin Using Molecular Dynamics Simulations in Combination with NMR. *J. Phys. Chem. B* **2019**, *123* (23), 4867–4877. <https://doi.org/10.1021/acs.jpcc.9b02370>.
- (111) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res.* **2019**, *47* (D1), D1102–D1109. <https://doi.org/10.1093/nar/gky1033>.
- (112) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52* (7), 1757–1768. <https://doi.org/10.1021/ci3001277>.
- (113) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>.
- (114) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- (115) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623–1641. <https://doi.org/10.1002/jcc.10128>.
- (116) Li, P.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10* (1), 289–297. <https://doi.org/10.1021/ct400751u>.
- (117) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg 2+ Ion Models Commonly Used in Biomolecular Simulations. *J. Comput. Chem.* **2015**, *36* (13), 970–982. <https://doi.org/10.1002/jcc.23881>.

- (118) Mongan, J.; Case, D. A.; McCammon, J. A. Constant PH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048. <https://doi.org/10.1002/jcc.20139>.
- (119) Maingi, V.; Jain, V.; Bharatam, P. V.; Maiti, P. K. Dendrimer Building Toolkit: Model Building and Characterization of Various Dendrimer Architectures. *J. Comput. Chem.* **2012**, *33* (25), 1997–2011. <https://doi.org/10.1002/jcc.23031>.
- (120) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, NA-NA. <https://doi.org/10.1002/prot.22711>.
- (121) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. <https://doi.org/10.1021/ct400341p>.
- (122) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (123) Peppel, K. A Tumor Necrosis Factor (TNF) Receptor-IgG Heavy Chain Chimeric Protein as a Bivalent Antagonist of TNF Activity. *J. Exp. Med.* **1991**, *174* (6), 1483–1489. <https://doi.org/10.1084/jem.174.6.1483>.
- (124) Sadava, D. M.; Hillis, D.; Heller, H. C.; Berenbaum, M. R. *Life, The Science Of Biology*, 9th ed.; Sadava, D. M., Ed.; Sinauer Associates, Inc. W. H. Freeman & Co., 2009.
- (125) Schmidt, S. R. Fusion Proteins as Biopharmaceuticals - Applications and Challenges. *Curr. Opin. Drug Discov. Devel.* **2009**.
- (126) Yang, H.; Liu, L.; Xu, F. The Promises and Challenges of Fusion Constructs in Protein Biochemistry and Enzymology. *Appl. Microbiol. Biotechnol.* **2016**, *100* (19), 8273–8281. <https://doi.org/10.1007/s00253-016-7795-y>.
- (127) Peters, T. The Albumin Molecule. In *All About Albumin*; Elsevier, 1995; pp 9–II. <https://doi.org/10.1016/B978-012552110-9/50004-0>.
- (128) Weimer, T.; Metzner, H. J.; Schulte, S. Recombinant Albumin Fusion Proteins. In *Fusion Protein Technologies for Biopharmaceuticals*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; pp 163–178. <https://doi.org/10.1002/9781118354599.ch10>.
- (129) Trujillo, J. M.; Nuffer, W. Albiglutide. *Ann. Pharmacother.* **2014**, *48* (11), 1494–1501. <https://doi.org/10.1177/1060028014545807>.
- (130) Lyseng-Williamson, K. A. Coagulation Factor IX (Recombinant), Albumin Fusion Protein (Albutrepenonacog Alfa; Idelvion®): A Review of Its Use in Haemophilia B. *Drugs* **2017**, *77* (1), 97–106. <https://doi.org/10.1007/s40265-016-0679-8>.
- (131) Hui, H.; Farilla, L.; Merkel, P.; Perfetti, R. The Short Half-Life of Glucagon-like Peptide-1 in Plasma Does Not Reflect Its Long-Lasting Beneficial Effects. *Eur. J. Endocrinol.* **2002**, *146* (6), 863–869. <https://doi.org/10.1530/eje.0.1460863>.

- (132) Roques, B. P.; Noble, F.; Dauge, V.; Fournié-Zaluski, M. C.; Beaumont, A. Neutral Endopeptidase 24.11: Structure, Inhibition, and Experimental and Clinical Pharmacology. *Pharmacol. Rev.* **1993**, *45* (1), 87–146.
- (133) Erdös, E. G.; Skidgel, R. A. Neutral Endopeptidase 24.11 (Enkephalinase) and Related Regulators of Peptide Hormones. *FASEB J.* **1989**, *3* (2), 145–151. <https://doi.org/10.1096/fasebj.3.2.2521610>.
- (134) Howell, S.; Nalbantoglu, J.; Crine, P. Neutral Endopeptidase Can Hydrolyze β -Amyloid(1–40) but Shows No Effect on β -Amyloid Precursor Protein Metabolism. *Peptides* **1995**, *16* (4), 647–652. [https://doi.org/10.1016/0196-9781\(95\)00021-B](https://doi.org/10.1016/0196-9781(95)00021-B).
- (135) Cordes, A. A.; Platt, C. W.; Carpenter, J. F.; Randolph, T. W. Selective Domain Stabilization as a Strategy to Reduce Fusion Protein Aggregation. *J. Pharm. Sci.* **2012**, *101* (4), 1400–1409. <https://doi.org/10.1002/jps.23049>.
- (136) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2007**, *129* (17), 5656–5664. <https://doi.org/10.1021/ja069124n>.
- (137) Ettori, C.; Righetti, P. G.; Chiesa, C.; Frigerio, F.; Galli, G.; Grandi, G. Purification of Recombinant Human Growth Hormone by Isoelectric Focusing in a Multicompartment Electrolyzer with Immobilized Membranes. *J. Biotechnol.* **1992**, *25* (3), 307–318. [https://doi.org/10.1016/0168-1656\(92\)90163-4](https://doi.org/10.1016/0168-1656(92)90163-4).
- (138) Sønderby, P.; Bukrinski, J. T.; Hebditch, M.; Peters, G. H. J.; Curtis, R. A.; Harris, P. Self-Interaction of Human Serum Albumin: A Formulation Perspective. *ACS Omega* **2018**, *3* (11), 16105–16117. <https://doi.org/10.1021/acsomega.8b02245>.
- (139) Wenskowsky, L.; Schreuder, H.; Derau, V.; Matter, H.; Volkmar, J.; Nazaré, M.; Opatz, T.; Petry, S. Identification and Characterization of a Single High-Affinity Fatty Acid Binding Site in Human Serum Albumin. *Angew. Chemie Int. Ed.* **2018**, *57* (4), 1044–1048. <https://doi.org/10.1002/anie.201710437>.
- (140) Moss, S.; Subramanian, V.; Acharya, K. R. High Resolution Crystal Structure of Substrate-Free Human Neprilysin. *J. Struct. Biol.* **2018**, *204* (1), 19–25. <https://doi.org/10.1016/j.jsb.2018.06.004>.
- (141) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
- (142) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (Web Server), W665–W667. <https://doi.org/10.1093/nar/gkh381>.
- (143) Ando, M.; Kira, H.; Hayashida, S. Changes in the Autonomic Nervous System and Moods of Advanced Cancer Patients by Mindfulness Art Therapy Short Version. *J. Cancer Ther.* **2016**, *07* (01), 13–16. <https://doi.org/10.4236/jct.2016.71002>.

- (144) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**. <https://doi.org/10.1021/ja981844+>.
- (145) Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004**, *25* (2), 238–250. <https://doi.org/10.1002/jcc.10379>.
- (146) Tan, P.-N.; Steinbach, M.; Kumar, V. Chapter 8: Cluster Analysis: Basic Concepts and Algorithms. In *Introduction to Data Mining*; 2005. [https://doi.org/10.1016/0022-4405\(81\)90007-8](https://doi.org/10.1016/0022-4405(81)90007-8).
- (147) Wolf, A.; Kirschner, K. N. Principal Component and Clustering Analysis on Molecular Dynamics Data of the Ribosomal L11·23S Subdomain. *J. Mol. Model.* **2013**, *19* (2), 539–549. <https://doi.org/10.1007/s00894-012-1563-4>.
- (148) Nelson, D. L.; Cox, M. M.; Lehninger, A. L. *Lehninger Principles of Biochemistry*, 4th ed.; Nelson, D. L., Ed.; Freeman, W. H. & Company: New York, 2004.
- (149) Sadava, D. M.; Hillis, D.; Heller, H. C.; Berenbaum, M. R. Summary for Policymakers. In *Climate Change 2013 - The Physical Science Basis*; Intergovernmental Panel on Climate Change, Ed.; Cambridge University Press: Cambridge, 2011; pp 1–30. <https://doi.org/10.1017/CBO9781107415324.004>.
- (150) Boyle, J. *Lehninger Principles of Biochemistry (4th Ed.)*: Nelson, D., and Cox, M. *Biochem. Mol. Biol. Educ.* **2005**, *33* (1), 74–75. <https://doi.org/10.1002/bmb.2005.494033010419>.
- (151) Vidarsson, G.; Dekkers, G.; Rispens, T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Front. Immunol.* **2014**, *5*. <https://doi.org/10.3389/fimmu.2014.00520>.
- (152) Flaherty, D. *Immunology for Pharmacy*; Elsevier, 2012. <https://doi.org/10.1016/C2009-0-40304-0>.
- (153) Cohenuram, M.; Saif, M. W. Panitumumab the First Fully Human Monoclonal Antibody: From the Bench to the Clinic. *Anticancer. Drugs* **2007**, *18* (1), 7–15. <https://doi.org/10.1097/CAD.0b013e32800feecb>.
- (154) Riechmann, L.; Clark, M.; Waldmann, H.; Winter, G. Reshaping Human Antibodies for Therapy. *Nature* **1988**, *332* (6162), 323–327. <https://doi.org/10.1038/332323a0>.
- (155) Parren, P. W. H. I.; Carter, P. J.; Plückthun, A. Changes to International Nonproprietary Names for Antibody Therapeutics 2017 and beyond: Of Mice, Men and More. *MAbs* **2017**, *9* (6), 898–906. <https://doi.org/10.1080/19420862.2017.1341029>.
- (156) Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. Antibody Structure, Instability, and Formulation. *J. Pharm. Sci.* **2007**, *96* (1), 1–26. <https://doi.org/10.1002/jps.20727>.
- (157) He, F.; Hogan, S.; Latypov, R. F.; Narhi, L. O.; Razinkov, V. I. High Throughput Thermostability Screening of Monoclonal Antibody Formulations. *J. Pharm. Sci.* **2010**, *99* (4), 1707–1720. <https://doi.org/10.1002/jps.21955>.

- (158) Tian, X.; Langkilde, A. E.; Thorolfsson, M.; Rasmussen, H. B.; Vestergaard, B. Small-Angle X-Ray Scattering Screening Complements Conventional Biophysical Analysis: Comparative Structural and Biophysical Analysis of Monoclonal Antibodies IgG1, IgG2, and IgG4. *J. Pharm. Sci.* **2014**, *103* (6), 1701–1710. <https://doi.org/10.1002/jps.23964>.
- (159) Lee, J. C.; Timasheff, S. N. The Stabilization of Proteins by Sucrose. *J. Biol. Chem.* **1981**, *256* (14), 7193–7201.
- (160) Timasheff, S. N. Protein-Solvent Preferential Interactions, Protein Hydration, and the Modulation of Biochemical Reactions by Solvent Components. *Proc. Natl. Acad. Sci.* **2002**, *99* (15), 9721–9726. <https://doi.org/10.1073/pnas.122225399>.
- (161) Akazawa-Ogawa, Y.; Nagai, H.; Hagihara, Y. Heat Denaturation of the Antibody, a Multi-Domain Protein. *Biophys. Rev.* **2018**, *10* (2), 255–258. <https://doi.org/10.1007/s12551-017-0361-8>.
- (162) Ionescu, R. M.; Vlasak, J.; Price, C.; Kirchmeier, M. Contribution of Variable Domains to the Stability of Humanized IgG1 Monoclonal Antibodies. *J. Pharm. Sci.* **2008**, *97* (4), 1414–1426. <https://doi.org/10.1002/jps.21104>.
- (163) Garber, E.; Demarest, S. J. A Broad Range of Fab Stabilities within a Host of Therapeutic IgGs. *Biochem. Biophys. Res. Commun.* **2007**, *355* (3), 751–757. <https://doi.org/10.1016/j.bbrc.2007.02.042>.
- (164) Gentiluomo, L.; Svilenov, H.; Augustijn, D.; El Bialy, I.; Greco, M. L.; Kulakova, A.; Indrakumar, S.; Mahapatra, S.; Morales, M.; Pohl, C.; et al. Boosting Therapeutic Protein Development by Publicly Available Datasets Including Comprehensive Computational and Biophysical Characterization. **2019**. (Manuscript submitted)
- (165) Finnis, C. J.; Payne, T.; Hay, J.; Dodsworth, N.; Wilkinson, D.; Morton, P.; Saxton, M. J.; Tooth, D. J.; Evans, R. W.; Goldenberg, H.; et al. High-Level Production of Animal-Free Recombinant Transferrin from *Saccharomyces Cerevisiae*. *Microb. Cell Fact.* **2010**, *9* (1), 87. <https://doi.org/10.1186/1475-2859-9-87>.
- (166) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; et al. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *J. Appl. Crystallogr.* **2017**, *50* (4), 1212–1225. <https://doi.org/10.1107/S1600576717007786>.
- (167) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3784–3788. <https://doi.org/10.1093/nar/gkg563>.

Appendix A

All tables and figures are numbered according to the corresponding chapter: tables and figures S4 correspond to the chapter 4 (structural classification of protein therapeutics), S5 to chapter 5 (transferrin), S6 to chapter 6 (albumin-neprilysin fusion protein), and S7 to chapter 7 (monoclonal antibodies).

Table S4.1: Structural classification of approved peptide- and protein-drugs according SCOP database, include class, fold, superfamily and family. Additional information, *e.g.* targets, bioactivity indicators, disease, CAS number, associated PDB files and URL address to relevant databases (*e.g.* SCOP, DrugBank, KEGG, Drug, and FDA) available, but not shown.

