Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

Jephthah, Stephanie

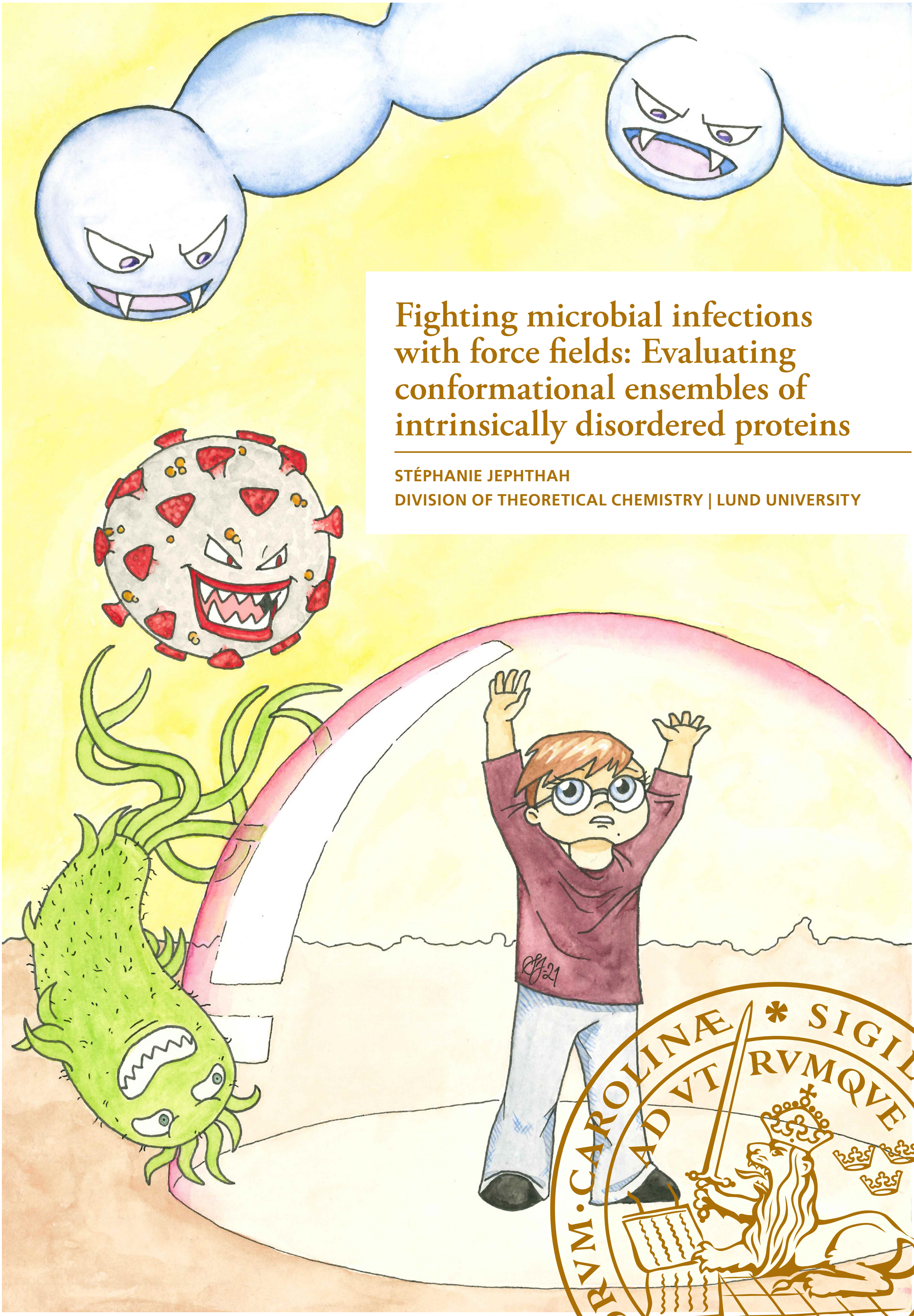
2021

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

Link to publication

Citation for published version (APA):
Jephthah, S. (2021). *Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins*. Lund University.

Total number of authors:
1



Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

STÉPHANIE JEPHTHAH

DIVISION OF THEORETICAL CHEMISTRY | LUND UNIVERSITY

LUND
UNIVERSITY

ISBN: 978-91-7422-806-9

Theoretical Chemistry
Department of Chemistry
Faculty of Science
Lund University

9 789174 228069

Printed by Media-Tryck, Lund 2021 NORDIC SWAN ECOLABEL 3041 0903

Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

by Stéphanie Jephthah

LUND
UNIVERSITY

DOCTORAL THESIS

by due permission of the Faculty of Science, Lund University, Sweden. To be defended on Friday 28th of May 2021 at 13.00 in lecture hall A at Kemicentrum.

Faculty opponent
Dr. Robert B. Best, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.

DOKUMENTDATABLAD enl SIS 61 41 21

Organization **LUND UNIVERSITY** Kemicentrum Box 124 SE–221 00 Lund, Sweden	Document name **DOCTORAL THESIS**
	Date of disputation 2021-05-28
	Sponsoring organization NordForsk's Nordic Neutron Science Programme (grant number 81912)
Author(s) Stéphanie Jephthah	

Title and subtitle

Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

Abstract

The main goal of this compilation thesis has been to investigate the conformational ensemble of the intrinsically disordered protein (IDP) histatin 5 by using a mixture of experimental and computational techniques including, but not limited to, small-angle X-ray scattering (SAXS), circular dichroism (CD) spectroscopy, atomistic molecular dynamics (MD) simulations, and coarse-grained Monte Carlo (MC) simulations. Histatin 5 is a peptide of particular interest for two reasons. Firstly, histatin 5 is naturally found in the saliva, where it helps protect the teeth enamel from abrasion and degradation by being part of the dental pellicle. Additionally, it also possess candidacidal properties and thus acts as a first line of defense against fungal infections. Therefore, by studying the conformational ensemble histatin 5, we hope to aid in the development of saliva substitutes and new antimicrobial agents. Secondly, histatin 5 acts as a typical IDP that is relatively easy to work with in the lab, which makes it a perfect model IDP for evaluating different MD force fields and other simulation techniques.

In the first study of this thesis, simulations were used to investigate how the conformational ensemble of histatin 5 is affected by protonation of its histidine residues. The study also continues to investigate how this affects the interaction between histatin 5 and a negatively charged surface. This study is followed by an investigation of how histatin 5 is affected by temperature, and of how well modern simulation methods used for simulating IDPs can capture temperature-induced conformational changes. The results of this study led to another project, in which four different force fields were evaluated based on their ability to simulate histatin 5 and four other peptides that are known to possess polyproline II structure. In the final histatin 5 study, a shorter but more potent version of histatin 5 was conjugated to the polyamine spermidine, and was then investigated using a combination of experiments and simulation techniques. The aim of this study was to investigate if the increased candidacidal effect could be connected to the conformational properties of the histatin 5-spermidine conjugate. The final study of this thesis concerned a completely different peptide, namely the intrinsically disordered region of magnesium transporter A (MgtA) found in *Escherichia coli*. This study was mainly experimental and the aim was to investigate how this intrinsically disordered region contributes to the function of MgtA, and the knowledge about how bacterial proteins work might help to develop new antibiotics in the future.

Key words

molecular dynamics, force field, Monte Carlo, coarse-grained modeling, simulation, intrinsically disordered proteins, histatin 5, small-angle X-ray scattering, circular dichroism

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language English
ISSN and key title		ISBN 978-91-7422-806-9 (print) 978-91-7422-807-6 (pdf)
Recipient's notes	Number of pages 214	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Stephanie Jephthah Date 2021-04-23

Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

by Stéphanie Jephthah

LUND
UNIVERSITY

DOCTORAL THESIS

Fighting microbial infections with force fields: Evaluating conformational ensembles of intrinsically disordered proteins

Funding information: This work is funded by NordForsk's Nordic Neutron Science Programme in the project "The structure of membrane proteins under solution conditions" (grant number 81912).

Parts of this doctoral thesis have been published before in the licentiate thesis *Two aspects of simulating disorder* by Stéphanie Jephthah (2019).

Faculty of Science, Lund University, Sweden, Kemicentrum

ISBN: 978-91-7422-806-9 (print)
ISBN: 978-91-7422-807-6 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2021

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is an environmentally
certified and ISO 14001:2015 certified
provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN

"I can feel bits of my brain falling away like a wet cake"
– Bernard Black

"Hey, you know what this calls for? A party!"
– Pinkie Pie

Preface

My path though the world of chemistry might seem to have been a straight journey, but for me it feels more like I have been stumbling on a winding path. In the end I guess it has been a little bit of both. The science route started already during my lower secondary education but became more defined during the International Baccalaureate (IB) Diploma Programme, which feels like an eternity ago, although I still have nightmares about the exams. I had grand ambitions; in the beginning I thought I might become a medical doctor, but towards the end I was more interested in forensics. I decided to enroll in the bachelor's programme in chemistry and molecular biology—that way I could still qualify for studying forensics and get a broader education at the same time. However, I realized pretty soon that I was more interested in physical chemistry than in biochemistry, at least when it came to lab work and the way of how the different subjects are studied; I did not want to learn all amino acid structures by heart—I wanted to understand the concepts and principles that make chemistry work in the way it does. After a course about surface and colloid chemistry, I was informed that I could apply for working on a summer research project at the division of physical chemistry or at the division of theoretical chemistry. *Theoretical chemistry*. What does that really mean? I wanted to know more about this mysterious subject and sent an application, which was accepted. This is where I got my first impression of theoretical chemistry. Computer simulations, programming, scripting, FORTRAN, coarse-grained modeling, supercomputers—all of which were completely new concepts for me. I was lost in a new world, but also intrigued. Then it was time for the bachelor's degree project. I had yet not decided if I wanted to do it in physical or theoretical chemistry, but when I suddenly got a project offer at theoretical chemistry I was hooked. My soul now belonged to the division of theoretical chemistry and this is where I have stayed from then until now, as I am finishing my PhD project. In total, I have spent most of the last seven years at the division of theoretical chemistry in Lund (although this thesis only comprises the work of the last four and a half years) and it is finally time to let go and continue my journey elsewhere. I have learned a lot over the past years, especially during the PhD project, and I have met so many different and interesting people. Naturally, there have been hardships on the way, but I am still here and that counts for something. In the end I am happy with how far I have come during this time, and I am proud of my progress—not only within the field of chemistry, but also on a personal level. I am also grateful to everyone who has inspired me on my chemical journey.

To my *dad*, who have always supported my science interests and educational choices, and who encourages positive thinking (Foten i kläm? Jajjamensan!).

To my first chemistry teacher, *Therese Persson*, for providing an encouraging introduction to chemistry; and to my class mentor, *Anders Östling*, who taught me to take my studies more seriously.

To my IB chemistry teachers with spookily similar names, *Gunnar Nilsson*, *Gunilla Åkesson-Nilsson* and *Helene Åkesson*, for maintaining my interest of chemistry; and to my other IB teachers, *Sten Bernhardsson* and *Karin Kulka*, for inspiring me to do science.

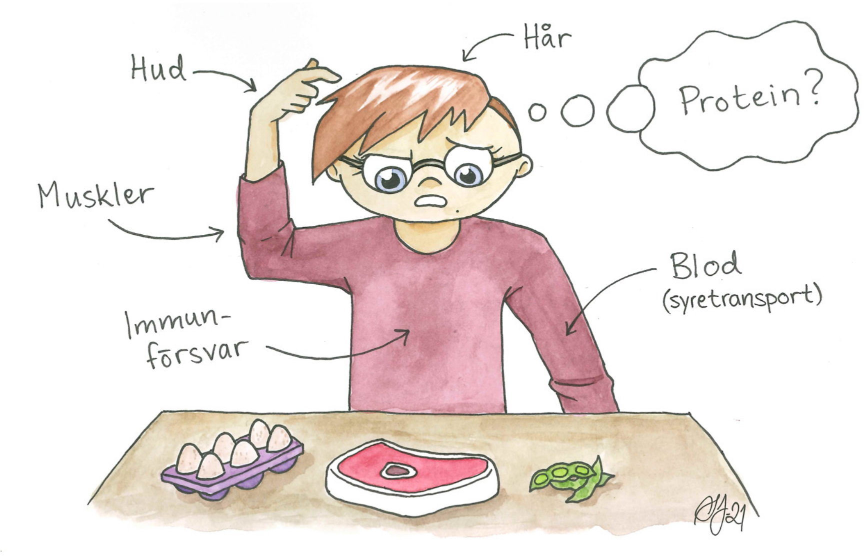
To *João Henriques* for teaching me everything I needed to know for starting my own PhD studies; and to my supervisor, *Marie Skepö*, for everything—from the start of my bachelor's degree project to the end of this PhD project.

Stéphanie Jephthah
Lund, March–April 2021

Contents

Populärvetenskaplig sammanfattning på svenska

Vad är ett protein? För någon som inte är särskilt insatt i kemi kanske ordet för tankarna till gymmet, där hurtbullarna häver i sig proteindrycker. Kanske går tankarna vidare till kött, ägg, soja och liknande produkter som går att hitta på tallriken. Till vardags används ordet protein främst i samband med näringsintag, och de flesta vet att det är viktigt att få i sig ordentligt med proteiner för att kunna bygga muskler. Vad folk kanske inte vet är att proteiner är inblandade i så mycket mer. Proteiner finns överallt i kroppen, och inte bara som byggstenar till våra organ. De hjälper även till med att bryta ner den mat vi äter, transportera syre i blodet, och skydda oss från sjukdomar.



Figur 1: Proteiner är inte bara näringsämnen. Proteiner är inblandade i nästan allt i kroppen: de bygger upp många olika typer av vävnader, transporterar syre och många andra viktiga atomer, hjälper till att bryta ner och omvandla molekyler, samt tar del av kroppens immunförsvar på många olika sätt.

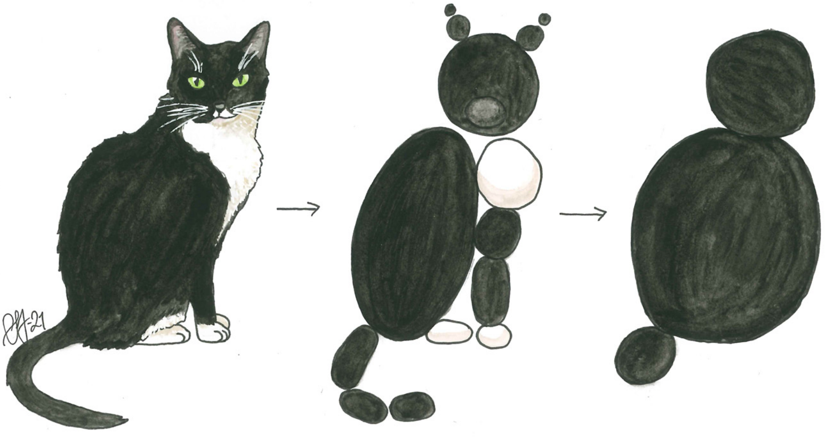
Proteiner kan liknas vid långa kedjor där varje länk är en mindre molekyl som kallas för en aminosyra. Det finns drygt tjugo olika aminosyror, och de är alla olika. Beroende på vilka aminosyror som finns i ett protein, och i vilken ordning de sitter, får proteinet specifika egenskaper. Bland annat är det vanligt att proteinet viker sig och på så sätt får en unik 3D-struktur som kan vara betydande för att proteinet ska kunna utföra sina funktioner i kroppen. Under en längre tid utgick forskare från att det var nödvändigt för proteiner att ha en bestämd 3D-struktur för att ha någon biologisk funktion. Sedan dess har vi insett att även proteiner som inte bildar sådana strukturer har viktiga funktioner i kroppen. Sådana proteiner är ofta väldigt flexibla och ändrar sin struktur konstant, och kallas därför för oordnade proteiner. Tyvärr är det också vanligt att oordnade proteiner är inblandade i otäcka sjukdomar som till exempel Parkinsons och

Alzheimers. Eftersom de oordnade proteinerna finns överallt och både kan hjälpa och stjälpa i kroppen, är det viktigt att vi lär oss hur de fungerar. Om vi kan förstå hur ett hjälpsamt protein fungerar kan vi kanske använda det i läkemedel. På likande sätt, om vi förstår hur ett skadligt protein fungerar, kan vi komma på andra läkemedel som förhindrar skadorna.