Year	Active ingredient	Drug	Class	Fold	Superfamily	Family	Marketing status
1974	Gonadotropin	Chorionic gonadotropin	g	g.17	g.17.1	g.17.1.4	Prescription
1989	Epoietin alfa	EPOGEN/PROCRIT	a	a.26	a.26.1	a.26.1.2	Prescription
1989	Interferon alfa-N3	Alferon N Injection	a	a.26	a.26.1	a.26.1.3	Prescription
1989	Interferon alfa-2b	Intron A	a	a.26	a.26.1	a.26.1.3	Prescription
1989	Human Serum Albumin	Flexbumin	a	a.126	a.126.1	a.126.1.1	Prescription
1990	Pegademase bovine	Adagen	c	c.1	c.1.9	c.1.9.1	Prescription
1974	Gonadotropin	Chorionic gonadotropin	g	g.17	g.17.1	g.17.1.4	Prescription
1991	Sargramostim	Leukine	a	a.26	a.26.1	a.26.1.2	Prescription
1991	Onabotulinumtoxin A	BOTOX	Complex active ingredient				Prescription
1991	Filgrastim	Neupogen	a	a.26	a.26.1	a.26.1.1	Prescription
1992	Aldesleukin	Proleukin	a	a.26	a.26.1	a.26.1.2	Prescription
1992	Muromonab-CD3	Orthoclone OKT3	b	b.1	b.1.1	AB	Prescription
1993	Dronase alfa	Pulmozyme	d	d.151	d.151.1	d.151.1.1	Prescription
1993	Interferon beta-1b	BETASERON	a	a.26	a.26.1	a.26.1.3	Prescription
1993	Abciximab	ReoPro	b	b.1	b.1.1	BD	Prescription
1994	Asparaginase Escherichia coli	ELSPAR	c	c.88	c.88.1	c.88.1.1	Prescription
1994	Imiglucerase	Cerezyme	Complex active ingredient				Prescription
1994	Pegaspargase	Oncaspar	c	c.88	c.88.1	c.88.1.1	Prescription
1995	Somatotropin recombinant	Genotropin	a	a.26	a.26.1	a.26.1.1	Prescription
1996	Insulin lispro	Humalog	g	g.1	g.1.1	g.1.1.1	Prescription
1996	Interferon beta-1a	AVONEX	a	a.26	a.26.1	a.26.1.3	Prescription

1996	Alteplase	Activase		Complex active ingredient			Prescription
1996	Retepase	Retavase		Complex active ingredient			Prescription
1996	Insulin susp protamine zinc	Protamine zinc and iletin I	g	g.1	g.1.1	g.1.1.1	Prescription
1997	Rituximab	Rituxan	b	b.1	b.1.1	AB	Prescription
1997	Becaplermin	Regranex	g	g.17	g.17.1	g.17.1.1	Prescription
1997	Coagulation Factor IX	BeneFIX		Complex active ingredient			Prescription
1997	Daclizumab	Zenapax	b	b.1	b.1.1	ABD	Prescription
1997	Oprelvekin	Neumega		Not classified in scope			Prescription
1998	Sacrosidase	Sucraid		Not classified in scope			Prescription
1998	Basiliximab	Simulect	b	b.1	b.1.1	AB	Prescription
1998	Palivizumab	Synagis	b	b.1	b.1.1	A	Prescription
1998	Etanercept	Enbrel	b	b.1	b.1.1	B	Prescription
1998	Infliximab	Remicade	b	b.1	b.1.1	no sequence	Prescription
1998	Glucagon recombinant	Glucagon	j	j.6	j.6.1	j.6.1.1	Prescription
1998	Thyrotropin alfa	Thyrogen	g	g.17	g.17.1	g.17.1.4	Prescription
1999	Trastuzumab	Herceptin	b	b.1	b.1.1	AB	Prescription
1999	Interferon gamma- 1b	Actimmune	a	a.26	a.26.1	a.26.1.3	Prescription
1999	Antihemophilic factor	ReFacto		Not classified in scope			Prescription
1999	Denileukin diftitox	Ontak		Complex active ingredient			Prescription
2000	Botulinum Toxin Type B	Myobloc		Complex active ingredient			Prescription
2000	Insulin aspart	NovoLog	g	g.1	g.1.1	g.1.1.1	Prescription
2000	Choriogonadotropin alfa	Ovidrel	g	g.17	g.17.1	g.17.1.4	Prescription
2000	Tenecteplase	TNKase		Complex active ingredient			Prescription
2001	Darbepoetin alfa	Aranesp	a	a.26	a.26.1	a.26.1.2	Prescription
2001	Peginterferon alfa- 2b	PegIntron	a	a.26	a.26.1	a.26.1.3	Prescription
2001	Nesiritide Recombinant	Natreacor		Not classified in scope			Prescription
2001	Alemtuzumab	Campath	b	b.1	b.1.1	BD	Prescription
2001	Anakinra	Kineret	b	b.42	b.42.1	b.42.1.2	Prescription
2001	Drotrecogin alfa	Xigris		Complex active ingredient			Prescription
2002	Peginterferon alfa- 2a	Pegasys	a	a.26	a.26.1	a.26.1.3	Prescription
2002	Rasburicase	Elitek	d	d.96	d.96.1	d.96.1.4	Prescription
2002	Ibritumomab tiuxetan	Zevalin	b	b.1	b.1.1	ABD	Prescription
2002	Pegfilgrastim	Neulasta	a	a.26	a.26.1	a.26.1.1	Prescription
2002	Teriparatide recombinant human	Forteo	j	j.15	j.15.1	j.15.1.1	Prescription
2002	Adalimumab	Humira	b	b.1	b.1.1	ABD	Prescription
2003	Pegvisomant	Somavert	a	a.26	a.26.1	a.26.1.1	Prescription
2003	Omalizumab	Xolair	b	b.1	b.1.1	ABD	Prescription
2003	Agalsidase beta	Fabrazyme		Complex active ingredient			Prescription
2003	Enfuvirtide	FUZEON	h	h.3	h.3.2	h.3.2.1	Prescription
2003	Laronidase	Aldurazyme		Not classified in scope			Prescription
2003	Tositumomab	Bexxar	b	b.1	b.1.1	ABD	Prescription
2003	Alefacept	Amevive	b	b.1	b.1.1	AB	Prescription
2004	Bevacizumab	Avastin	b	b.1	b.1.1	ABD	Prescription

2004	Cetuximab	Erbix	b	b.1	b.1.1	ABD	Prescription
2004	Insulin glulisine	Apidra	g	g.1	g.1.1	g.1.1.1	Prescription
2004	Follitropin beta	Follistim AQ	g	g.17	g.17.1	g.17.1.4	Prescription
2004	Menotropins (FSH;LH)	MENOPUR	g	g.17	g.17.1	g.17.1.4	Prescription
2004	Immune Globulin Human	Octagam	b	b.1	b.1.1	AB	Prescription
2004	Natalizumab	Tysabri	b	b.1	b.1.1	AB	Prescription
2004	Palifermin	Kepivance	b	b.42	b.42.1	b.42.1.1	Prescription
2005	Pramlintide acetate	Symlin		Not classified in scope			Prescription
2005	Exenatide	BYETTA	j	j.6	j.6.1	j.6.1.1	Prescription
2005	Insulin detemir	Levemir	g	g.1	g.1.1	g.1.1.1	Prescription
2005	Mecasermin recombinant	Increlex	g	g.1	g.1.1	g.1.1.1	Prescription
2005	Galsulfase	Naglazyme	c	c.76	c.76.1	c.76.1.2	Prescription
2005	Calcitonin Salmon	Fortical	j	j.6	j.6.1	j.6.1.1	Prescription
2005	Hyaluronidase	HYLENEX		Not classified in scope			Prescription
2005	Abatacept	Orencia	b	b.1	b.1.1	AB	Prescription
2006	Idursulfase	Elaprase		Not classified in scope			Prescription
2006	Ranibizumab	Lucentis	b	b.1	b.1.1	AB	Prescription
2006	Panitumumab	Vectibix	b	b.1	b.1.1	ABD	Prescription
2007	Eculizumab	Soliris	b	b.1	b.1.1	ABD	Prescription
2008	Romiplostim	Nplate	b	b.1	b.1.1	B	Prescription
2008	Certolizumab pegol	Cimzia	b	b.1	b.1.1	ABD	Prescription
2008	Riloncept	Arcalyst	b	b.1	b.1.1	BD	Prescription
2009	Ecaltantide	Kalbitor	g	g.8	g.8.1	g.8.1.1	Prescription
2009	Abobotulinumtoxin A	Dysport		Complex active ingredient			Prescription
2009	Ofatumumab	Arzerra	b	b.1	b.1.1	ABD	Prescription
2009	Canakinumab	Ilaris	b	b.1	b.1.1	AB	Prescription
2009	Pancrealipase	Pancreaze		Complex active ingredient			Prescription
2009	Ustekinumab	Stelara	b	b.1	b.1.1	ABD	Prescription
2009	Golimimumab	Simponi	b	b.1	b.1.1	no sequence	Prescription
2010	Incobotulinumtoxin A	Xeomin		Complex active ingredient			Prescription
2010	Collagenase Clostridium Histolyticum	XIAFLEX		Not classified in scope			Prescription
2010	Liraglutide	Victoza	j	j.6	j.6.1	j.6.1.1	Prescription
2010	Pegloticase	Krystexxa	d	d.96	d.96.1	d.96.1.4	Prescription
2010	Velaglycerase alfa	VPRIV		Complex active ingredient			Prescription
2010	Denosumab	XGEVA	b	b.1	b.1.1	AB	Prescription
2010	Alglucosidase alfa	Lumizyme		Complex active ingredient			Prescription
2010	Tocilizumab	Actemra	b	b.1	b.1.1	ABD	Prescription
2011	Brentuximab vedotin	Adcetris	b	b.1	b.1.1	no sequence	Prescription
2011	Belimumab	Benlysta	b	b.1	b.1.1	no sequence	Prescription
2011	Asparaginase Erwinia chrysanthemi	Erwinaze	c	c.88	c.88.1	c.88.1.1	Prescription
2011	Ipilimumab	Yervoy	b	b.1	b.1.1	AB	Prescription
2011	Aflibercept	Eylea	b	b.1	b.1.1	BC	Prescription
2011	Belatacept	Nulojix	b	b.1	b.1.1	AB	Prescription
2012	Taliglucerase alfa	Elelyso		Complex active ingredient			Prescription

2012	Coagulation Factor VIIa	NovoSeven		Complex active ingredient			Prescription
2012	Pertuzumab	Perjeta	b	b.1	b.1.1	AB	Prescription
2012	Raxibacumab	Raxibacumab	b	b.1	b.1.1	no sequence	Prescription
2012	Glucarpidase	Voraxaze		Complex active ingredient			Prescription
2012	Ocriplasmin	Jetrea	b	b.47	b.47.1	b.47.1.2	Prescription
2013	Ado-Trastuzumab Emtansine	Kadcyla	b	b.1	b.1.1	no sequence	Prescription
2013	Obinutuzumab	Gazyva	b	b.1	b.1.1	ABD	Prescription
2014	Vedolizumab	Entyvio	b	b.1	b.1.1	AB	Prescription
2014	Ramucirumab	Cyramza	b	b.1	b.1.1	ABD	Prescription
2014	Metreleptin	Myalept	a	a.26	a.26.1	a.26.1.1	Prescription
2014	Blinatumomab	Blinicyto	b	b.1	b.1.1	AD	Prescription
2014	Pembrolizumab	Keytruda	b	b.1	b.1.1	ABD	Prescription
2014	Siltuximab	Sylvant	b	b.1	b.1.1	AB	Prescription
2014	Peginterferon beta-1a	Plegridy	a	a.26	a.26.1	a.26.1.3	Prescription
2014	Elosulfase alfa	Vimizim		Not classified in scope			Prescription
2015	Alirocumab	Praluent	b	b.1	b.1.1	AB	Prescription
2015	Insulin glargine	Basaglar	g	g.1	g.1.1	g.1.1.1	Prescription
2015	Insulin aspart/degludec	Ryzodeg (70/30)	g	g.1	g.1.1	g.1.1.1	Prescription
2015	Insulin degludec	Tresiba	g	g.1	g.1.1	g.1.1.1	Prescription
2015	Evolocumab	REPATHA	b	b.1	b.1.1	ABD	Prescription
2015	Asfotase alfa	Strensiq		Complex active ingredient			Prescription
2015	Sebelipase alfa	Kanuma	c	c.69	c.69.1	c.69.1.6	Prescription
2015	Daratumumab	Darzalex	b	b.1	b.1.1	AB	Prescription
2015	Elotuzumab	Empliciti	b	b.1	b.1.1	AB	Prescription
2015	Dinutuximab	Unituxin	b	b.1	b.1.1	AB	Prescription
2015	Necitumumab	Portrazza	b	b.1	b.1.1	ABD	Prescription
2015	Secukinumab	Cosentyx	b	b.1	b.1.1	AB	Prescription
2015	Idarucizumab	Praxbind	b	b.1	b.1.1	ABD	Prescription
2015	Mepolizumab	NUCALA	b	b.1	b.1.1	ABD	Prescription
2015	Nivolumab	Opdivo	b	b.1	b.1.1	AB	Prescription
2016	Obiltoximab	Anthim	b	b.1	b.1.1	ABD	Prescription
2016	Ixekizumab	Taltz	b	b.1	b.1.1	AB	Prescription
2016	Reslizumab	Cinqair	b	b.1	b.1.1	no sequence	Prescription
2016	Atezolizumab	Tecentriq	b	b.1	b.1.1	ABD	Prescription
2016	Nusinersen	Spinraza		No sequence			Prescription
2016	Bezlotoxumab	Zinplava	b	b.1	b.1.1	No sequence	Prescription
2016	Olaratumab	Lartruvo	b	b.1	b.1.1	AB	Prescription
2017	Semaglutide	Ozempic	j	j.6	j.6.1	j.6.1.1	Prescription
2017	Emizumab	Hemlibra	b	b.1	b.1.1	ABD	Prescription
2017	Vestronidase alfa	MEPSEVII					Prescription
2017	Benralizumab	FASENRA	b	b.1	b.1.1	ABD	Prescription
2017	Inotuzumab ozogamicin	Besponsa		No sequence			Prescription
2017	Guselkumab	TREMFYA	b	b.1	b.1.1	ABD	Prescription
2017	Sarilumab	KEVZARA	b	b.1	b.1.1	ABD	Prescription
2017	Durvalumab	IMFINZI	b	b.1	b.1.1	AB	Prescription
2017	Abaloparatide	TYMLOS	j	j.15	j.15.1	j.15.1.1	Prescription
2017	Cerliponase alfa	BRINEURA	c	c.41	c.41.1	c41.1.2	Prescription
2017	Ocrelizumab	OCREVUS		No sequence			Prescription

2017	Dupilumab	DUPIXENT			No sequence		Prescription
2017	Avelumab	BAVENCIO	b	b.1	b.1.1	ABD	Prescription
2017	Brodalumab	SILIQ			No sequence		Prescription
	OspA lipoprotein	LYMerix	b	b.76	b.76.1	b.76.1.1	Prescription
1966	Insulin Pork	Iletin I	g	g.1	g.1.1	g.1.1.1	Discontinued
1979	Insulin Purified Pork	Iletin II	g	g.1	g.1.1	g.1.1.1	Discontinued
1980	Insulin zinc susp purified pork	Lente	g	g.1	g.1.1	g.1.1.1	Discontinued
1980	Insulin Purified beef	Regular Iletin II	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin SUSP Isophane Semisynthetic purified pork	Insulin Insulatard NPH Nordisk	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin purified pork, insulin susp isophane purified pork	Insulin Nordisk Mixtard	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin Zinc SUSP purified beef/pork	Lentard	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin zinc susp purified beef	Lente	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin zinc susp beef	Lente Iletin II	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin zinc susp pork	Lente Iletin II (PORK)	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	insulin susp isophane beef/pork	NPH Iletin I (beef-pork)	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin zinc SUSP Prompt purified pork	Semilente	g	g.1	g.1.1	g.1.1.1	Discontinued
1982	Insulin zinc susp extnended purified beef	Ultralente	g	g.1	g.1.1	g.1.1.1	Discontinued
1986	Insulin susp isophane	Insulatard NPH human	g	g.1	g.1.1	g.1.1.1	Discontinued
1990	Sermorelin acetate	Geref			Not classified in scope		Discontinued
1991	Alglucerase	Ceredase			Complex active ingredient		Discontinued
1992	Interferin alfa-2a	Roferon A	a	a.26	a.26.1	a.26.1.3	Discontinued
1993	Aprotinin	Trasylol	g	g.8	g.8.1	g.8.1.1	Discontinued
1994	Streptokinase	Streptase	d	d.15	d.15.5	d.15.5.1	Discontinued
1997	Interferon alfacon-1	INFERGEN	a	a.26	a.26.1	a.26.1.3	Discontinued
1998	Lepirudin	Refludan	g	g.3	g.3.15	g.3.15.2	Discontinued
2000	Gemtuzumab ozogamicin	Mylotarg	b	b.1	b.1.1	AD	Discontinued
2003	Efalizumab	Raptiva	b	b.1	b.1.1	BD	Discontinued
2004	Lutropin alfa	Luveris	g	g.17	g.17.1	g.17.1.4	Discontinued
2006	Insulin recombinant human	Exubera	g	g.1	g.1.1	g.1.1.1	Discontinued
2012	Lucinactant	Surfaxin			Not classified in scope		Discontinued

Note: A- b.1.1.1; B – b.1.1.2; C – b.1.1.3; D – b.1.1.4; Complex active ingredient includes multiple proteins that belong to different Family.