Alla experimentella metoder som vanligtvis används för att bestämma strukturen av ett protein fungerar inte för de oordnade proteinerna eftersom de byter struktur hela tiden. Därav blir man begränsad till metoder som enbart ger medelvärden av olika egenskaper som har med strukturen att göra. Till exempel går det att avgöra om de oordnade proteinerna i snitt tenderar att vara mer utsträckta eller om de hellre håller sig lite lätt vikta. Som ett komplement till att använda experimentella metoder för att studera oordnade proteiner kan man använda sig av datorsimuleringar. Med hjälp av datorsimuleringar går det att se exakt vilka strukturer proteinet växlar mellan, samtidigt som man också kan få ut samma medelvärden som de experimentella metoderna. Simuleringar kan även förutse vissa trender—till exempel hur ett protein påverkas av salthalt, pH och temperatur—vilket gör att man slipper göra vissa onödiga och dyra experiment.

Så varför gör man inte alltid simuleringar? För det första måste man vara säker på att simuleringsmetoden ger samma resultat som de experimentella metoderna. Det är relativt lätt att simulera ett vanligt protein i lösning, men ett oordnat protein fungerar annorlunda, så när de började simuleras vek dem sig alldeles för mycket och gav inte samma resultat som experimenten. Sedan dess har simuleringsmetoderna stegvis förbättrats, men ofta fungerar simuleringsmetoderna bara för ett snävt val av system. Det finns ytterligare ett problem med simuleringar: oftast går det lätt att simulera en ensam proteinkedja, men en ensam proteinkedja representerar inte alltid verkligheten. I ett naturligt system kan det finnas många andra saker i närheten av proteinet; till exempel kan det finnas fler proteiner, andra proteiner och molekyler, ytor, cellmembran, etc. Att ta hänsyn till allt detta gör simuleringarna extremt mycket mer komplicerade och tidskrävande, och därför är det i dagsläget orealistiskt att göra så komplicerade simuleringar på en större skala. Lyckligtvis finns det genvägar, bland annat genom att förenkla systemet som ska simuleras. En vanlig förenkling är att göra systemet grovkornigt. Det innebär att man antar att en grupp av atomer kan representeras av en större partikel. För grovkorniga proteiner är det vanligt att varje aminosyra representeras av en sfär, vilket gör att proteinet ser ut som ett pärlhalsband. Därav behöver man bara ta hänsyn till hur sfärerna växelverkar istället för hur alla atomer i systemet växelverkar. Att förenkla systemet på detta sätt ger såklart inte exakta strukturer men istället erhålls annan användbar information som inte går att få tag på om man enbart simulerar en ensam proteinkedja.

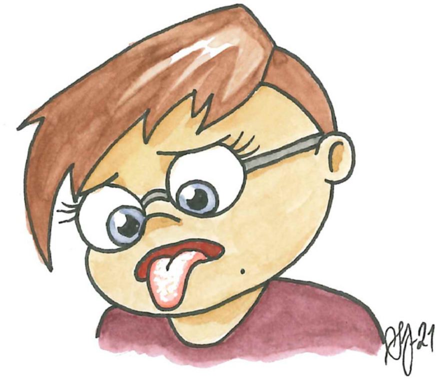
I den här avhandlingen har jag främst arbetat med att förstå hur det oordnade proteinet



Figur 2: Om jag skulle försöka göra min katt, Otto von Bismarck, grovkornig skulle det kunna gå till som på den här bilden. Det finns olika nivåer av grovkornighet; till exempel kan man göra något bara lite grovkornigt som i mittenbilden, där man fortfarande kan ana att det ska föreställa en katt. Annars kan man göra en mycket grövre approximation som i den högra bilden. Med en så pass grovkornig version är det omöjligt att avgöra om det är en katt, en hund, eller något djur över huvud taget.

histatin 5 fungerar genom att studera dess strukturella egenskaper. Histatin 5 finns vanligtvis i vår saliv, där det hjälper till att skydda tänderna genom att tillsammans med andra proteiner bilda en hinna på emaljen. Utöver det skyddar histatin 5 även mot svampinfektioner som till exempel torsk. På grund av dess intressanta biologiska egenskaper är det av intresse att förstå hur histatin 5 fungerar för att kunna optimera dess effekt och för att kunna utveckla nya och bättre läkemedel i framtiden. Histatin 5 är även intressant ur ett simuleringsperspektiv eftersom det är ett litet protein, vars strukturella egenskaper är väldigt typiska för oordnade proteiner. Därför är det fördelaktigt att använda just histatin 5 som ett modellprotein när man testar simuleringsmetoder som används specifikt för oordnade proteiner.

Till att börja med har vi studerat histatin 5 på en lite mer grundläggande nivå, genom att undersöka hur dess struktur ändras beroende på hur laddat det är. Vidare undersökte vi hur bra moderna simuleringsmetoder kan förutsäga de temperaturberoende strukturförändringar som sker för histatin 5. Resultaten av den studien gjorde oss nyfikna på att undersöka hur mycket de strukturella egenskaperna skiljs åt beroende på vilken simuleringsmetod som används. Detta gjordes inte bara för histatin 5, utan även för några andra små proteiner. Den sista studien av histatin 5 i den här avhandlingen handlar egentligen om en kortare variant av histatin 5 som är bunden till en annan molekyl för att öka den svampdödande effekten som histatin 5 har. Studien gick ut på att



Figur 3: Utan proteiner som histatin 5 i saliven får man lätt svampinfektioner. Om du drabbas av candidiasis, eller torsk som det också kallas, kan du få en vit beläggning på tungan. Hos väldigt sjuka personer, eller hos personer med nedsatt immunförsvar kan candidiasis sprida sig till blodet eller till och med benvävnaden.

försöka förstå hur strukturen av den här histatin 5-varianten bidrar till den förbättrade effekten. Avslutningsvis har vi även undersökt ett helt annat oordnat protein som egentligen är en del av ett större protein som finns i bakterier. Det är vår förhoppning att om vi kan förstå de strukturella egenskaperna hos detta protein, kan det i längden leda till nya alternativ för antibiotika.

Men vad har allt detta med kraftfält att göra? Det står “force fields” i titeln, och det betyder ju kraftfält. Dessutom är det en bild av ett kraftfält på avhandlingens omslag. Tyvärr är kraftfälten i den här avhandlingen inte lika flashiga som i science fiction. För vissa typer datorsimuleringar behövs information om alla olika parametrar och variabler som behövs för att beräkna krafterna som växelverkar mellan alla atomer i systemet—det är samlingen av all denna information som kallas för ett kraftfält. Det är på så sätt använder jag kraftfält i kampen mot infektioner.

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **Structural Characterization of Histatin 5–Spermidine Conjugates: A Combined Experimental and Theoretical Study**

S. Jephthah, J. Henriques, C. Cragnell, S. Puri, M. Edgerton, M. Skepö
Journal of Chemical Information and Modeling, 2017, 57(6), pp. 1330–1341

II **Temperature dependence of IDPs in simulations: What are we missing?**

S. Jephthah, L. Staby, B. B. Kragelund, M. Skepö
Journal of Chemical Theory and Computation, 2019, 15(4), pp. 2672–2683

III **Physicochemical Characterisation of KEIF – The Intrinsically Disordered N-Terminal Region of Magnesium Transporter A**

S. Jephthah, L. K. Månsson, D. Belić, J. P. Morth, M. Skepö
Biomolecules, 2020, 10(4), pp. 623

IV **Force field effects in simulations of proteins with varying polyproline II propensity**

S. Jephthah, F. Pesce, K. Lindorff-Larsen, M. Skepö
Manuscript

V **The effect of histidine protonation on the conformational ensemble in simulations of histatin 5**

S. Jephthah, M. Polimeni, M. Lund, M. Skepö
Manuscript

All papers are reproduced with permission of their respective publishers.

Publications not included in this thesis:

Spontaneous Formation of Cushioned Model Membranes Promoted by an Intrinsically Disordered Protein

Y. Gerelli, A. E. Skog, **S. Jephthah**, R. J. L. Welbourn, A. Klechikov, M. Skepö
Langmuir, 2020, 35(15), pp. 3997–4004

Author contributions

Paper I: Structural Characterization of Histatin 5–Spermidine Conjugates: A Combined Experimental and Theoretical Study

I planned and initiated the project together with my co-authors. The simulations and the SAXS experiments, as well as the analysis of the data, were done with assistance from my co-authors. I wrote the manuscript together with my co-authors, and was involved in the submission and revision process.

Paper II: Temperature dependence of IDPs in simulations: What are we missing?

I planned the project together with my supervisor. I performed and analyzed the simulations, as well as the SAXS and the CD experiments. I wrote the manuscript with input from my co-authors, and was responsible for the submission and revision process.

Paper III: Physicochemical Characterisation of KEIF – The Intrinsically Disordered N-Terminal Region of Magnesium Transporter A

I planned and initiated the project together with my co-authors. I performed and analyzed the simulations, as well as the SAXS experiments. I wrote the manuscript together with my co-authors, and was responsible for the submission and revision process.

Paper IV: Force field effects in simulations of proteins with varying polyproline II propensity

I planned and initiated the project together with my supervisor. I performed and analyzed the simulations together with my co-authors. I wrote the first draft of the manuscript and edited the manuscript together with the co-authors.

Paper V: The effect of histidine protonation on the conformational ensemble in simulations of histatin 5

I planned and initiated the project together with my supervisor. I performed the MD simulations and I analyzed all simulations together with my co-authors. I wrote the first draft of the manuscript and edited the manuscript together with the co-authors.

Acknowledgements

To begin, I express my gratitude to my supervisor *Marie Skepö* for her support and advice, and for keeping me in her research group all the way until today. I also want to express my appreciation for all former and current colleagues in our research group and at the division. Special thanks goes to *João* for providing me with a solid foundation for performing simulations, to *Carolina* for teaching me about SAXS, and to *Sandeep* for making the experimental lab work more fun. I would also like to give extra thanks to *Maria*, *Ellen* and *Junhao*, for everything we've done together during these years—courses, conferences, experiments, discussions, etc. It has been fun. To *Erik* and *Sam*, thank you for keeping my spirits up.

Additional thanks goes to everyone in my personal life that has supported me during these years, and especially during these last few weeks. To *mamma Mia* and *pappa Marcus*; to *mormor Anette*, *farmor Lisbeth*, and *Margaretha*; to my siblings *Mathilda*, *Ottilia*, *Saga*, *Ella*, *Dante*, *Vannah*, and *Emma*; to my nerd friends, *Irene*, *David*, *Viktoria*, *Ben*, *Niklas*, *Pernilla*, *Hans*, and *Arvid*; and to my closest friends, *Jonas*, *Zoé*, *Sebastian* and *Jennifer*. I love you all.

Finally, I express my deepest gratitude to my fiancé *Martin*, for forcing me out of bed in the morning, for feeding me when I'm hangry, and for always being there for me. Du är min bästa vän och min klippa. Jag älskar dig.

1. Introduction

Summarized in this compilation thesis are four and a half years of research. The aim has been to investigate the conformational ensembles of different proteins, of which the two primary proteins are involved in microbial infections. By gaining a holistic understanding of how their conformational properties contribute to their biological functions, new alternatives for fighting microbial infections can be invented, which will advance medical research. Another aspect of the research was quickly developed by the start of this work as it was discovered that different simulation methods produced different results. For this reason, two of the studies have focused on investigation and evaluation of different simulation methods as well.

The thesis begins with a general background that introduces several basic concepts, as well as the main components that have been investigated throughout these past years. The aim of this part is to provide the reader with a basic understanding of important concepts, as well as to emphasize the importance of the study. Following the general background is a theoretical background that explains the more complicated theories that are needed to understand the underlying physics and chemistry that form the basis of the methods, and that are needed to correctly interpret the results. Next, the two different models that have been used in the simulations are presented, followed by the simulation methods. Connected to these sections are the simulation analyses that are subsequently presented. A few experimental techniques have also been utilized to complement the simulation results, and the background of these techniques, as well as different approaches for analyzing and interpreting their results, are illustrated in a section of its own. Finally, the actual research is summarized and discussed, followed by the research papers and the manuscripts.

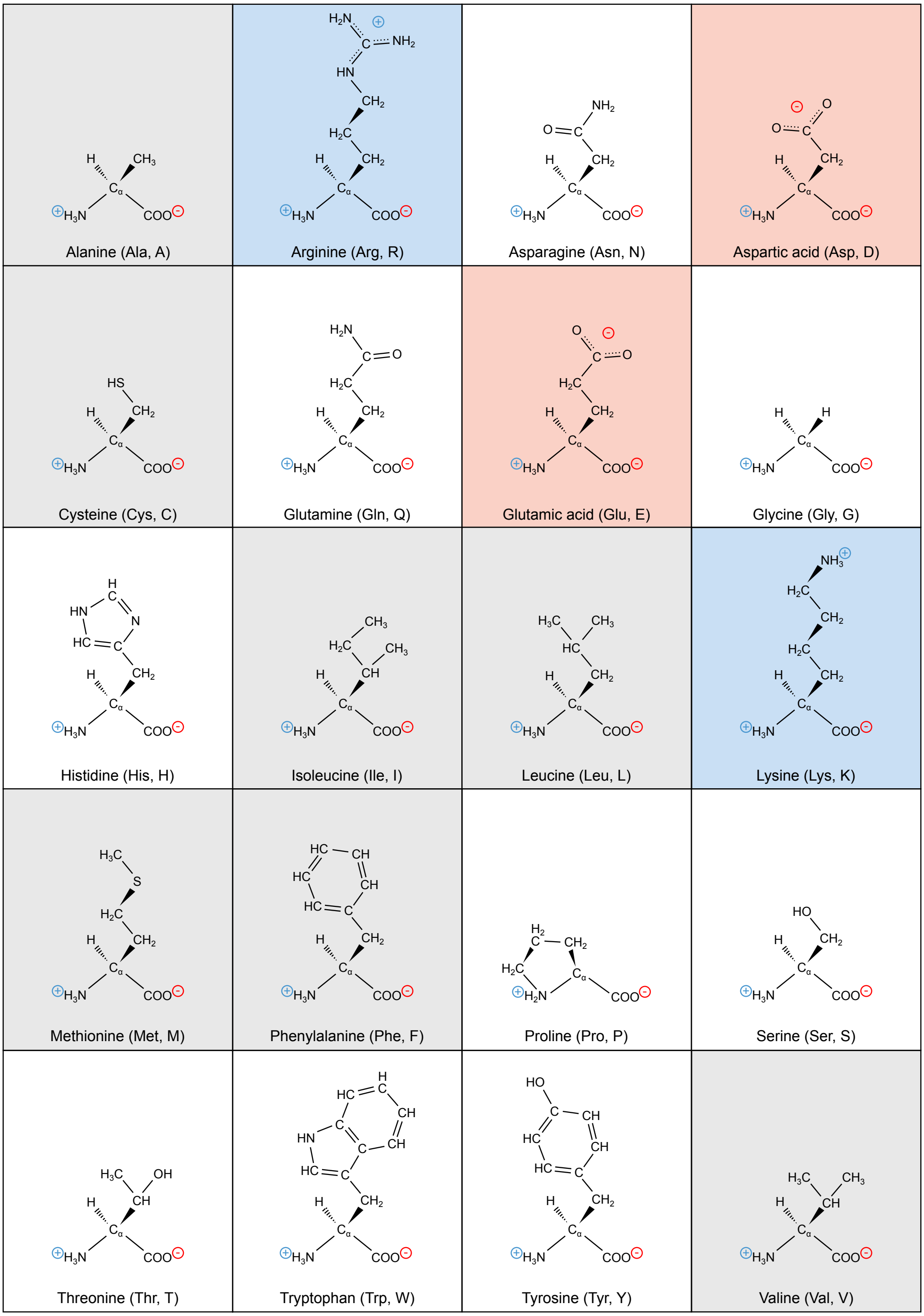
2. Background

2.1 Proteins

Proteins are biological macromolecules that are built up by subunits called *amino acids*. There are several types of amino acids, but all amino acids have the same base structure, consisting of a central carbon known as the α-carbon, which is linked to an amino group (–NH_2), a carboxylic acid group (–COOH), a hydrogen atom (H), and a side chain (R) that is unique for each individual amino acid. At neutral pH, most amino acids exist as so-called *zwitterions*, which means that the amino group is protonated (–NH_3^+) and the carboxylic acid group is deprotonated (–COO^-). The structure of the twenty most common amino acids, as well as their respective 3- and 1-letter codes, are depicted in Figure 2.1.