Table S4.2: Classification of protein-based drugs identified with complex active ingredient

Drug	Active ingredient	Class	Fold	Superfamily	Family
Pancrealipase	Chymotrypsin B	b	b.47	b.47.1	b.47.1.2
	Pancreatic alpha amylase	b	b.71	b.71.1	b.71.1.1
	Pancreatic alpha amylase	c	c.1	c.1.8	c.1.8.1
	Pancreatic triacylglycerol lipase	c	c.69	c.69.1	c.69.1.19
Dysport, Xeomin, BOTOX	Botulinum toxin A	d	d.92	d.92.1	d.92.1.7
		b	b.29	b.29.1	b.29.1.6
		b	b.42	b.42.4	b.42.4.2
		h	h.4	h.4.2	h.4.2.1
Ontak	Denileukin diftitox	a	a.26	a.26.1	a.26.1.2
		d	d.166	d.166.1	d.166.1.1
		f	f.1	f.1.2	f.1.2.1
Fabrazyme	Agalsidase beta	c	c.1	c.1.8	c.1.8.1
		b	b.71	b.71.1	b.71.1.1
Lumizyme	Alglucosidase alfa	b	b.71	b.71.1	b.71.1.1
		c	c.1	c.1.8	c.1.8.3
		b	b.30	b.30.5	b.30.5.11
		c	c.1	c.1.8	c.1.8.13
		b	b.71	b.71.1	b.71.1.4
		b	b.150	b.150.1	b.150.1.1
Activase	Alteplase	g	g.27	g.27.1	g.27.1.1
		g	g.3	g.3.11	g.3.11.1
		g	g.14	g.14.1	g.14.1.1
		b	b.47	b.47.1	b.47.1.2
Myobloc	Botulinum Toxin Type B	d	d.92	d.92.1	d.92.1.7
		h	h.4	h.4.2	h.4.2.1
		b	b.29	b.29.1	b.29.1.6
BeneFIX	Coagulation Factor IX	b	b.42	b.42.4	b.42.4.2
		g	g.3	g.3.11	g.3.11.1
NovoSeven	Coagulation Factor VIIa	b	b.47	b.47.1	b.47.1.2
		g	g.32	g.32.1	g.32.1.1
		g	g.3	g.3.11	g.3.11.1
Xigris	Drotrecogin alfa	b	b.47	b.47.1	b.47.1.2
		g	g.3	g.3.11	g.3.11.1
Voraxaze	Glucarpidase	c	c.56	c.56.5	c.56.5.4
		d	d.58	d.58.19	d.58.19.1
Cerezyme	Imiglucerase	b	b.71	b.71.1	b.71.1.2
		c	c.1	c.1.8	c.1.8.3
Retavase	Retepase	b	b.47	b.47.1	b.47.1.2
		g	g.14	g.14.1	g.14.1.1
Elelyso	Taliglucerase alfa	c	c.1	c.1.8	c.1.8.3
		b	b.71	b.71.1	b.71.1.2
TNKase	Tenecteplase	g	g.27	g.27.1	g.27.1.1
		g	g.3	g.3.11	g.3.11.1
		g	g.14	g.14.1	g.14.1.1
		b	b.47	b.47.1	b.47.1.2

VPRIV	Velaglucerase alfa	b	b.71	b.71.1	b.71.1.2
		c	c.1	c.1.8	c.1.8.3
Ceredase	Alglucerase	c	c.1	c.1.8	c.1.8.3
		b	b.71	b.71.1	b.71.1.2
Strensiq	Asfotase alfa	b	b.1	b.1.1	b.1.1.2
		c	c.76	c.76.1	c.76.1.1

Appendix B

Table S5.1: SAXS data collection – overview of the different physicochemical conditions tested.

Buffer	pH	Additives	$c_{\text{additives}}$ (mM)	c_{TrF} (g/L)
10 mM histidine	5.0	-	-	1.0, 2.0, 4.9, 6.8 and 9.4
		NaCl	0, 35, 70 and 140	1.0 5.69, 5.25, 5.28 and 5.17
		Arginine-HCl	0, 35, 70 and 140	3.5, 3.6, 3.7 and 3.7
		GuHCl	0, 1000, 1500 and 2000	1.0, 1.0, 1.0 and 1.0
	6.5	-	-	1.0, 2.0, 5.0, 7.0 and 10.0
		NaCl	0, 35, 70 and 140	1.0, 1.0, 1.0 and 0.9 5.53, 5.20, 4.98 and 5.02
10 mM acetate	4.0	-	-	0.69, 1.87, 4.80, 6.73 and 9.65
	5.0	-	-	1.8, 3.5, 8.5, 12.0 and 17.5
		Arginine-HCl	0, 35, 70 and 140	1.5, 1.3, 1.2 and 1.1
10 mM tris	8.0	-	-	1.0, 2.0, 4.7, 6.4 and 9.0
		NaCl	0, 35, 70 and 140	1.0, 1.0, 0.9 and 1.0

Table S5.2: SAXS - experimental details.

a) Sample details				
	Recombinant Transferrin ¹⁶⁵			
Organism	<i>Saccharomyces cerevisiae</i>			
Source	Albumedix Ltd			
Extinction coefficient (A_{280} , $M^{-1}cm^{-1}$)	85115			
Molecular weight from chemical composition (kDa)	75.135			
b) SAXS data collection				
Instrument	P12 BioSAXS beamline (PETRAIII)			BM29 BioSAXS
Date	12.17	07.18	12.18	12.17
Detector	PILATUS2M	PILATUS6M		PILATUS1M
Wavelength (\AA)		0.124402		0.99
Beam size (mm^2)	0.2 \times 0.12			0.7 \times 0.7
Detector distance (m)	3.000			2.867
q -measurement range (nm^{-1})	0.027-5.078	0.026 -7.288	0.0261-7.263	0.039-4.9335
Absolute scaling method	Comparison with scattering from pure H_2O		Comparison with scattering from BSA	Comparison with scattering from pure H_2O
Normalization	To transmitted intensity by beam-stop counter			
Monitoring for radiation damage	Frame-by-frame comparison			
Exposure time (s)	20 \times 0.05		30 \times 0.095	10 \times 1.00
Sample configuration	Quartz glass capillary			
Sample temperature ($^{\circ}C$)	20			
c) Software employed for SAXS data reduction, analysis and interpretation				
SAS data reduction	<i>PRIMUSqt</i> ⁴⁶ from <i>ATSAS</i> 2.8.3 ¹⁶⁶			
Extinction coefficient estimate	<i>ExPaSy</i> ¹⁶⁷			
Basic analyses: Guinier, $p(r)$, V_P	<i>PRIMUSqt</i> ⁴⁶			
Volume fractions of components	<i>OLIGOMER</i> ⁴⁶			
Molecular graphics	<i>PyMOL</i> (version 1.8.2.3, Schrödinger, LLC)			

Table S5.3: Structural parameters derived from SAXS experiments.

	C _r TrF (g/L)	C _{NaCl} (mM)	C _{arginine} (mM)	Guinier		<i>p</i> (<i>r</i>)			Apparent MW (kDa)	
				<i>I</i> (0)/ <i>c</i>	<i>R</i> _g (nm)	<i>I</i> (0)/ <i>c</i>	<i>R</i> _g (nm)	<i>D</i> _{max} (nm)	Guinier	<i>p</i> (<i>r</i>)
10 mM acetate pH 4.0	1.00	-	-	-	-	14310*	3.67	13.60	-	96
	2.10	-	-	-	-	13620*	3.46	12.00	-	91
	4.80	-	-	-	-	13500*	3.40	12.00	-	90
	6.69	-	-	-	-	13090*	3.30	11.57	-	87
	9.72	-	-	-	-	12500*	3.19	10.89	-	83
10 mM acetate pH 5.0	1.76	-	-	0.060	3.52	0.060	3.54	11.36	83	83
	3.53	-	-	0.075	3.55	0.060	3.57	11.90	90	83
	8.46	-	-	0.062	3.37	0.060	3.42	11.00	86	83
	11.96	-	-	0.061	3.34	0.060	3.39	11.00	84	83
	17.48	-	-	0.060	3.27	0.060	3.33	10.70	83	83
	1.49	-	0	0.064	3.45	0.060	3.53	12.10	89	83
	1.32	-	35	0.070	3.86	0.080	3.94	13.13	109	111
	1.22	-	70	0.083	3.99	0.080	4.02	13.00	115	111
	1.05	-	140	0.084	3.96	0.080	3.95	12.28	116	111
10 mM histidine pH 5.0	1.00	-	-	0.063	3.36	0.060	3.29	9.68	87	83
	1.98	-	-	0.064	3.41	0.060	3.44	11.27	89	83
	4.92	-	-	0.066	3.51	0.060	3.55	12.30	91	83
	6.76	-	-	0.066	3.50	0.060	3.65	13.87	91	83
	9.39	-	-	0.066	3.56	0.060	3.67	14.09	91	83
	1.66	-	0	0.063	3.45	0.060	3.53	12.00	87	83
	1.60	-	35	0.063	3.52	0.060	3.56	11.53	87	83
	1.42	-	70	0.064	3.61	0.060	3.65	11.90	89	83
	1.15	-	140	0.065	3.70	0.070	3.69	11.50	90	97
	5.53	0	-	11976*	3.53	12070*	3.67	14.57	80	81
	5.50	35	-	12919*	3.72	12980*	3.85	14.91	86	87
	5.26	70	-	13824*	3.95	13880*	4.09	16.70	92	93
	5.07	140	-	14931*	4.05	15160*	4.31	17.00	100	101
10 mM histidine pH 6.5	1.00	-	-	62**	3.37	63**	3.40	11.00	69	70
	2.00	-	-	64**	3.35	65**	3.41	11.20	71	72
	5.00	-	-	66**	3.20	67**	3.29	11.52	73	74
	7.00	-	-	66**	3.19	67**	3.29	11.59	73	74
	10.00	-	-	64**	3.14	66**	3.27	11.56	71	73
	5.26	0	-	11279*	3.29	11410*	3.41	12.60	75	76
	5.20	35	-	11592*	3.47	11670*	3.61	14.40	77	78
	5.20	70	-	11288*	3.37	11590*	3.56	14.00	75	77
	4.92	140	-	11703*	3.58	11719*	3.65	15.00	78	78
	10 mM tris pH 8.0	1.00	-	-	0.059	3.54	0.060	3.47	11.00	82
1.95		-	-	0.057	3.40	0.060	3.41	11.00	79	83
4.71		-	-	0.055	3.22	0.050	3.30	11.65	76	69
6.37		-	-	0.054	3.18	0.050	3.26	11.00	75	69
8.98		-	-	0.053	3.13	0.050	3.24	10.93	73	69

Note: *DESY – BSA calibration; ** - data collected in ESRF

Table S5.4: OLIGOMER data analysis.

	C_rTrF (g/L)	C_{NaCl} (mM)	C_{arginine} (mM)	Open	Partially open	Closed	Dimer	χ²
10 mM acetate pH 4.0	1.00	-	-	0.04 (±0.04)	0.60 (±0.05)	0.15 (±0.02)	0.201 (±0.006)	1.05
	2.10	-	-	0.08 (±0.02)	0.62 (±0.03)	0.12 (±0.01)	0.178 (±0.004)	1.04
	4.80	-	-	0.122 (±0.009)	0.59 (±0.01)	0.116 (±0.006)	0.176 (±0.002)	1.21
	6.69	-	-	0.156 (±0.007)	0.550 (±0.009)	0.139 (±0.005)	0.156 (±0.002)	1.27
	9.72	-	-	0.19 (±0.01)	0.52 (±0.02)	0.178 (±0.009)	0.109 (±0.003)	1.09
10 mM acetate pH 5.0	1.76	-	-	-	0.38 (± 0.03)	0.41 (± 0.04)	0.236 (± 0.005)	0.98
	3.53	-	-	-	0.34 (± 0.02)	0.43 (± 0.01)	0.230 (± 0.004)	1.01
	8.56	-	-	0.01 (± 0.02)	0.30 (± 0.03)	0.50 (± 0.02)	0.186 (± 0.004)	1.03
	11.96	-	-	0.007 (± 0.017)	0.31 (± 0.02)	0.51 (± 0.01)	0.173 (± 0.003)	0.96
	17.48	-	-	0.06 (± 0.01)	0.23 (± 0.02)	0.558 (± 0.010)	0.150 (± 0.002)	1.03
	1.49	-	0	-	0.38 (±0.05)	0.41 (±0.06)	0.207 (±0.006)	0.99
	1.32	-	35	-	0.57 (±0.03)	-	0.434 (±0.003)	1.27
	1.22	-	70	0.04 (±0.004)	0.48 (±0.04)	-	0.480 (±0.004)	1.42
	1.05	-	140	-	0.489 (±0.005)	-	0.511 (±0.004)	1.62
	10 mM histidine pH 5.0	1.00	-	-	-	0.52 (±0.02)	0.28 (±0.02)	0.206 (±0.002)
1.98		-	-	-	0.68 (±0.01)	0.121 (±0.014)	0.199 (±0.002)	1.21
4.92		-	-	-	0.551 (±0.006)	0.234 (±0.006)	0.215 (±0.001)	1.52
6.76		-	-	-	0.519 (±0.005)	0.258 (±0.005)	0.223 (±0.001)	1.68
9.39		-	-	-	0.509 (±0.006)	0.261 (±0.004)	0.229 (±0.001)	2.41
5.53		0	-	-	0.578 (±0.004)	0.201 (±0.004)	0.221 (±0.001)	2.77
5.50		35	-	-	0.637 (±0.004)	0.073 (±0.005)	0.290 (±0.001)	4.65
5.26		70	-	-	0.644 (±0.005)	0.003 (±0.005)	0.353 (±0.001)	6.16
5.07		140	-	-	0.546 (±0.001)	-	0.454 (±0.001)	8.00

	1.66	-	0	-	0.52 (±0.02)	0.28 (±0.02)	0.206 (±0.002)	1.15
	1.69	-	35	-	0.746 (±0.003)	-	0.254 (±0.002)	1.16
	1.62	-	70	-	0.684 (±0.004)	-	0.316 (±0.003)	1.05
	1.64	-	140	-	0.612 (±0.005)	-	0.388 (±0.004)	1.66
10 mM histidine pH 6.5	1.00	-	-	-	-	0.829 (±0.003)	0.171 (±0.002)	1.13
	2.00	-	-	-	-	0.812 (±0.002)	0.188 (±0.002)	1.61
	5.00	-	-	-	-	0.893 (±0.001)	0.107 (±0.001)	4.72
	7.00	-	-	-	-	0.904 (±0.001)	0.096 (±0.001)	9.28
	10.00	-	-	-	-	0.929 (±0.001)	0.071 (±0.001)	19.33
	5.26	0	-	-	-	0.859 (±0.001)	0.141 (±0.001)	1.52
	5.20	35	-	-	-	0.835 (±0.001)	0.165 (±0.001)	2.64
	5.20	70	-	-	-	0.834 (±0.001)	0.166 (±0.001)	2.76
	4.92	140	-	-	-	0.835 (±0.001)	0.165 (±0.001)	2.75
10 mM tris pH 8.0	1.00	-	-	-	-	0.807 (±0.004)	0.193 (±0.003)	1.07
	1.95	-	-	-	-	0.831 (±0.002)	0.169 (±0.002)	1.07
	4.71	-	-	-	-	0.875 (±0.001)	0.125 (±0.001)	1.24
	6.37	-	-	-	-	0.956 (±0.001)	0.044 (±0.001)	9.23
	8.98	-	-	-	-	0.945 (±0.001)	0.055 (±0.001)	4.02