Because the amino acids are designed so that there are four different groups connected to the central α-carbon, all amino acids (except glycine) are *chiral*. Chiral molecules are their own mirror images that can never be superimposed onto each other. A more relatable analogue would be our hands, which are mirroring each other but can never look exactly the same no matter how we turn them. The two chiral forms of amino acids are called the L isomer and the D isomer, but proteins are exclusively comprised of L amino acids [2]. Figure 2.2a-b depicts the amino acid structure of both chiral forms at different sides of a mirror plane.

Amino acids can bond together through *peptide bonds* to form longer chains known as *polypeptides* (also commonly referred to as peptides). If the polypeptide is long enough it is called a protein. The continuous chain structure of proteins is comprised of the α-carbons, the nitrogen atoms of the amino groups, and the carbon atoms of the carboxylic acid groups, and is commonly referred to as the protein *backbone*. Illustrations of the peptide bond and the protein backbone are depicted in Figure 2.2c. The order of the amino acids in a protein is referred to as the amino acid sequence, which constitutes the *primary structure* of the protein, see Figure 2.3a. Hydrogen bonds



Alanine (Ala, A)
Arginine (Arg, R)
Asparagine (Asn, N)
Aspartic acid (Asp, D)
Cysteine (Cys, C)
Glutamine (Gln, Q)
Glutamic acid (Glu, E)
Glycine (Gly, G)
Histidine (His, H)
Isoleucine (Ile, I)
Leucine (Leu, L)
Lysine (Lys, K)
Methionine (Met, M)
Phenylalanine (Phe, F)
Proline (Pro, P)
Serine (Ser, S)
Threonine (Thr, T)
Tryptophan (Trp, W)
Tyrosine (Tyr, Y)
Valine (Val, V)

Figure 2.1: The structure of the individual amino acids at physiological pH. Their 3- and 1-letter codes are included within parenthesis after their names. Amino acids with a net positive charge ($+1e$) are shown with blue background, and amino acids with a net negative charge ($-1e$) are shown with red background. Hydrophobic amino acids (according to Kyte & Doolittle (1982) [1]) are shown with gray background.

between the amino hydrogens and the carboxylic oxygens in the protein backbone can cause local folding of the protein chain. These local rearrangements make up the *secondary structure* of the protein. There are several different types of secondary structures, depending on the relative positions of the amino acids in the sequence, and they can be defined by either (i) the hydrogen bonding pattern, or (ii) the backbone dihedral angles (ϕ, ψ). Examples of secondary structure elements are the α-helix (Figure 2.3b) and the β-sheet (Figure 2.3c). Secondary structure can be visualized in a so-called Ramachandran plot, where the dihedral angles of each amino acid residue of a protein is plotted against each other. A protein can also fold into a fixed three-dimensional structure, called the *tertiary structure*, see example in Figure 2.3d. Spatial arrangement of several protein chains into larger structures or complexes is called a *quaternary structure*. A typical example of a quaternary structure is that of hemoglobin, which consists of four protein subunits, see Figure 2.3e.

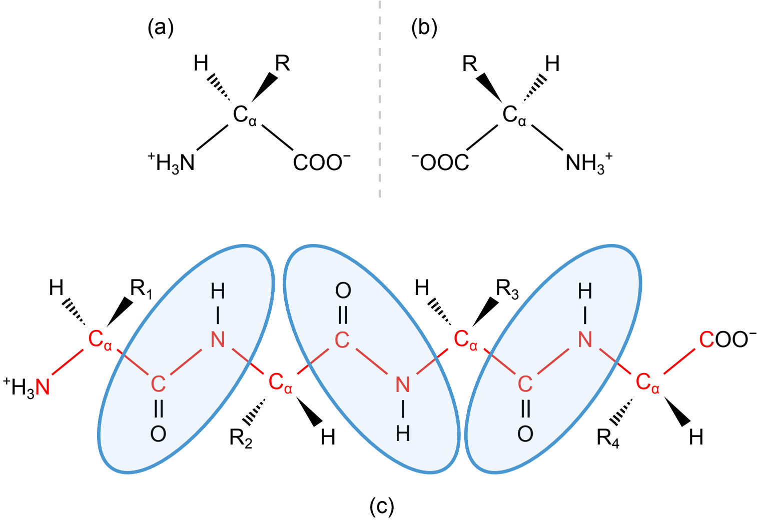


Figure 2.2: The general structure of an amino acid at neutral pH shown as (a) the L isomer and (b) the D isomer at different sides of a mirror plane. (c) A schematic polypeptide illustrating the peptide bonds (blue areas) and the protein backbone (red).

Because proteins can be created of all different sizes, and with all imaginable combinations of primary structure, they are considered to be the most versatile macromolecules. Apart from being essential nutrients and building blocks for most tissue, proteins are also responsible for a great variety of important functions in the body, such as catalysis,

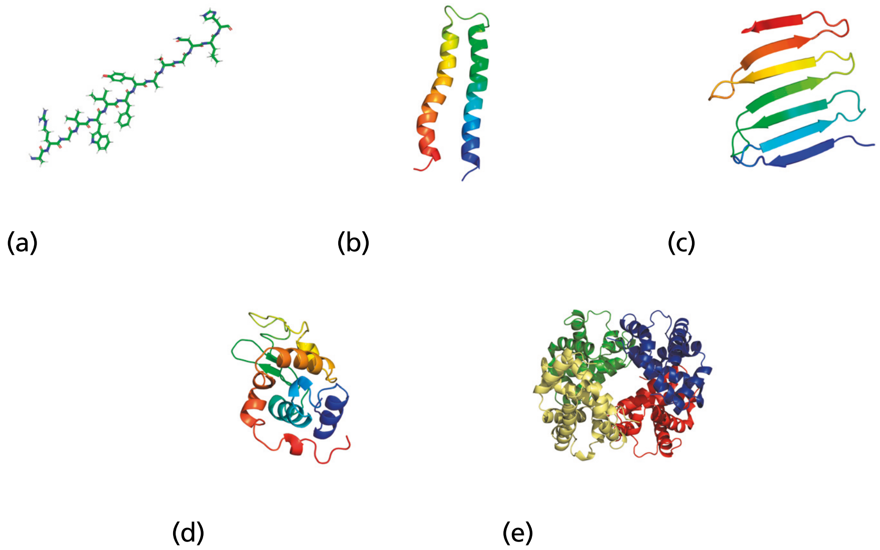


Figure 2.3: Examples of different levels of protein structure. (a) Primary structure representation, the secondary structures (b) α-helix[1], (c) β-sheet[2], (d) tertiary structure[3], and (e) quaternary structure with four different subunits[4]. All structures except for the primary structure were obtained from RCSB PDB (`https://www.rcsb.org`) [3] and are shown in cartoon representation. Some structures have been modified for visual enhancement.

molecule transport, protection, signaling, and regulation to name a few.

2.2 Intrinsically disordered proteins

For a long time, it was considered a requirement for a protein to have a well-defined folded structure in order for it to have any biological function. Nevertheless, during the last 20 years, plenty of evidence has emerged revealing proteins and protein regions that challenge this structure–function paradigm [7]. This class of proteins and protein regions are commonly referred to as *intrinsically disordered proteins* (IDPs) and *intrinsically disordered regions* (IDRs). These IPDs and IDRs are characterized by their extreme flexibility and lack of tertiary structure in aqueous solution. Additionally, they often lack secondary structure, have low hydrophobicity, and a high net charge [8, 9]. Despite being described as "unstructured", IDPs do not have to completely lack sec-

---

[1] 2I7U; Ma et al. (2006) Structural and dynamical analysis of a four-alpha-helix bundle with designed anesthetic binding pockets; doi: 10.2210/pdb2i7u/pdb [4].