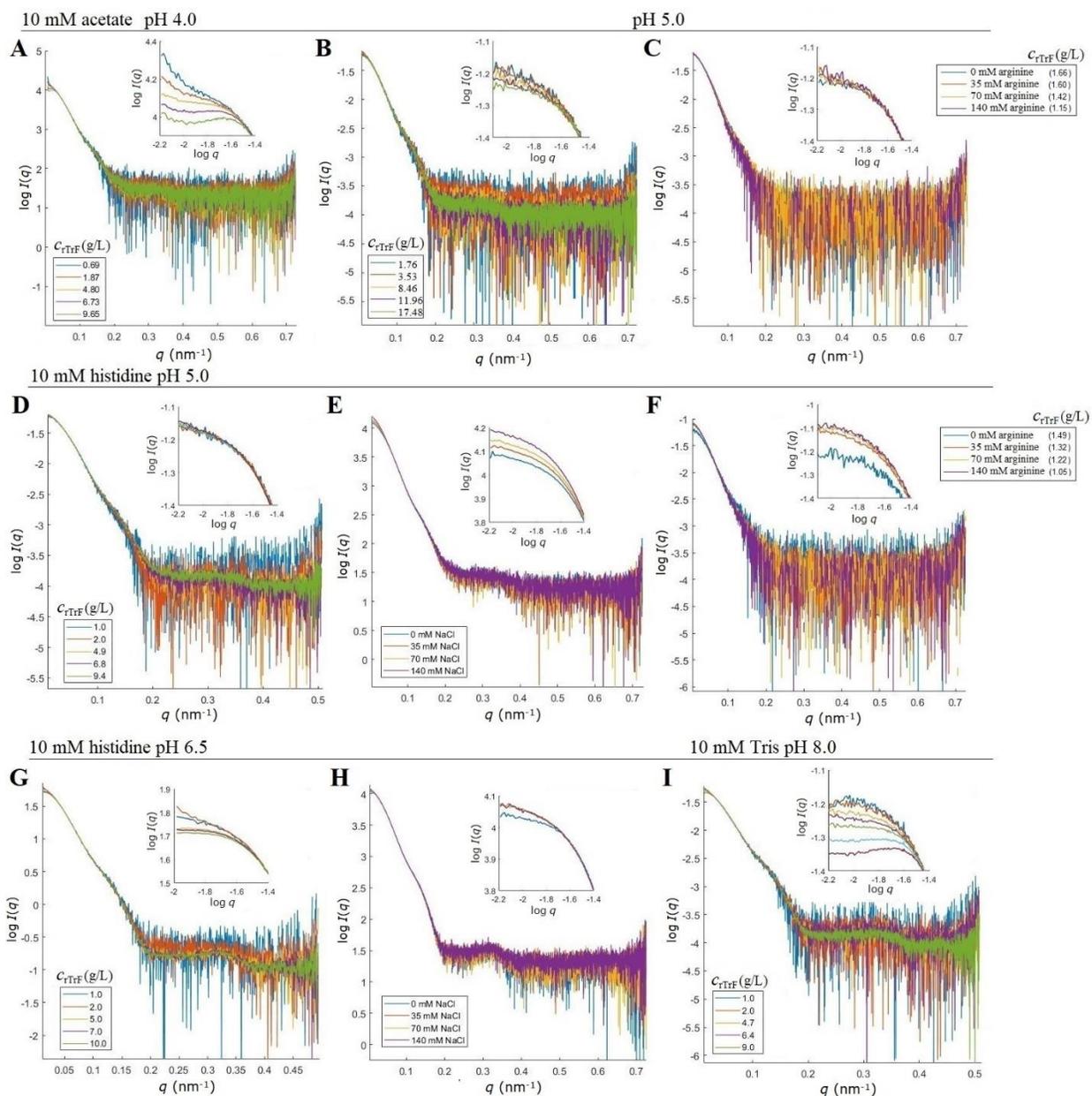


Figure S5.1: SAXS scattering curves for concentration series at A: 10 mM acetate pH 4.0; B: 10 mM acetate pH 5.0; D: 10 mM histidine pH 5.0; G: 10 mM histidine pH 6.5; I: 10 mM tris pH 8.0. SAXS scattering curves with varying c_{NaCl} at E: 10 mM histidine pH 5.0 and H: 10 mM histidine pH 6.5 with c_{Trif} around 5-5.5 g/L (see inset); and varying c_{arginine} at C (see inset): 10 mM acetate pH 5.0 and F: 10 mM histidine pH 5.0 with c_{Trif} around 1-1.5 g/L (see inset).

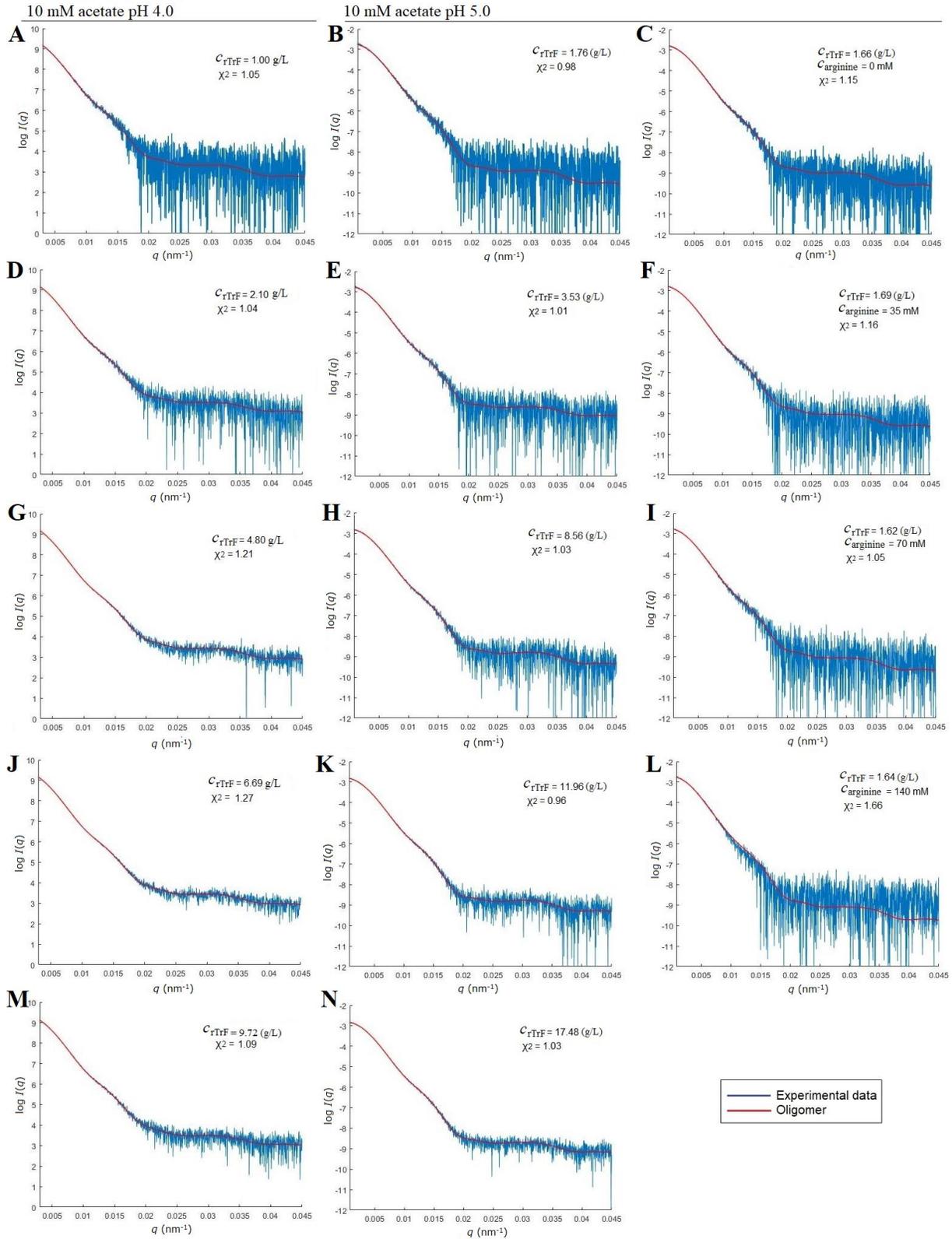


Figure S5.2: Fit plots from *OLIGOMER* analysis for c_{rTrF} at 10 mM acetate pH 4.0 (A, D, G, J, and M) and 10 mM acetate pH 5.0 (B, C, E, F, H, I, K, L, and N).

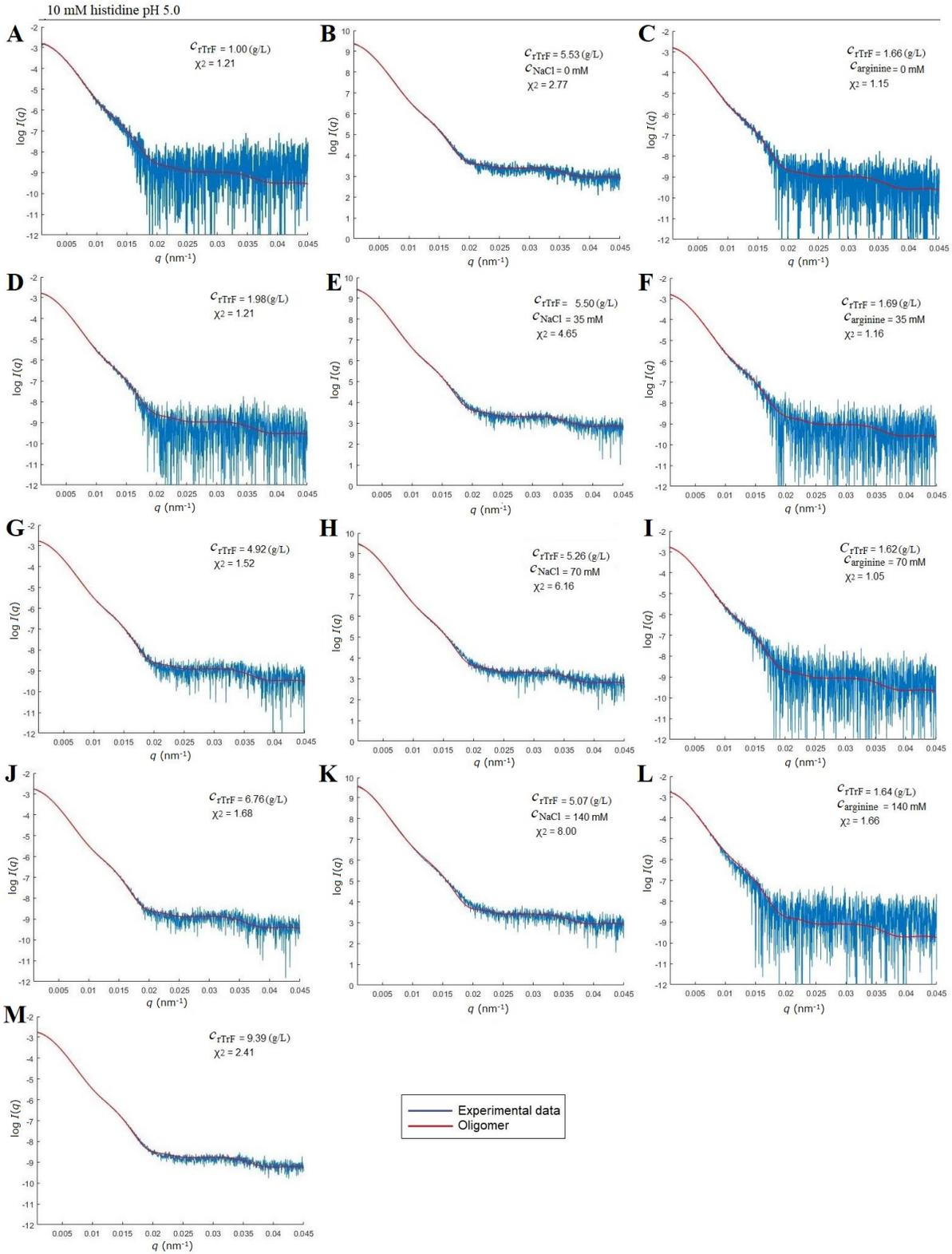


Figure S5.3: Fit plots from *OLIGOMER* analysis for varying c_{rTrF} at 10 mM histidine pH 5.0 (A, D, G, J, and M), varying c_{NaCl} at 10 mM histidine pH 5.0 (B, E, H, and K), and varying $c_{arginine}$ at 10 mM histidine pH 5.0 (C, F, I, and L).

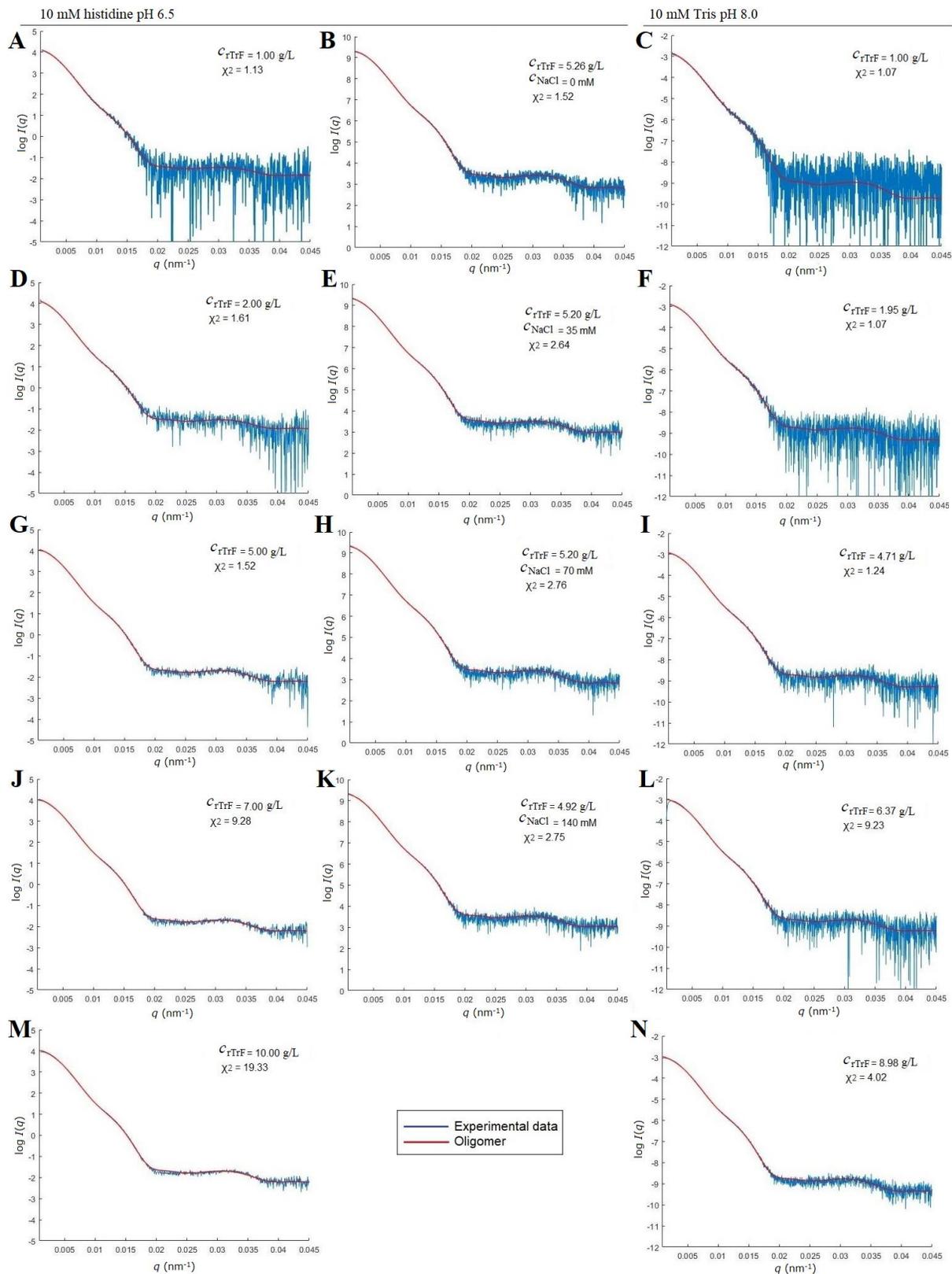


Figure S5.4: Fit plots from *OLIGOMER* analysis for varying c_{rTrF} at 10 mM histidine pH 6.5 (A, D, G, J, and M), and at 10 mM tris pH 8.0 (C, F, I, L, and N), and varying c_{NaCl} at 10 mM histidine pH 5.0 (B, E, H, and K).

Table S5.5: SEC-MALS experiments - overview of molecular weight (*MW*) and %Area at different pH and salt concentrations.

10 mM His pH 5.0		10 mM His pH 6.5		10 mM Tris pH 8.0		Oligomeric state
<i>MW</i> (kDa)	% Area	<i>MW</i> (kDa)	% Area	<i>MW</i> (kDa)	% Area	
0 mM NaCl						
79.1	85.2	75.3	85.4	74.9	85.2	Monomer
157.3	12.6	152.2	12.3	149.3	12.5	Dimer
241.6	1.9	240.3	1.9	229.0	1.9	Trimer
70 mM NaCl						
81.0	83.9	75.0	85.3	74.8	85.4	Monomer
157.5	14.0	148.9	12.5	148.7	12.4	Dimer
241.5	1.9	226.3	1.9	227.7	1.9	Trimer
140 mM NaCl						
83.8	86.0	75.0	85.5	74.9	85.7	Monomer
167.6	11.8	149.2	12.2	149.7	12.2	Dimer
267.3	1.7	225.5	1.9	230.6	1.8	Trimer

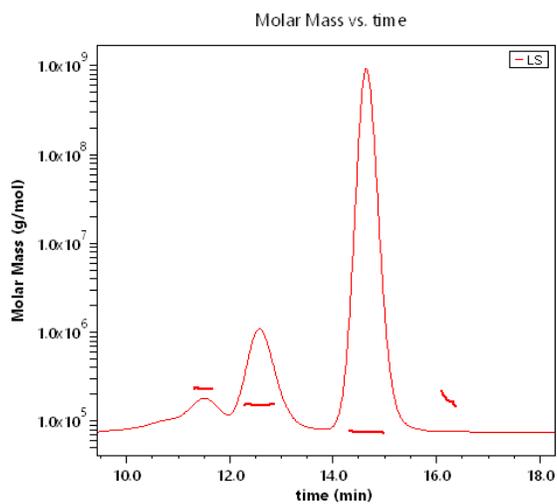


Figure S5.5: SEC-MALS result at 10 mM histidine pH 5.0.

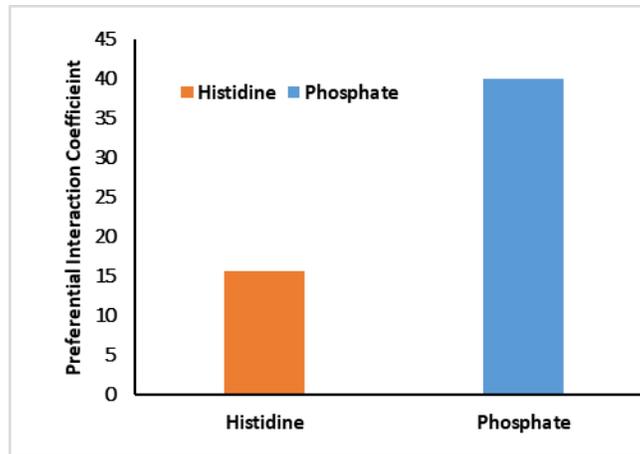


Figure S5.6: Preferential interaction coefficient (PIC) of histidine and phosphate to rTrF at pH 6.5. The tendency of phosphate to bind to rTrF is significantly higher than for histidine; $PIC_{\text{phosphate}} \sim 40$ vs $PIC_{\text{histidine}} \sim 15$.

Sequence¹⁶⁵

VPDKTVRWCAVSEHEATKCQSFDRDHMKSVIPSDGPSVACVKKASYLDCIRAIANEADAVTL DAGLVYDA
 YLAPNNLKPVVAEFYGSKEDPQTFYYAVAVVKKDSGFQMNQLRGKKSCHTGLGRSAGWNIPIGLLYCDLP
 EPRKPLEKAVANFFSGSCAPCADGTDFFPQLCQLCPGCGCSTLNQYFGYSGAFKCLKDAGDVAFVKHSTIF
 ENLANKADRQYELLCLDNTRKPVDEYKDCHLAQVPSHTVVVARSMGGKEDLIWELLNQAQEHFGKDKSK
 EFQLFSSPHGKDLLFKDSAHGFLKVPPRMDAKMYLGYEYVTAIRNLREGTCPEAPTDECKPVKWCALSHH
 ERLKCDEWSVNSVGKIECVSAETTEDCIAKIMNGEADAMSLDGGFVYIAGKCGLVPVLAENYNKADNCED
 TPEAGYFAVAVVKKASDLTWDNLKGGKSCHTAVGRTAGWNIPMGLLYNKINHCRFDEFFSEGCAPGSK
 KDSSLCKLCMGSGLNLCEPNNKEGYGYTGAFRCLVEKGDVAFVKHQTPQNTGGKNPDPWAKNLNEK
 DYELLCLDGTRKPVVEEYANCHLARAPNHAVVTRKDKACVHKILRQQHLFGSNVADCSGNFCLFRSETK
 DLLFRDDTVCLAKLHDRNTYEKYLGEYVAVGNLRKCSTSSLLEACTFRRP

Appendix C

Table S6.1: SAXS data collection overview.

Buffer	pH	Additives	$c_{\text{additives}}$ (M)	$c_{\text{HSA-NEP}}$ (g/L)
10 mM histidine	5.0	-	-	1.05, 1.62, 3.72, 5.39, 7.42, 10.12, and 14.09
		NaCl	0, 0.035, 0.070, and 0.140	5.83, 6.06, 6.08, and 5.70
	5.5	-	-	1.20, 2.03, 4.76, 6.62, 9.12, 12.45, and 16.66
		Urea	0, 1, 1.5, 3, 5.5, and 8	6.15, 5.39, 5.7, 5.85, 5.73, and 5.56
	6.5	-	-	1.77, 3.50, 8.74, 12.07, and 17.30
		NaCl	0, 0.035, 0.070, and 0.140	6.05, 5.83, 5.69, and 5.44
		Arginine	0, 0.035, 0.070, and 0.140	2.80, 2.11, 2.20, and 1.82
7.5	-	-	1.16, 2.40, 5.65, 8.11, 11.30, 17.02, and 22.69	
10 mM phosphate	6.5	-	-	2.09, 4.11, 9.59, 15.26, and 20.52
10 mM tris	8.5	-	-	1.02, 2.29, 5.49, 7.56, and 10.49

Table S6.2: SAXS experimental details.

a) Sample details			
Organism	HSA-NEP fusion protein Chinese Hamster Ovary (CHO) cells		
Source	AstraZeneca		
Extinction coefficient (Abs 0.1% = 1 g/L)	1.04		
Molecular weight from chemical composition (kDa)	146.7		
b) SAXS data collection			
Instrument	P12 BioSAXS beamline (PETRAIII)		
Date	12.17	07.18	12.18
Detector	PILATUS2M	PILATUS6M	
Wavelength (Å)	0.124402		
Beam size (mm ²)	0.2 × 0.12		
Detector distance (m)	3.000		
q-measurement range (nm ⁻¹)	0.027-5.078	0.026 -7.288	0.0261-7.263
Absolute scaling method	Comparison with scattering from pure H ₂ O		Comparison with scattering from BSA
Normalization	To transmitted intensity by beam-stop counter		
Monitoring for radiation damage	Frame-by-frame comparison		
Exposure time (s)	20 x 0.05		30 x 0.095
Sample configuration	Quartz glass capillary		
Sample temperature (°C)	20		
c) Software employed for SAXS data reduction, analysis and interpretation			
SAS data reduction	<i>PRIMUSqt</i> ⁴⁶ from <i>ATSAS</i> 2.8.3 ¹⁶⁶		
Basic analyses: Guinier, $p(r)$, V_p	<i>PRIMUSqt</i> ⁴⁶		
Ensemble representation of atomic models	<i>EOM</i> ^{45,136}		
Molecular graphics	<i>PyMOL</i> (version 1.8.2.3, Schrödinger, LLC)		
Figures	<i>MatLab</i>		

Table S6.3: Structural parameters derived from SAXS experiments.

	CHSA- NEP (g/L)	C _{NaCl} (mM)	C _{arginine} (mM)	Guinier		$p(r)$			Apparent MW (kDa)	
				$I(0)/c$	R_g (nm)	$I(0)/c$	R_g (nm)	D_{max} (nm)	Guinier	$p(r)$
10 mM histidine pH 5.0	1.05	-	-	0.13	4.91	0.13	5.07	16.00	180	180
	1.62	-	-	0.13	4.88	0.13	5.06	16.30	180	180
	3.72	-	-	0.13	4.77	0.13	4.97	16.16	180	180
	5.39	-	-	0.13	4.70	0.13	4.91	15.60	180	180
	7.42	-	-	0.13	4.63	0.13	4.88	15.10	180	180
	10.12	-	-	0.13	4.56	0.13	4.83	15.00	180	180
	14.09	-	-	0.13	4.51	0.13	4.81	14.90	180	180
	5.83	0	-	23808*	5.03	23990*	5.19	17.72	159	160
	6.06	35	-	23961*	5.14	24030*	5.25	17.70	160	160
	6.08	70	-	23662*	5.05	23950*	5.25	18.00	158	160
5.70	140	-	23984*	5.19	24010*	5.29	17.50	160	160	
10 mM histidine pH 5.5	1.20	-	-	0.11	4.86	0.11	5.05	15.60	152	152
	2.03	-	-	0.10	4.78	0.11	4.98	15.40	138	152
	4.76	-	-	0.11	4.76	0.11	5.08	15.70	152	152
	6.62	-	-	0.10	4.45	0.10	4.83	15.10	133	138
	9.12	-	-	0.09	4.33	0.10	4.75	15.10	129	138
	12.45	-	-	0.09	4.16	0.10	4.69	15.00	123	138
	16.66	-	-	0.08	3.68	0.09	4.45	14.50	107	125
	6.05	0	-	20350*	4.14	22620*	4.85	15.10	136	151
10 mM histidine pH 6.5	1.77	-	-	0.13	4.66	0.13	5.01	15.70	180	180
	3.50	-	-	0.12	4.57	0.13	4.95	15.70	166	180
	8.74	-	-	-	-	0.12	4.90	15.50	-	166
	12.07	-	-	-	-	0.11	4.89	15.40	-	152
	17.30	-	-	-	-	0.10	4.67	15.10	-	138
	5.83	35	-	21725*	4.65	22110*	4.89	15.50	145	148
	5.69	70	-	22255*	4.76	22540*	4.96	15.70	149	150
	5.44	140	-	22498*	4.79	22830*	5.00	15.80	150	152
	2.80	-	0	0.086	4.62	0.09	4.95	15.77	119	125
	2.11	-	35	0.093	5.00	0.09	5.13	15.80	129	125
2.20	-	70	0.093	5.03	0.09	5.17	16.00	129	125	
1.82	-	140	0.090	5.17	0.09	5.21	16.20	125	125	
10 mM phosphate pH 6.5	2.09	-	-	0.12	4.87	0.12	5.07	15.50	166	166
	4.11	-	-	0.11	4.59	0.11	4.90	15.50	152	152
	9.59	-	-	0.10	4.14	0.10	4.52	14.40	133	138
	15.26	-	-	0.09	3.95	0.09	4.33	14.20	125	125
	20.52	-	-	0.08	3.60	0.08	4.26	14.20	111	111
10 mM histidine pH 7.5	1.16	-	-	0.12	5.03	0.12	5.45	16.60	166	166
	2.40	-	-	0.12	4.92	0.12	5.43	16.60	166	166
	5.65	-	-	0.12	4.98	0.13	5.50	17.00	166	180
	8.11	-	-	0.12	4.92	0.13	5.42	17.00	166	180
	11.30	-	-	0.12	4.76	0.13	5.53	17.80	166	180
	17.02	-	-	-	-	0.13	5.22	16.60	-	180
	22.69	-	-	-	-	0.13	5.11	16.00	-	180
	2.80	-	0	0.086	4.62	0.09	4.95	15.77	119	125
10 mM tris pH 8.5	1.02	-	-	23795*	4.92	24490*	5.28	16.80	159	163
	2.29	-	-	22704*	4.70	24020*	5.18	16.20	152	160
	5.49	-	-	20674*	4.30	23280*	5.05	16.00	138	155
	7.56	-	-	-	-	24800*	5.21	16.50	-	166
	10.49	-	-	-	-	24350*	5.14	16.50	-	163

Note: * BSA calibration

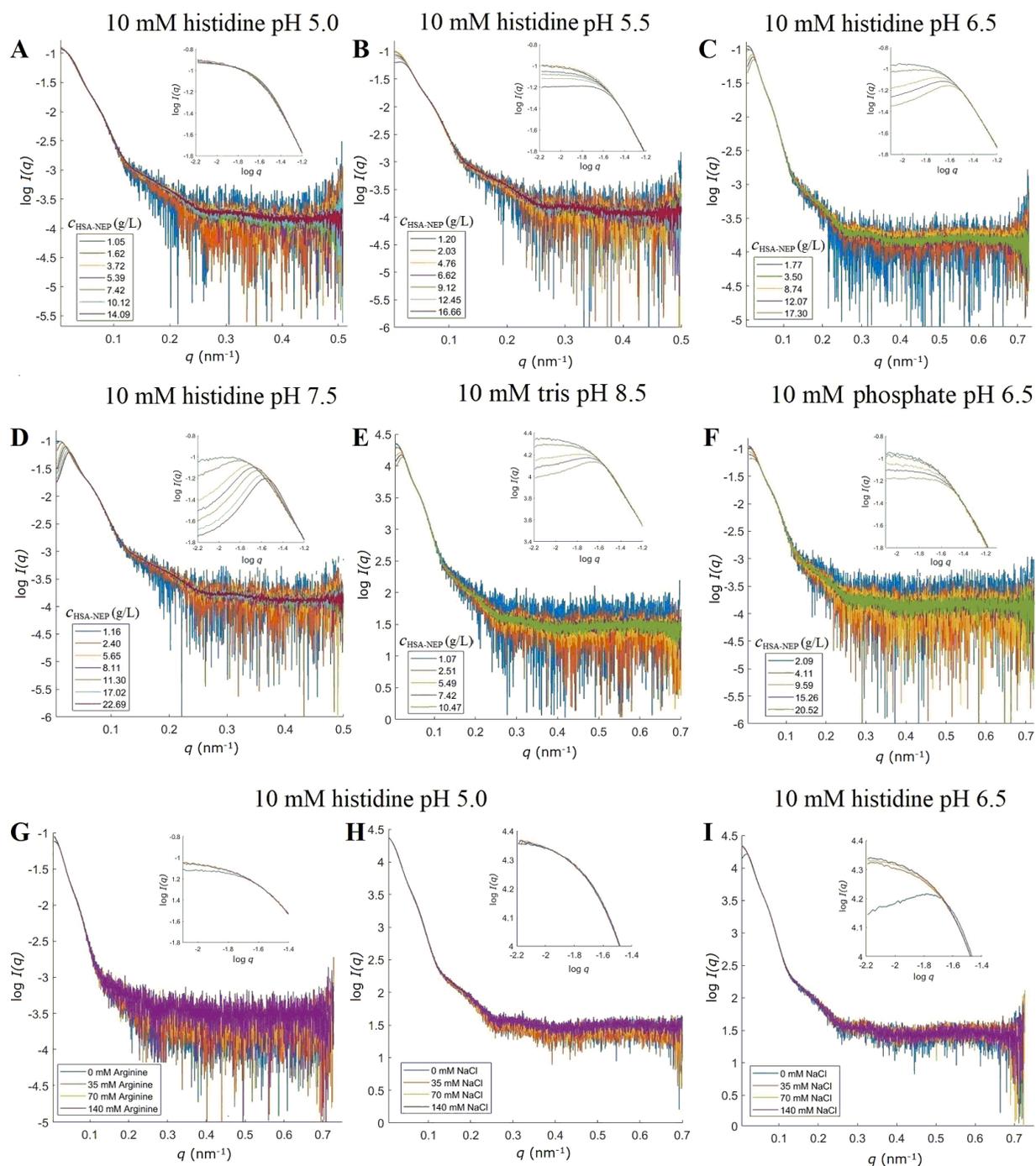


Figure S6.1: SAXS data collected at different pH with and without additives. Concentration series: A: 10 mM histidine pH 5.0; B: 10 mM histidine pH 5.5; C: 10 mM histidine pH 6.5; D: 10 mM histidine pH 7.5; E: 10 mM tris pH 8.5; F: 10 mM phosphate pH 6.5. Additives: 10 mM histidine pH 5.0 with G: arginine with $c_{\text{HSA-NEP}}$ around 2 g/L and H: NaCl with $c_{\text{HSA-NEP}}$ around 5.5-6 g/L; I: 10 mM histidine pH 6.5 with NaCl with $c_{\text{HSA-NEP}}$ around 6 g/L.