[2] 6J48; Makabe and Hori (2019) Glycine mutation on single layer beta-sheet of OspAsm1; doi: 10.2210/pdb6j48/pdb [5].

[3] 1AKI; Carter et al. (1997) The structure of the orthorhombic form of hen egg-white lysozyme at 1.5 Angstroms resolution; doi: 10.2210/pdb1aki/pdb.

[4] 1A3N; Tame and Vallone (1998) Deoxy human hemoglobin; doi: 10.2210/pdb1a3n/pdb [6].

ondary structure. In fact, it has been discovered that several "unstructured" proteins possess a degree of local order in the form of the lesser known *left-handed poly-*L*-proline type II (PPII) helix* [10–13]. The PPII helix is an extended secondary structure with dihedral angles of $(\phi, \psi) = (-75°, +145°)$, a helical pitch of 9.3 Å/turn, and 3.0 residues/turn [14, 15].

The amino acid content of IDPs/IDRs has been shown to differ significantly from that of structured proteins. For example, the IDPs usually possess a low content of "order-promoting" amino acids, as for example the hydrophobic Ile, Leu, Val, as well as the aromatic Trp, Tyr, and Phe [9, 16]. Cys is also uncommon in IDPs. Instead, the content of "disorder-promoting" amino acids is much higher for IDPs. Disorder-promoting amino acids are often small or charged, with examples including Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys [9, 16]. The structures of the order-promoting and the disorder-promoting amino acid are depicted in the previous Figure 2.1.

It is possible to predict the conformational properties of IDPs based on the number of charged residues in their amino acid sequence [17–19]. To predict the conformation, it is necessary to know the *fraction of charged residues* (FCR), and the *net charge per residue* (NCPR):

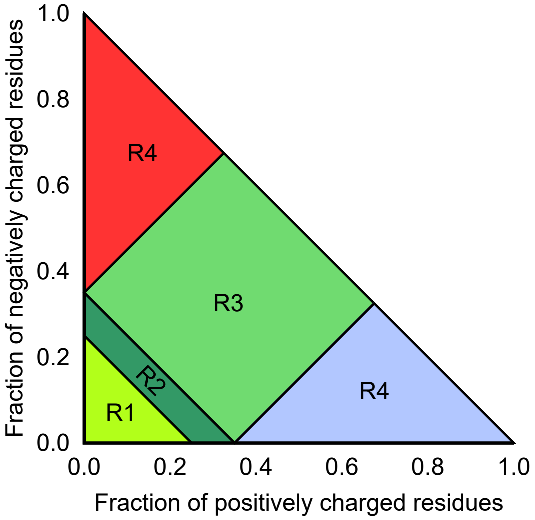
$$\mathrm{FCR} = f_+ + f_-, \tag{2.1}$$

$$\mathrm{NCPR} = |f_+ - f_-|, \tag{2.2}$$

where f_+ and f_- are defined as the fraction of positively and negatively charged residues, respectively. Depending on the combined values of the FCR and the NCPR, the protein is assigned to one of four predefined conformational classes, see Figure 2.4.

A common method for determining protein structure is *X-ray crystallography*. However, because of their extreme flexibility, IDPs are nearly impossible to crystallize [20]. Instead, other methods such as *circular dichroism* (CD), *Förster resonance energy transfer* (FRET), *nuclear magnetic resonance* (NMR), and *small-angle X-ray scattering* (SAXS) are more appropriate for obtaining structural information of IDPs [21–23]. Results obtained from these experiments capture averages of large ensembles of disordered conformations. Different types of radii, size distributions, general shapes, and secondary structure elements are a few examples of properties that can be obtained from experimental measurements by using these methods.

As previously stated, IDPs and IDRs are important for a variety of biological functions. A few examples are signaling, regulation, and recognition [24]. Because of their flexibility and ability to assume several different conformations, the same IDP can have several different functions in the body. While many IDPs/IDRs are essential for the body to function normally, it has been revealed that certain IDPs are involved in common diseases, such as Parkinson's disease, Alzheimer's disease, human papillomaviruses (HPVs), as well as several types of cancer [7, 24].



Class	FCR	NCPR	Expected conformation
R1	< 0.25	< 0.25	Globule
R2	$0.25-0.35$	≤ 0.35	Mix of globule and coils
R3	> 0.35	≤ 0.35	Polyampholytic coils or hairpins
R4	> 0.35	> 0.35	Polyelectrolytic semi-flexible rods or coils

Figure 2.4: Conformational classes of IDPs based on the FCR and the NCPR of the amino acid sequence according to Das et al. (2015) [18].

2.3 The protective functions of saliva

Saliva is the natural fluid that is present in the oral cavity. While the function of saliva may seem to be mainly lubricative and digestive, saliva also has several important functions involved in maintaining the oral health. Although saliva consists of more than 99 % water [25], the remaining constituents are also essential for the function of saliva. Less than 1 % of the saliva is constituted by proteins that are responsible for numerous different protective functions of saliva. For example, the lubricative function of saliva would be significantly less effective without the proteins known as *mucins* [25,26]. Additionally, by forming a so-called dental pellicle on the enamel surface, several proteins, including mucins, *proline-rich proteins* (PRPs), *statherin*, and *histatins*, aid in protecting the teeth from abrasion and acid degradation [25, 26]. Interestingly enough, these particular saliva proteins are also classified as IDPs. Several of these IDPs are involved in the remineralization process of the teeth (PRPs and histatins), and in the inhibition of mineral precipitation and crystal formation (PRPs and statherin) [25, 26]. A less trivial function of saliva is its antimicrobial action, which protects against microbial infections. This function also arises because of different proteins that are present in the saliva, with examples including *lysozyme*, *lactoferrin*, *peroxidases*, and histatins [25–29]. An overview of the different saliva functions and the responsible proteins is illustrated in Figure 2.5.

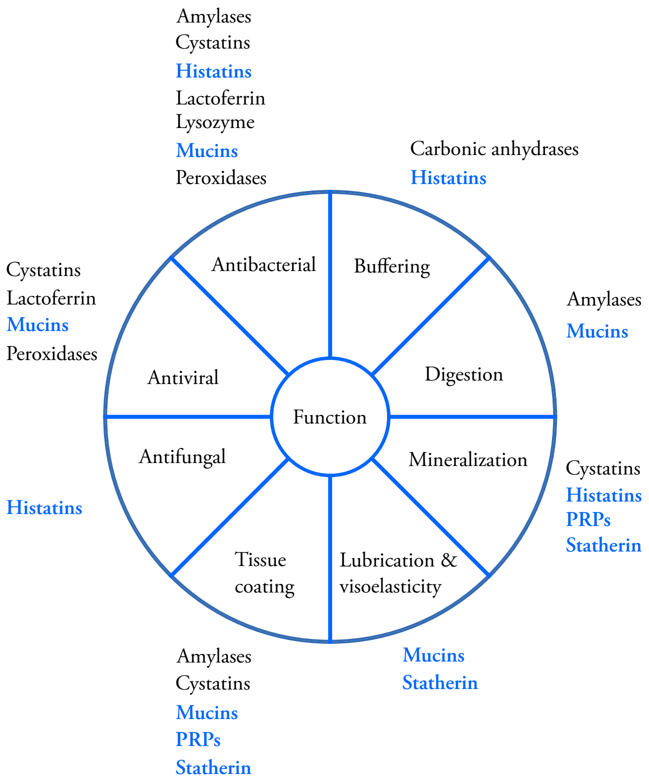


Figure 2.5: An overview of the different saliva functions and their responsible proteins. Proteins that are classified as IDPs are marked in blue. The figure is adapted from Levine (1993) [30].