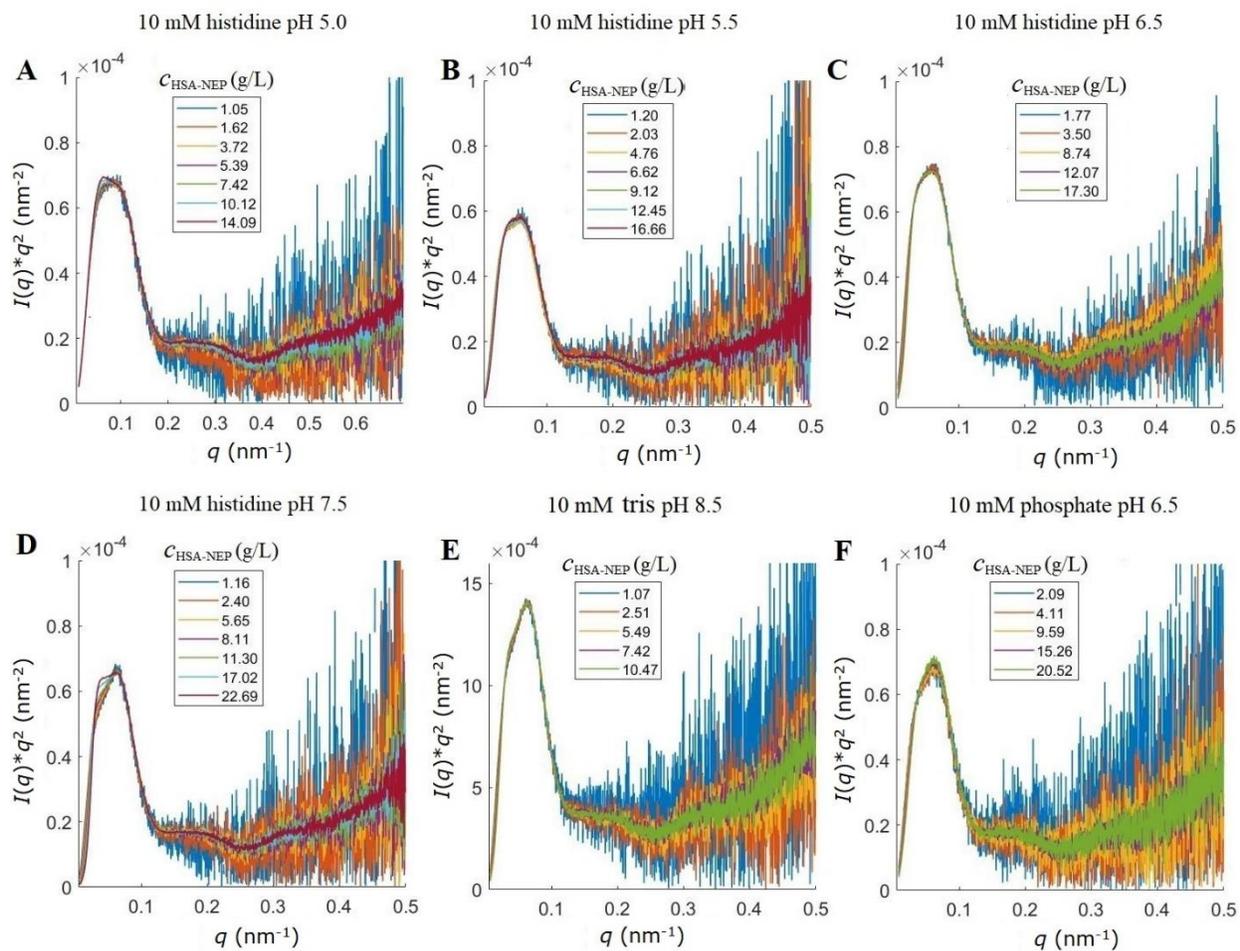


Figure S6.2: Kratky plot of a HSA-NEP at different pH: A: 10 mM histidine at pH 5.0; B: 10 mM histidine at pH 5.5; C: 10 mM histidine at pH 6.5; D: 10 mM histidine at pH 7.5; E: 10 mM tris at pH 8.5; F: 10 mM phosphate at pH 6.5.

Table S6.4: Structural parameters for merged files used for *EOM*.

	Histidine				Tris	Phosphate
	pH 5.0	pH 5.5	pH 6.5	pH 7.5	pH 8.5	pH 6.5
Guinier analysis						
$I(0)/c$ ($\text{Lg}^{-1}\text{cm}^{-1}$)	0.11	0.09	0.11	0.10	20250*	0.10
R_g (nm)	4.38	4.33	4.28	4.36	4.35	4.27
q_{\min} (nm^{-1})	0.2812	0.2839	0.2964	0.2875	0.2802	0.2964
MW from $I(0)$ (ratio to predicted) (kDa)	152 (1.04)	127 (0.86)	152 (1.04)	133 (0.91)	135 (0.92)	138 (0.94)
$p(r)$ analysis						
$I(0)$ (cm^{-1})	0.12	0.10	0.13	0.11	24520*	0.12
R_g (nm)	4.83	4.93	5.05	5.15	5.28	5.06
D_{\max} (nm)	15.00	15.30	16.00	16.80	16.40	16.00
q range (nm^{-1})	0.2812- 1.9013	0.2867- 1.9013	0.2964- 2.0127	0.2802- 1.9058	0.2930- 1.9082	0.2964- 1.9989
χ^2 (total estimate from <i>GNOM</i>)	2.897	2.650	3.622	2.908	3.083	1.395
MW from $I(0)$ (ratio to predicted) (kDa)	166 (1.13)	138 (0.94)	180 (1.23)	152 (1.04)	164 (1.12)	166 (1.13)
Porod volume (nm^{-3}) (ratio $V_p/\text{calculated } MW$)	235.2 (1.70)	243.7 (1.70)	245.0 (1.70)	241.4 (1.70)	245.03 (1.70)	236.43 (1.70)
V , MW using Fischer method (ratio of MW to expected)	138.4 (0.94)	143.4 (0.98)	144.1 (0.98)	142.0 (0.97)	144.1 (0.98)	139.1 (0.95)

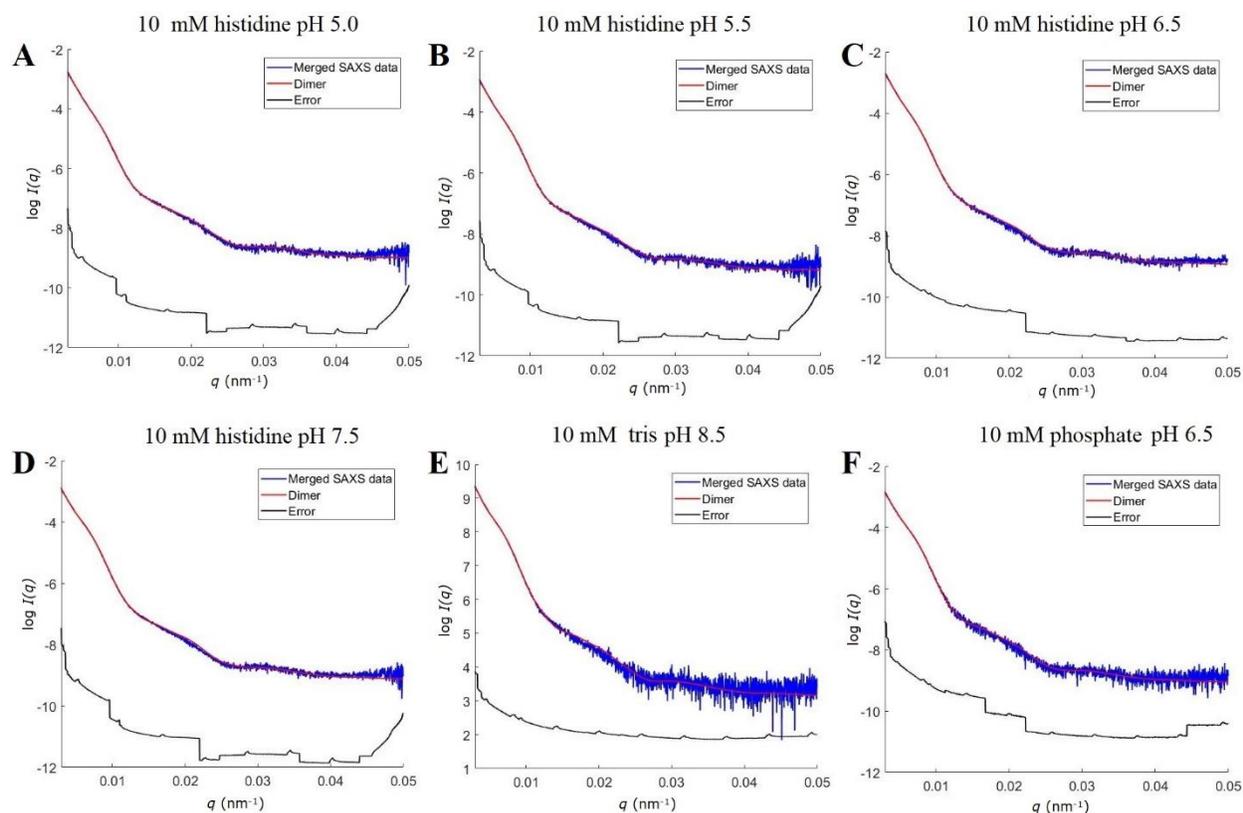


Figure S6.3: Fit plots from EOM analysis against merged experimental data. A: 10 mM histidine pH 5.0; B: 10 mM histidine pH 5.5; C: 10 mM histidine pH 6.5; D: 10 mM histidine pH 7.5; E: 10 mM tris pH 8.5; F: 10 mM phosphate pH 6.5.

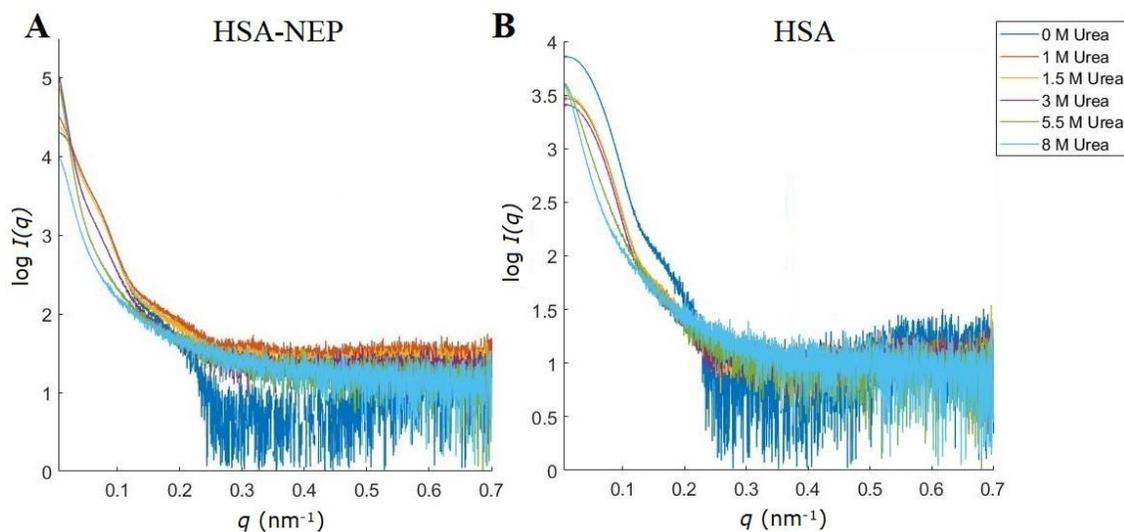


Figure S6.4: SAXS scattering curves for A: HSA-NEP and B: HSA in 10 mM histidine pH 5.5 at different concentrations of urea with $c_{\text{HSA-NEP}}$ around 5 g/L.

Sequence

DAHKSEVAHRFKDLGEENFKALVLIFAQYLQQSPFEDHVKLVNEVTEFAKTCVADESAENCDSLHTLF
GDKLCTVATLRETYGEMADCCAQKQEPERNECFLQHKDDNPPLRVLVRPEVDVMCTAFHDNEETFLKKYL
YEIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDELDEGKASSAKQRLKCASLQKFGER
AFKAWAVARLSQRFPKAEFAEVSKLVDTLTKVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECC
EKPLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVLLLR
AKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQLGEYKFQNALLVRYTKKVPQVST
PTLVEVSRNLGKVGSKCCKHPEAKRMPCAEDYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSA
LEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCK
ADDKETCFAEEGKKLVAASQAALGLGGGSYDDGICKSSDCIKSAARLIQNMDATTEPCTDFFKYACGGW
LKRNVIPETSSRYGNFDILRDELEVVLKDVLPQPKTEDIVAVQKAKALYRSCINESAIDSRGGEPLLKLLPDI
YGWPVATENWEQKYGASWTAEKAIQLNSKYGKKVLINLFGTDDKNSVNHVIHIDQPRGLPSRDYEE
CTGIYKEACTAYVDFMISVARLIRQEERLPIDENQLALEMNKVMEMEKEIANATAKPEDRNDPMLLYNKMT
LAQIQNNFSLEINGKPFWSLWLNFTNEIMSTVNISITNEEDVVVYAPEYLTCLKPILTKYSARDLQNLMSWRFIM
DLVSSLSRPTYKESRNAFRKALYVTTSETATWRRCANYVNGNMENAVGRLYVEAAFAGESKHVVEDLIAQI
REVFQTLDDLTWMDAETKKRAEEKALAIKERIGYPDDIVSNDNKLNNEYLELNYKEDEYFENIIQNLKFSQ
SKQLKKLREKVDKDEWISGAAVVNAFYSSGRNQIVFPAGILQPPFFSAQQSNSLNYGGIGMVGHEITHGFD
DNGRNFNKDGDLVDWWTQQSASNFKESQSCMVYQYGNFSWDLAGGQHLNGINTLGENIADNGGLGQA
YRAYQNYIKKNGEEKLLPGLDLNHLKQLFFLNFAQVWCGTYRPEYAVNSIKTDVHSPKNFRIIGTLQNSAEF
SEAFHCRKNSYMNPEKKCRVW

Linker

Appendix D

Table S7.1: Protein overview.

	Type	MW (kDa)	pI	Extinction coefficient A280 (L g ⁻¹ cm ⁻¹)	Sequence
mAb ₁	IgG ₁ λ YTE	146.2	8.99	1.76	Not available
mAb ₂	Human IgG ₁ κ TM	148.9	9.04	1.66	Available
mAb ₃	IgG ₂ κ	145.1	7.78	1.31	Available

Table S7.2: Refolding test for ICD studies.

	Without denaturant	c_{\max}^*	1:10 dilution of c_{\max}	Conclusion
	λ_{\max}			
mAb ₁	344	354	346	Reversible
mAb ₂	342	356	346	Partially reversible
mAb ₃	320	356	314	Reversible

*maximum concentration of denaturant

Table S7.3: SAXS data collection overview.

Buffer	pH	Additives	$c_{\text{additives}}$ (M)	c_{mAb} (g/L)
mAb₁				
10 mM histidine	6.0	-	-	1.16, 1.90, 6.98, 8.32, and 12.46
		Urea	0, 1, 2, 2.5, and 5	2.29, 2.25, 2.30, 2.33, and 2.25
mAb₂				
10 mM histidine	5.0	-	-	1.03, 2.03, 4.69, 6.37, 8.48, 11.17, and 15.15
		Arginine	0.14	2.63, 5.92, and 9.17
		GuHCl	0, 1, 1.5, 2, 3, and 4.5	1.08, 1.01, 0.96, 1.12, 1.12, and 1.04
	6.0	-	-	1.01, 2.29, 6.22, 9.24, 12.37, 18.54, 24.95, and 36.32
	6.5	-	-	1.52, 2.86, 7.47, 10.51, and 14.31
		Arginine	0.14	2.11, 5.51, and 8.77
	7.5	-	-	1.29, 2.55, 7.20, 10.48, 14.86, 22.02, and 28.80
mAb₃				
10 mM histidine	5.0	-	-	1.03, 1.54, 3.62, 4.76, 7.09, 9.80, 11.92, and 17.84
		NaCl	0, 0.035, 0.070, and 0.140	0.67, 1.00, 0.86, and 1.05
	6.0	-	-	1.57, 2.54, 5.59, 7.75, 10.82, 15.95, 21.68, and 31.50
	6.5	-	-	0.98, 1.81, 4.30, 6.06, 8.49, 12.92, 17.14, 27.20, and 41.16
		NaCl	0, 0.035, 0.070, and 0.140	1.27, 1.56, 1.50, and 1.48
		Arginine	0, 0.035, 0.070, and 0.140	1.90, 1.84, 1.61, and 1.60

Table S7.4: SAXS experimental details.

a) SAXS data collection			
Instrument	P12 BioSAXS beamline (PETRAIII)		
Date	06.17	12.17	07.18
Detector	PILATUS2M		PILATUS6M
Wavelength (Å)	1.241	1.240	0.124402
Beam size (mm ²)	0.2 × 0.12		
Detector distance (m)	3.000		
q-measurement range (nm ⁻¹)	0.025-5.036	0.027-5.078	0.026 -7.288
Absolute scaling method	Comparison with scattering from pure H ₂ O		
Normalization	To transmitted intensity by beam-stop counter		
Monitoring for radiation damage	Frame-by-frame comparison		
Exposure time (s)	20 x 0.05		
Sample configuration	Quartz glass capillary		
Sample temperature (°C)	20		
Measured proteins	mAb ₁ , mAb ₂ and mAb ₃	mAb ₂ and mAb ₃	mAb ₂ and mAb ₃
b) Software employed for SAXS data reduction, analysis, and interpretation			
SAS data reduction	<i>PRIMUSqt</i> ⁴⁶ from <i>ATSAS</i> 2.8.3 ¹⁶⁶		
Basic analyses: Guinier, $p(r)$, V_P	<i>PRIMUSqt</i> ⁴⁶		
Ensemble representation of atomic models	<i>CORAL</i> ⁴³		
Molecular graphics	<i>PyMOL</i> (version 1.8.2.3, Schrödinger, LLC)		
Figures	<i>MatLab</i>		

Table S7.5: Structural parameters for mAb₁ derived from SAXS experiments.

	c_{mAb} (g/L)	Guinier		$p(r)$			Apparent MW (kDa)		
		$I(0)/c$	R_g (nm)	$I(0)/c$	R_g (nm)	D_{max} (nm)	$I(0)/c$	R_g (nm)	$I(0)/c$
10 mM histidine pH 6.0	1.14	0.100	5.08	0.100	5.31	17.20	138	138	169
	1.87	0.110	5.20	0.110	5.32	17.00	152	152	172
	6.85	0.110	4.87	0.120	5.67	19.00	152	166	212
	8.17	0.110	4.53	0.120	5.61	18.50	152	166	209
	12.23	0.100	3.58	0.120	5.66	18.60	138	166	220

10 mM histidine pH 6.0

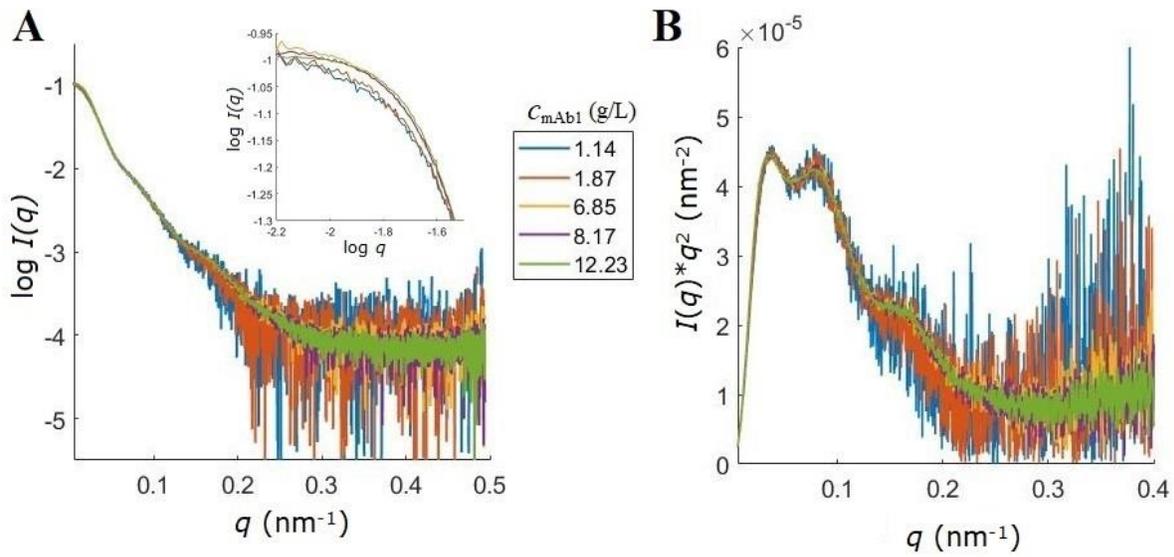


Figure S7.1: Concentration series SAXS data collected at different pH. A: log scale and B: the Kratky plot for mAb₁ in 10 mM histidine pH 6.0.

Table S7.6: Structural parameters for mAb₂ derived from SAXS experiments.

	C_{mAb} (g/L)	C_{NaCl} (mM)	C_{arginine} (mM)	Guinier			p(r)		Apparent MW (kDa)		
				<i>I</i> (0)/ <i>c</i>	<i>R_g</i> (nm)	<i>I</i> (0)/ <i>c</i>	<i>R_g</i> (nm)	<i>D_{max}</i> (nm)	<i>I</i> (0)/ <i>c</i>	<i>R_g</i> (nm)	<i>I</i> (0)/ <i>c</i>
10 mM histidine pH 5.0	1.03	-	-	0.092	4.93	0.090	5.07	15.60	127	125	146
	2.29	-	-	0.089	4.82	0.090	5.03	15.20	123	125	145
	5.29	-	-	0.074	3.32	0.090	4.91	14.80	102	125	139
	7.16	-	-	0.080	3.74	0.090	4.96	14.80	111	125	144
	9.57	-	-	-	-	0.090	4.88	14.70	-	125	139
	12.64	-	-	-	-	0.080	4.74	14.40	-	111	128
	2.63	-	140	0.110	6.01	0.100	5.44	19.10	152	138	168
	5.92	-	140	0.110	5.69	0.110	5.63	18.40	152	152	184
	9.17	-	140	0.110	6.01	0.110	6.05	20.80	152	152	203
10 mM histidine pH 6.0	1.01	-	-	0.100	4.92	0.100	5.06	15.60	138	138	151
	2.38	-	-	0.096	4.87	0.100	5.03	15.40	133	138	149
	6.50	-	-	0.090	4.36	0.100	5.01	15.40	125	138	153
	9.72	-	-	0.083	3.53	0.100	5.02	15.30	115	138	158
	13.03	-	-	0.079	2.68	0.090	4.99	15.60	109	125	155
10 mM histidine pH 6.5	1.52	-	-	0.110	5.22	0.110	5.29	17.50	152	152	170
	4.10	-	-	0.110	5.37	0.110	5.49	18.20	152	152	182
	10.70	-	-	0.140	6.19	0.140	6.23	21.50	194	194	251
	15.05	-	-	0.150	6.54	0.140	6.46	21.60	208	194	283
	2.11	-	140	0.110	5.44	0.100	5.38	18.40	152	138	169
	5.51	-	140	0.110	5.63	0.110	5.74	21.60	152	152	186
	8.77	-	140	0.110	6.01	0.110	6.03	23.20	152	152	202
10 mM histidine pH 7.5	1.29	-	-	0.110	5.20	0.110	5.28	17.50	152	152	175
	2.34	-	-	0.120	5.33	0.120	5.35	17.50	166	166	180
	6.61	-	-	0.130	5.50	0.130	5.77	19.90	180	180	214
	9.63	-	-	0.130	5.32	0.130	5.87	20.20	180	180	227
	13.66	-	-	0.130	4.88	0.140	5.98	21.60	180	194	237

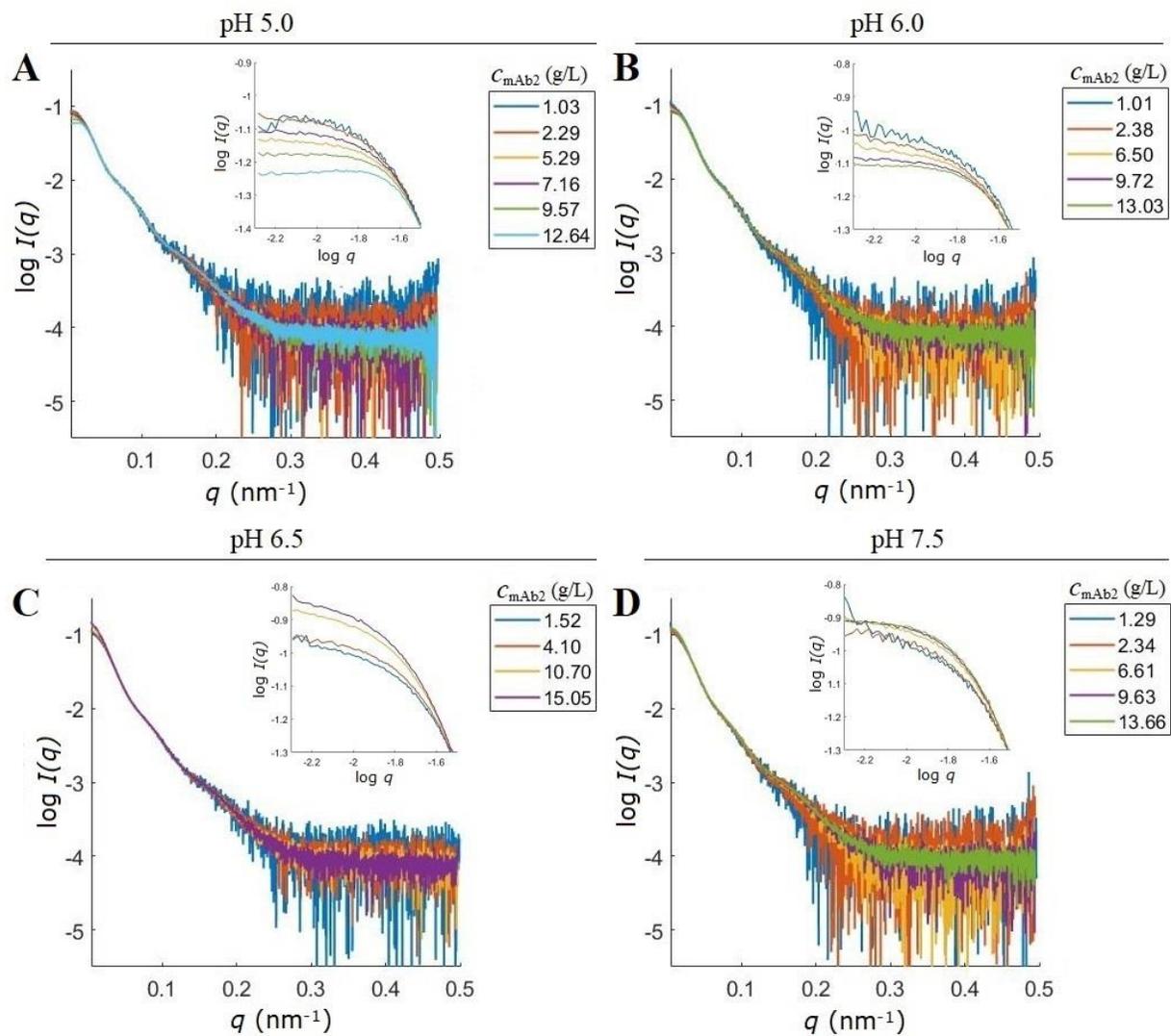


Figure S7.2: Concentration series SAXS data collected for mAb₂ in 10 mM histidine at different pHs. A: pH 5.0, B: pH 6.0, C: pH 6.5, and D: pH 7.5.

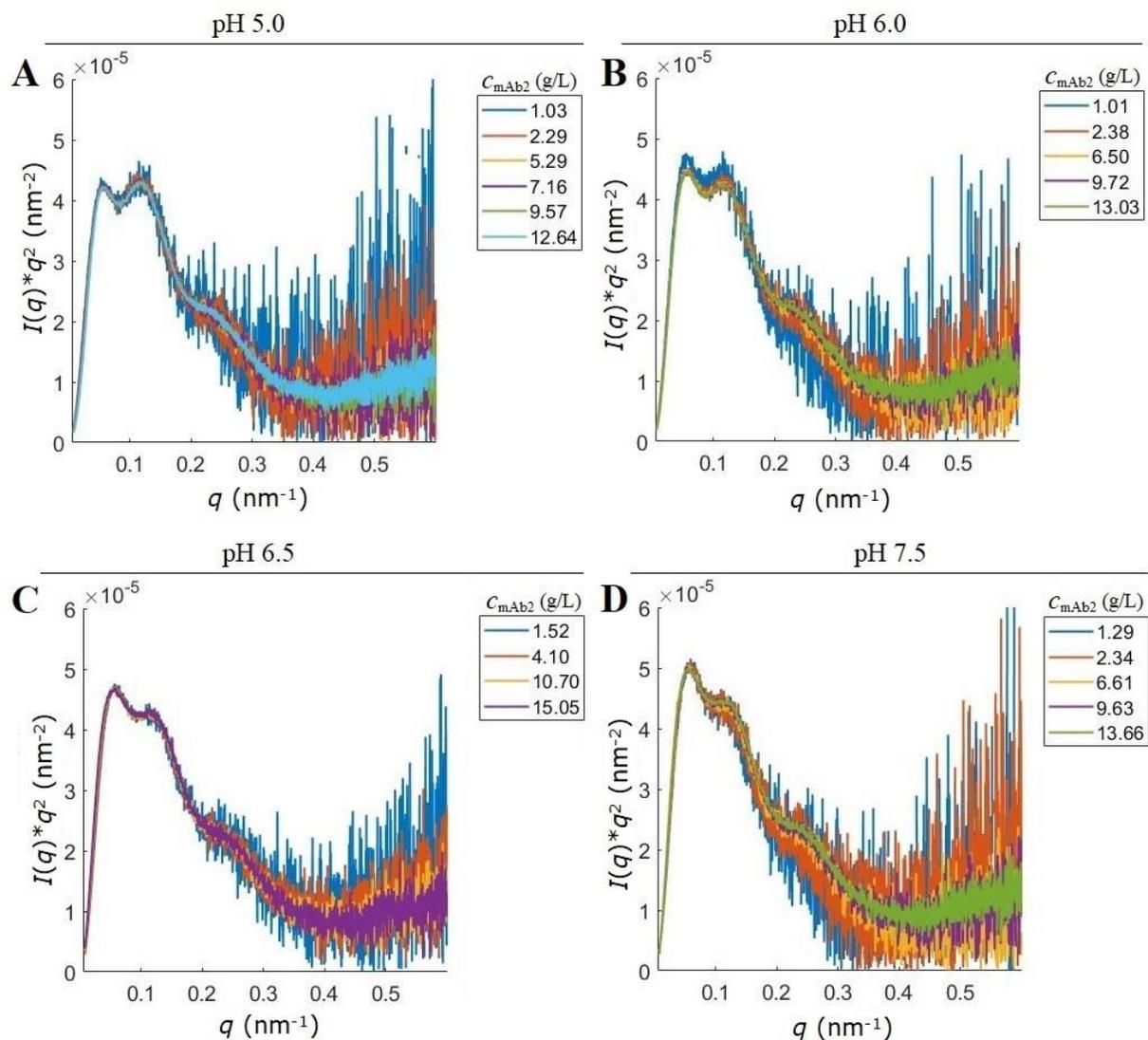


Figure S7.3: Kratky plot of a mAb₂ in 10 mM histidine buffer at A: pH 5.0, B: pH 6.0, C: pH 6.5, and D: pH 7.5.

Table S7.7: Structural parameters for mAb₃.