2.4 Xerostomia

A change in saliva composition or a reduction of saliva flow can lead to *xerostomia*, which is also known as *dry mouth*. Xerostomia is a very common syndrome that affects up to 46 % of the human population [31]. Apart from the sensation of dryness, other symptoms may include a difficulty in swallowing, burning mouth syndrome, loss or change of taste, and increased thirst [31–35]. Additionally, a lack of saliva causes the body to become more susceptible to tooth decay (dental caries) and oral candidiasis (thrush) [32, 35]. A plentitude of different factors can cause xerostomia. For example, it can be caused by stress, anxiety, and depression [36, 37], and it is a very common side effect of many prescription drugs, including antihistamines and antidepressants [37–39]. It can also be caused by radiation therapy in the head and neck region, which is a common treatment for some specific types of cancer [40]. Furthermore, xerostomia is a common symptom of several systematic diseases, examples including Sjögren's syndrome, human immunodeficiency virus (HIV), cystic fibrosis, and diabetes [31, 32, 34, 35, 41–43]. There is no general cure for xerostomia; the treatment is instead focused on relieving the symptoms. This can be done by avoiding drinking alcohol and smoking tobacco, by avoiding acidic and salty food and beverages, and by using salivary flow stimulants (for example sugar-free gum and hard candies) [32, 42, 44–46]. There are also saliva substitutes and oral lubricants available in the form of solutions, gels and mouth sprays [32, 45, 47]. More intensive treatment is usually long-term and demanding [42]. Further information regarding xerostomia can be found in the following references and papers therein: [31, 32, 34, 35, 42, 45].

2.5 Candidiasis

Candidiasis is a fungal infection caused by opportunistic yeast species belonging to the *Candida* genus. *Candida* spp. are part of the natural microbiota of the skin and the gastrointestinal tract [48], and are normally harmless in healthy individuals. Mucosal candidiasis can nevertheless manifest as for example thrush or vulvovaginitis [49]. These infections can normally be treated locally and with relative ease [50]. However, in immunocompromised or in critically ill individuals, *Candida* spp. can lead to life-threatening *Candida* infections known as *invasive* candidiasis. Invasive candidiasis refers to two types of infections: (i) candidemia (presence of *Candida* spp. in the bloodstream), and (ii) deep-seated tissue candidiasis such as intra-abdominal abscess (collection of infected fluid in a pocket of inflamed tissue in the abdomen), peritonitis (inflammation of the inner abdominal wall membrane), and osteomyelitis (bone infection) [51–56]. In 2015, the mortality rate of invasive candidiasis was stated to be as high as 40 % [55]. Several different *Candida* spp. can cause disease in humans, but the most prevalent species is *Candida albicans*, which by itself can account for up to 50 % of detected isolates [55, 56]. Because of the high mortality rate, there is a great interest in finding new measures for preventing and treating candidiasis. However, new antifungal resistant *Candida* spp. are constantly discovered, making the need for new antifungal alternatives even greater [56, 57].

2.6 Histatin 5 and variants

One model IDP that has been used in several studies of this work is *histatin 5*. Histatin 5 is a natural constituent of saliva, where it has several protective functions. For example, it is a part of the dental pellicle on the enamel [58] and has also shown bactericidal and fungicidal action. It is especially known for its candidacidal properties, or more specifically, its inhibition of growth and germination of *C. albicans* [59–62]. There are several types of histatins, and it has been shown that the candidacidal effect is inversely proportional to the size of the histatins [59]. Histatin 5 is the proteolytic product of histatin 3, in which it constitutes only the first 24 N-terminal amino acid residues. Thereby, histatin 5 is the most potent of the naturally occurring histatins when it comes to killing *C. albicans*.

The amino acid sequence of histatin 5 is rich in charged residues, giving the peptide a net charge of $+5e$ at physiological conditions, see Table 2.1. Additionally, $\sim$29 % of the amino acid residues are His residues, which gives histatin 5 the potential of charge regulation. It has been hypothesized that this might be a key feature in the mechanisms responsible for the antimicrobial properties of histatin 5 [63].

Because histatin 5 is a naturally occurring antimicrobial peptide, there is an interest in using it therapeutically in treatments against fungal infections [62]. However, since the saliva contains enzymes that can proteolytically degrade histatin 5, attempts have been made to find alternative histatin 5 variants that are more resistant to degradation. In these attempts, it was discovered that the candidacidal function is maintained in the shorter C-terminal fragments of histatin 5, *C14* and *C16* [60], as well as in the *P-113* fragment consisting of only 12 amino acids (residues 4-15) [64]. The amino acid sequences of all the mentioned histatin variants are presented in Table 2.1.

Table 2.1: The amino acid sequence of different histatin variants. Blue residues are positively charged, red residues are negatively charged, and histidine residues are represented in green.

Variant	Amino acid sequence
	1 5 10 15 20 25 30
Histatin 3	DSHAKRHHGYKRKFHEKHHSHRGYRSNYLYDN
Histatin 5	DSHAKRHHGYKRKFHEKHHSHRGY
C16	GYKRKFHEKHHSHRGY
C14	KRKFHEKHHSHRGY
P-113	AKRHHGYKRKFH

Although it is known *that* histatin 5 has antimicrobial properties, there is a lack of consensus regarding *how* it performs its functions. One common way for antimicrobial peptides to kill microbial cells is disrupting the cell membrane by for example forming membrane-spanning pores. However, because of its high charge and relatively low amphipathicity, histatin 5 is unable to form such classical pores [65]. Whether or not histatin 5 can disrupt the cell membrane at all has also been debated. Although studies have shown the lack of lytic activity on membranes by histatin 5 [66, 67], more recent studies have determined that histatin 5 can cause membrane disruption at higher physiological concentrations [68], or when bonded to zinc ions [69].

It has also been established that histatin 5 targets the mitochondrion of *C. albicans* [70], but the peptide has to enter the cells to do so. Several studies have shown that histatin 5 alone has a limited ability to translocate membranes [71–73]. It has however been pointed out that this might only be the case when there is a lack of membrane potential [68], which would also explain observed lack of interactions between histatin 5 and model bilayers [72]. Kumar et al. (2011) discovered that histatin 5 can utilize the polyamine transporters Dur3 and Dur31 because of its similarity to polyamine substrates [74]. This was further proved in a study where *C. albicans* Dur3 and Dur31 was expressed in the more resistant yeast species *Candida glabrata*, which increased the histatin 5 uptake and toxicity [75]. Because of this discovery, Tati et al. (2014) conjugated the active P-113 fragment to the polyamine spermidine, which resulted an increased candidacidal effect, as well as an increased resistance against proteolytic degradation compared to the native histatin 5 [76].

2.7 Magnesium ion transport in bacteria

There are three classes of magnesium ion (Mg^{2+}) transporters that have been identified in prokaryotes. These are (i) the *CorA* magnesium transporter, (ii) *magnesium transporter E* (MgtE), and (iii) the P-type ATPase magnesium transporters, *magnesium transporter A and B* (MgtA and MgtB) [77, 78]. While the first two classes are classified as ion channels [79, 80], the P-type ATPase magnesium transporters are classified as Mg^{2+} pumps, which are close relatives of the proton pumps [81]. However, in contrast to most P-type ATPase magnesium transporters that uses ATP to mediate the *efflux* of cations against an electrochemical gradient, MgtA and MgtB instead mediate the *influx* of Mg^{2+} ions down the electrochemical gradient [77, 78].

The N-terminal regions of both MgtA and MgtB from *Salmonella typhimurium* are intrinsically disordered, and Subramani et al. (2016) predicted that this also was the case for the N-terminal region of MgtA from *Escherichia coli (E. coli)* [82]. While the study mainly highlighted that the anionic phospholipid *cardiolipin* (CL) is essential for activating MgtA *in vitro*, it also divulged the N-terminal part most likely is not required for activation by CL, nor is it involved in membrane trafficking or targeting to the CL domains. This does however give rise to the question: what is the function of the N-terminal IDR in MgtA?

The N-terminal IDR in MgtA from *E. coli* constitutes the 33 first amino acids of the MgtA sequence and is referred to as *KEIF* in this work (named after the first readable part in its amino acid sequence), see Table 2.2. At physiological pH, KEIF has a net charge of $+3e$ and the majority of the charged residues are evenly distributed in the first half of the sequence.