	C_{mAb} (g/L)	C_{NaCl} (mM)	C_{arginine} (M)	Guinier		<i>p</i>(<i>r</i>)		Apparent <i>MW</i> (kDa)			
				<i>I</i>(0)/<i>c</i>	<i>R_g</i> (nm)	<i>I</i>(0)/<i>c</i>	<i>R_g</i> (nm)	<i>D_{max}</i> (nm)	Guinier	<i>p</i>(<i>r</i>)	Porod Volume
10 mM histidine pH 5.0	1.03	-	-	0.110	4.85	0.110	4.91	15.00	152	152	139
	1.59	-	-	0.100	4.73	0.100	4.87	14.80	138	138	135
	3.74	-	-	0.092	4.05	0.100	4.82	14.70	127	138	138
	4.92	-	-	0.085	3.53	0.100	4.77	14.30	118	138	135
	7.32	-	-	0.075	2.47	0.100	4.74	14.30	104	138	134
	10.12	-	-	-	-	0.090	4.63	14.10	-	125	124
	12.31	-	-	-	-	0.090	4.53	13.80	-	125	116
	1.00	0	-	0.110	4.87	0.110	4.93	15.20	152	152	138
	1.35	35	-	0.110	5.01	0.110	5.01	15.60	152	152	147
	1.17	70	-	0.110	5.14	0.110	5.06	15.60	152	152	151
1.41	140	-	0.110	5.13	0.110	5.07	15.60	152	152	154	
10 mM histidine pH 6.0	1.57	-	-	0.110	4.84	0.110	4.89	14.90	152	152	142
	2.53	-	-	0.100	4.71	0.100	4.86	14.40	138	138	137
	5.58	-	-	0.097	4.37	0.100	4.81	14.50	134	138	136
	7.73	-	-	0.094	4.21	0.100	4.76	14.30	130	138	133
	10.24	-	-	0.089	3.76	0.100	4.77	14.30	123	138	136
	14.54	-	-	0.084	3.42	0.100	4.72	14.20	116	138	132
10 mM histidine pH 6.5	0.98	-	-	0.110	4.96	0.110	4.99	15.20	152	152	148
	1.64	-	-	0.110	4.93	0.110	5.00	15.10	152	152	147
	3.88	-	-	0.110	4.84	0.110	4.95	15.20	152	152	144
	5.46	-	-	0.110	4.84	0.110	4.95	15.20	152	152	146
	7.66	-	-	0.110	4.77	0.110	4.94	15.10	152	152	144
	11.66	-	-	0.110	4.89	0.110	5.00	15.50	152	152	148
	15.47	-	-	0.110	4.80	0.110	4.97	15.50	152	152	145
	1.27	0	-	0.110	4.95	0.110	5.00	15.20	152	152	146
	1.56	35	-	0.110	5.14	0.110	5.04	15.40	152	152	150
	1.50	70	-	0.110	5.14	0.110	5.07	15.40	152	152	147
	1.48	140	-	0.110	5.21	0.110	5.07	15.40	152	152	157
	1.90	-	0	0.110	4.95	0.110	5.01	15.20	152	152	149
	1.79	-	35	0.110	5.33	0.110	5.04	15.50	152	152	156
	1.56	-	70	0.110	5.37	0.110	5.02	15.50	152	152	154
1.55	-	140	0.110	4.61	0.110	5.07	16.40	152	152	157	

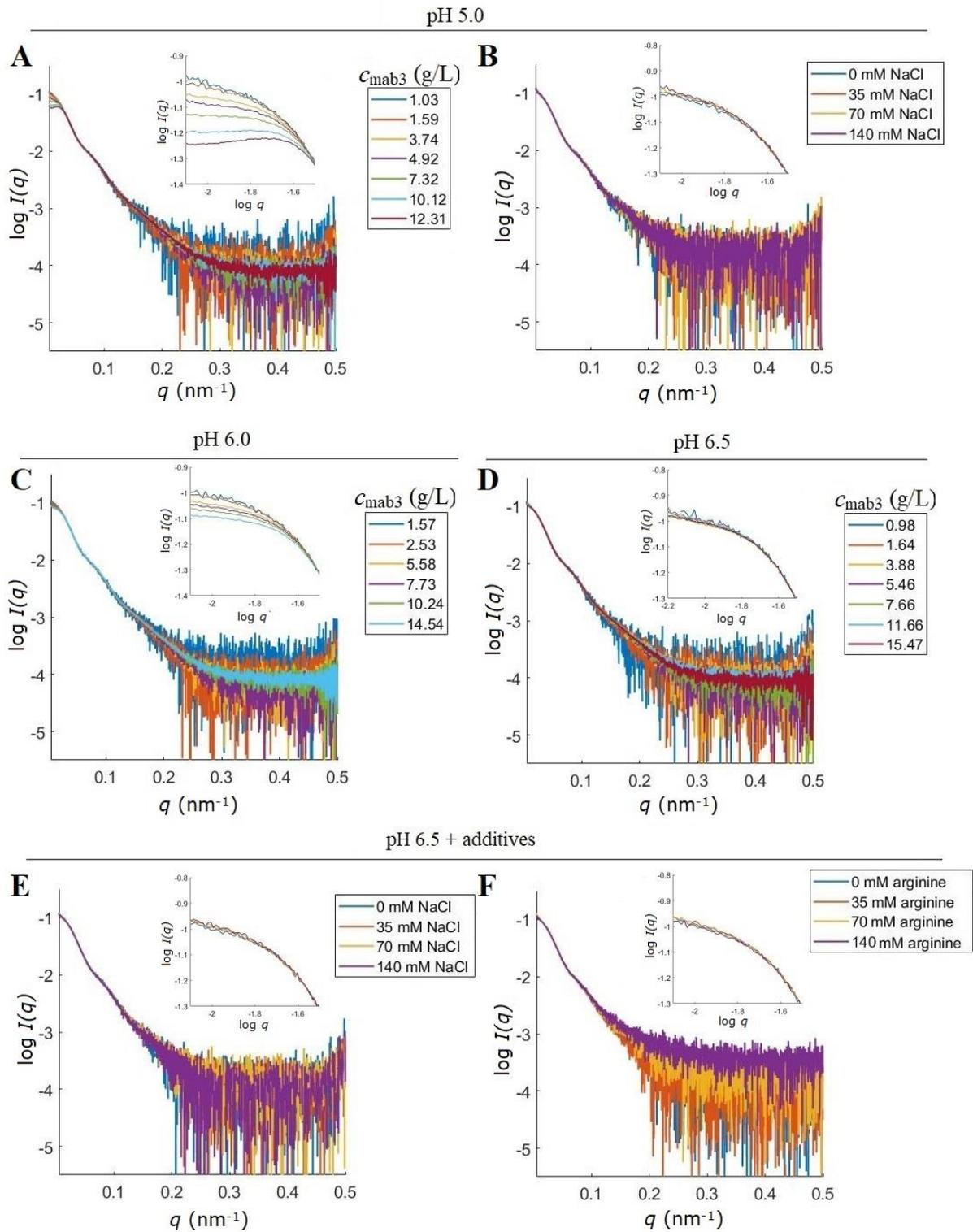


Figure S7.4: SAXS data collected for mAb₃ in 10 mM histidine at different pH with and without additives. Concentration series: A: pH 5.0, C: pH 6.0, and D: pH 6.5. Additives: B: pH 5.0 with NaCl, E: pH 6.5 with NaCl, and F: pH 6.5 with arginine with c_{mAb} around 1 g/L.

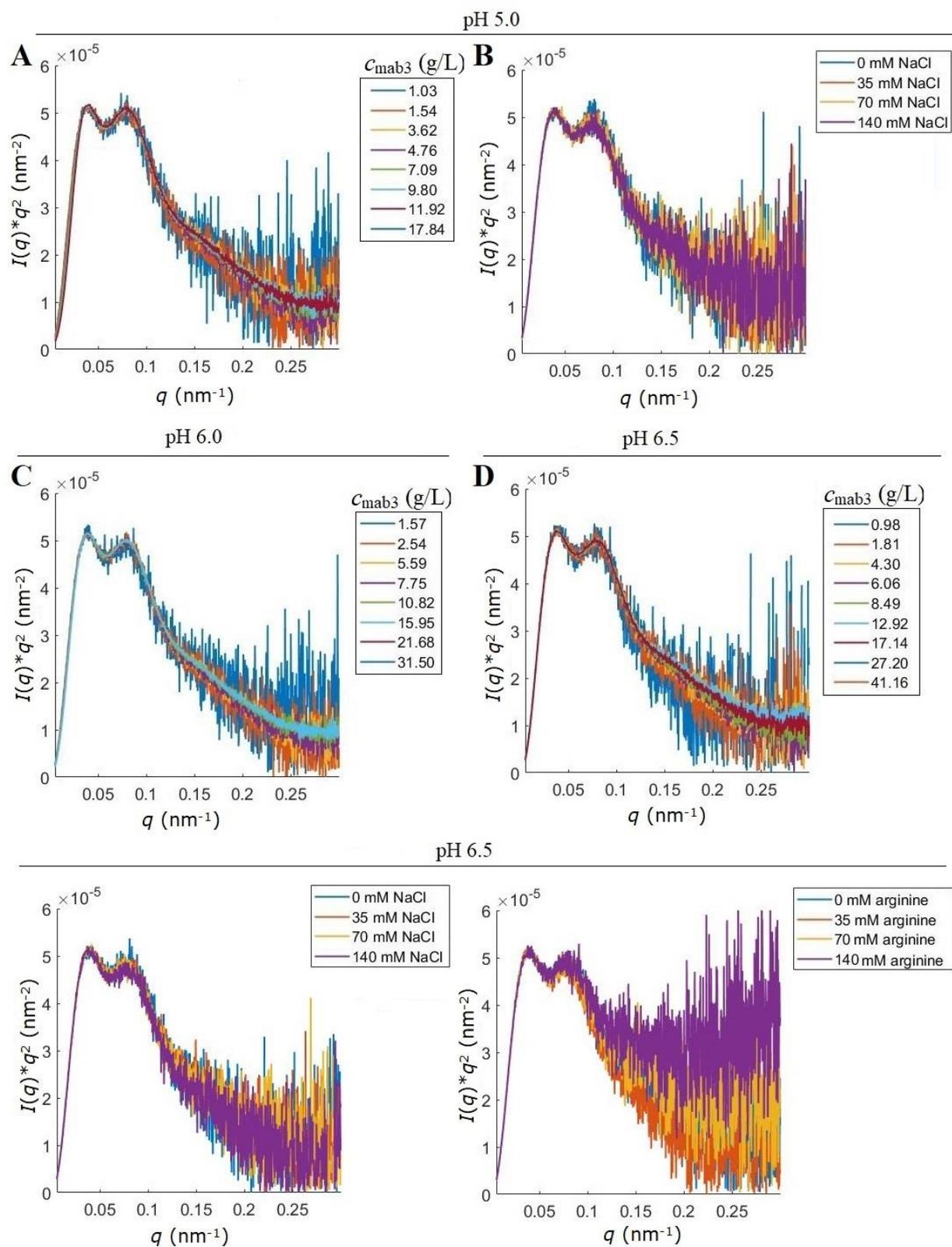


Figure S7.5: Kratky plot of a mAb₃ in 10 mM histidine buffer at A: pH 5.0, B: pH 5.0 with NaCl, C: pH 6.0, D: pH 6.5, E: pH 6.5 with NaCl, and F: pH 6.5 with arginine with c_{mAb} around 1 g/L.

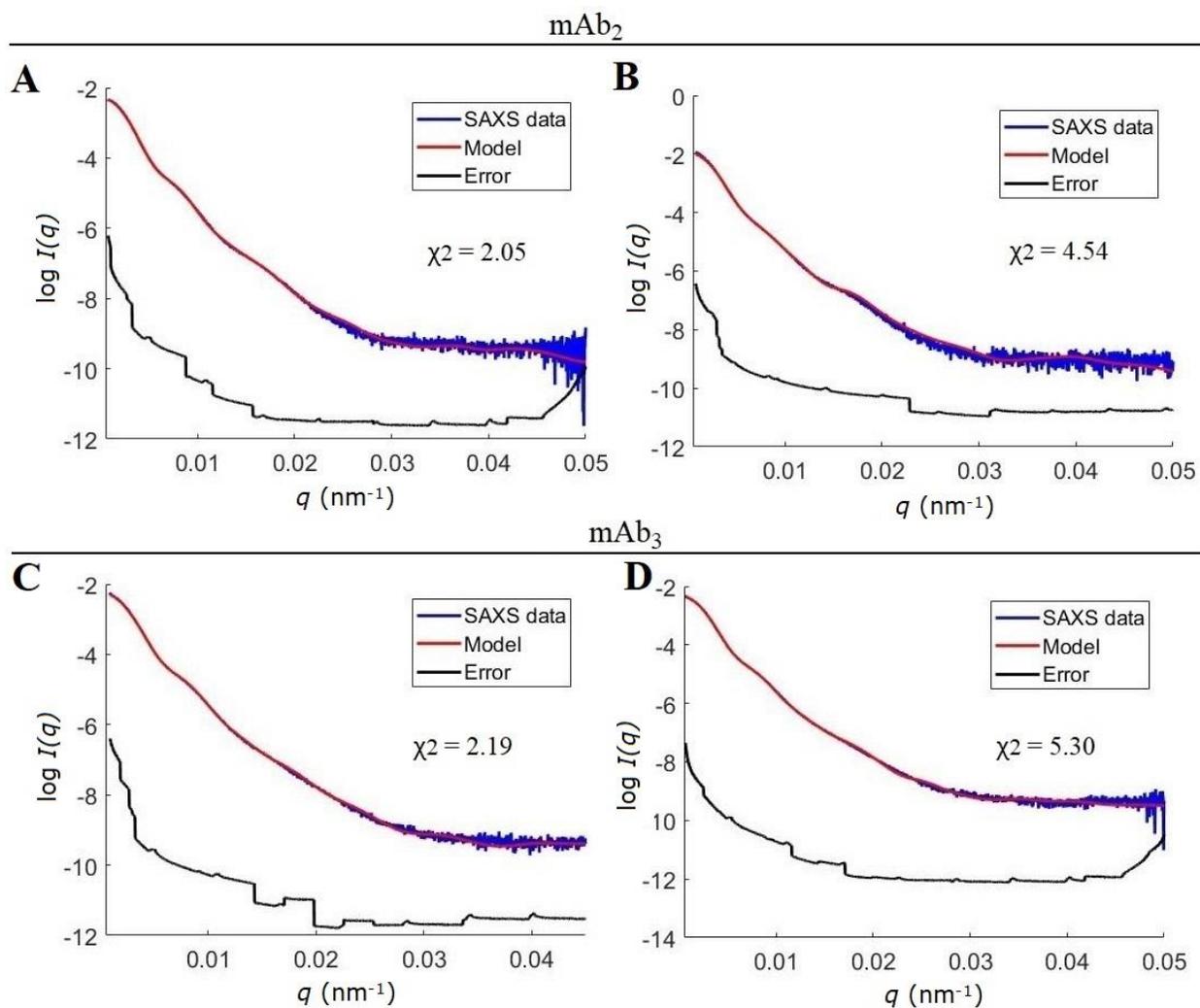


Figure 7.17: Fit plots for rigid body models from *CORAL* to the experimental data. Rigid body models in 10 mM histidine for mAb₂ at A: pH 5.0 and B: pH 6.5; and for mAb₃ C: pH 5.0 and D: pH 6.5.

Sequences:

mAb₂

Heavy chain

QVQLQESGPGLVKPSSETLSLTCTVSGGSISADGYYSWIRQPPGKGLEWIGSLYYSGSTYYNPSLKGRVTIS
GDTSKNQFSLKLSSVTAADTAVYYCARTPAYFGQDRDFFDVWGRGTLTVSSASTKGPSVFPLAPSSKST
SGGTAALGCLVKDYFPEPVTVSWNSGALTSKVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPS
NTKVDKRVEPKSCDKTHTCPPCPAPEFEGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWY
VDGVEVHNAKTKPREEQY^NSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPASIIEKTISKAKGQPREPQV
YTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQ
GNVFSCSVMEALHNHYTQKSLSLSPGK

Light chain

DIQMTQSPSTLSASVGDRVTITCRASQGISSWLAWYQQKPGKAPKVLIIYKASTLESVPSRFRSGSGSGTEFT
LTISSLQPEDFATYYCQQSHHPWTFGQGTKLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAK
VQWKVDNALQSGNSQESVTEQDSKDYSLSSLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Linker EPKSCDKTHT^CPP^CPAPEFEGG

mAb₃

Heavy chain

QVQLVESGGGLVKGGLRLSCLASGFTFSYYMNWIRQAPGKGLEWVSYISSSGSIIYYADSVKGRFTISR
DNAKNSLYLQMNLSRAEDTAVYYCAREGRIARGMDVWVGQGTTVTVSSASTKGPSVFPLAPCSRSTSEST
AALGCLVKDYFPEPVTVSWNSGALTSKVHTFPAVLQSSGLYSLSSVTVPSSNFGTQTYTCNVDHKPSNTK
VDKTVKRCCKVECP^CPP^CAPPVAAGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVQFNWYVDGVETK
PREEQFNSTFREEQFNSTFRVVSVLTVHVDWLNKEYKCKGLPAPIEKTISKTKGQPREPQVYTLPPSREE
MTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFSCSV
MHEALHNHYTQKSLSLSPGK

Light chain

DIQMTQSPSSLSASVGDRVTITCRPSQSFSRYYINWYQQKPGKAPKLLIYAASSLVGGVPSRFRSGSGSGTDFTL
TISSLQPEDFATYYCQQYTSNPPITFGQGRLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKV
QWKVDNALQSGNSQESVTEQDSKDYSLSSLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Linker RKCC^CVE^CPP^CAPPVA

N-glycosylation site

^C – cysteines that form S-S bond