Table 2.2: The amino acid sequence of the N-terminal IDR of MgtA, KEIF. Blue residues are positively charged, red residues are negatively charged, and histidine residues are represented in green.

```
1   5    10   15   20   25   30
MFKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV
```

Knowing your enemy is essential in war [83], and this also applies in the battle against bacterial infections. In order to develop new antibiotics, it is necessary to not only understand the antibacterial agents, but also the bacteria it is supposed to work against. It is our hope that by gaining a holistic understanding of the biological function of KEIF in MgtA, this will aid for future discovery of new effective antibiotics.

2.8 Simulations of IDPs

Because of the relatively recent realization that IDPs have significant biological functions (see Section 2.2), interest of simulating these types of proteins has naturally followed. By performing atomistic *molecular dynamics* (MD) simulations of proteins (see Section 5.2 for more information on MD simulations), it is theoretically possible to obtain both structural and dynamical properties. Because most experimental methods used for IDP properties only capture ensemble averages, atomistic simulations become an important tool for studying proteins in more detail. For example, with the help of well-tuned simulations it is possible to obtain structural and dynamical information that is unfeasible to obtain from experiments alone. Additionally, by using simulations it is possible to predict experimental results.

Atomistic MD simulations of IDPs have historically not been straight forward. The collection of equations and potentials used for atomistic simulations are called *force fields*. There is a large diversity of force fields available, and they are usually specialized for the specific system that they are used for. Thus, initial simulations of IDPs that used the then currently available force fields for "regular" globular proteins revealed major flaws. Not only did these force fields display a strong bias for sampling α-helical or β-sheet structures in IDPs [84], but they were also predisposed to sample too compact IDP conformations [85,86]. Consequently, several new options for atomistic IDP simulations have emerged during this past decade, mostly as completely new force fields [87–98], but an updated water model [99] has also been developed to work for IDPs in combination with older force fields. To optimize IDP force fields, both of the previously mentioned flaws need to be dealt with. A common solution to the problem of too strong secondary structure propensities is to adjust the protein backbone dihedral parameters, by for example the addition of a two-dimensional grid-based energy correction map (CMAP) [93]. To avoid sampling too compact IDP conformations it is necessary to fine-tune the interaction between the protein and the water in the simulations [90]. Use of the Kirkwood–Buff theory of solutions has also been proposed for solving this problem [100]. More information regarding force field development for IDPs can be found in the following reviews and references within: [101–103].

Development of force fields for IDP simulations is constantly ongoing. When a new force field is tested it is unfeasible to try it for all different kinds of proteins, and often they are mainly tested for a few particular proteins of interest. Additionally, the force field may only have been validated for specific system parameters. Therefore, it is particularly important to not take any force field for granted when initiating simulations with a new IDP. To obtain accurate simulations, it is always necessary to compare the simulation results to experimental data. It is also important to investigate if the force field in use actually is able to capture the desired observables accurately.

3. Theoretical background

3.1 Statistical thermodynamics

> *"Som namnet föreslår, statistisk termodynamik är en blandning av stat, istik, termos och dynamik."*
>
> – Jonas Ravelid

One fundamental theory behind the methods that are used for simulating proteins is *statistical thermodynamics*, which enables calculations of macroscopic properties from interpretation of the microscopic, or molecular, properties. While non-mechanical thermodynamical properties (such as temperature, entropy, free energy, chemical potential, etc.) are treated implicitly by classical thermodynamics, the mechanical properties (pressure, energy, volume, etc.) need to be calculated as *ensemble averages*. An ensemble is a collection of systems that are identical on the macroscopic level but can be very different on the molecular level. The systems in the ensemble are representations of all the possible states available for the thermodynamic system. Thus, properties can be calculated for each individual system, but these may vary significantly. Instead, it is more useful to calculate an ensemble average, which is the mean value calculated from each system at any instant of time.

The *first postulate of statistical mechanics* states that an ensemble average can be used in place of the time average of a real system, provided that the ensemble contains a large number of systems [104]. In the *second postulate of statistical mechanics*, which is also known as the principle of *a priori* probabilities, it is stated that the probability of finding a system in a specific state is the same for all possible states in the ensemble. The combination of the two postulates is called the *ergodic hypothesis*, which implies that an isolated system in an ensemble spends equal amounts of time in each available state [104, 105].

Thermodynamic systems with macroscopic constraints define the statistical ensembles. For example, an isolated system of fixed volume, V, containing a constant number of particles, N, while having a constant total energy, E, is called a *micro-canonical ensemble*, or an *NVE* ensemble. Other constraints give different ensembles such as the *grand canonical ensemble (μVT)*, where the constant variables instead are the chemical potential, μ, and the temperature, T, in addition to the fixed volume. In the work presented in this thesis, the *canonical ensemble (NVT)* and the *isobaric-isothermal ensemble (NpT)* were used. The number of particles, the volume, and the temperature is kept constant in the canonical ensemble, while the pressure, p, is constant instead of the volume in the isobaric-isothermal ensemble.

3.2 The radial distribution function

To evaluate the internal structure of particle systems, it is common to use the *radial distribution function*, $g(r)$. The radial distribution function is a measure of how the density of particles, ρ, in the system deviates from the bulk density as a function of the distance, r, from a reference particle [104]. Thus, $g(r) \to 0$ for $r \to 0$, and $g(r) \to 1$ for $r \to \infty$. Schematic illustrations of the radial distribution function are depicted in Figure 3.1. Thermodynamic properties such as energy, pressure, and chemical potential can also be obtained from the radial distribution function [105], although the function itself cannot be determined from thermodynamic measurements [104]. It is however possible to calculate the radial distribution function from scattering experiments, where the relation between the distance distribution function and the structure factor, $S(q)$, is given by [105–107]:

$$S(q) = 1 + \int_0^\infty 4\pi r^2 \rho (g(r) - 1) \frac{\sin(qr)}{qr} dr, \tag{3.1}$$

where q is the magnitude of the scattering vector. More theory about scattering and the structure factor is presented in section 7.1.

3.3 Intermolecular interactions

There are two categories of forces responsible for interactions on the molecular level: (i) the *intra*molecular forces that acts *within* molecules and that can connect atoms to form molecules, and (ii) the *inter*molecular forces, which are responsible for interactions *between* atomic and molecular species without creating chemical bonds between them. The intermolecular forces are central when trying to explain the structural and conformational properties of proteins. There are several types of intermolecular forces,

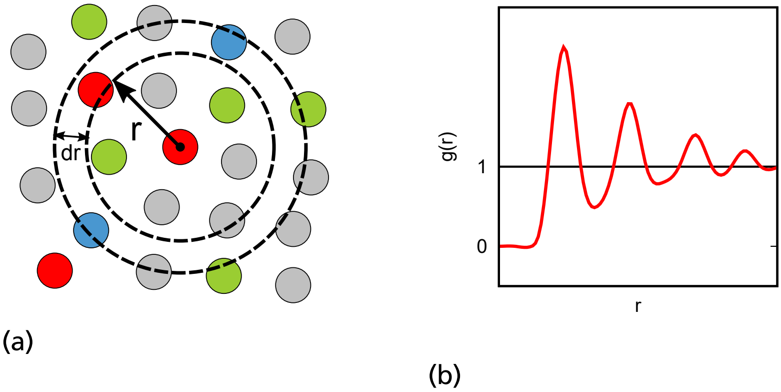


Figure 3.1: (a) A 2D representation of how the radial distribution function is obtained, and (b) a schematic plot of the radial distribution function of a liquid.

and they can be classified and grouped in numerous ways. The interactions that are relevant for this work are summarized in the following subsections, and all equations in these subsections are based on those given by Israelachvili (2011) [108].

3.3.1 Coulomb interactions

Charged particles in any medium give rise to electric fields, that act on the other charged particles in the system by the *Coulomb force*. For two charges, q_1 and q_2, at distance r, the magnitude of the force, $F(r)$, is given by:

$$F(r) = \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon_r r^2}. \tag{3.2}$$

The vacuum permittivity is given by ε_0, and ε_r is the dielectric constant. From this equation, an expression for the free energy of the *Coulomb interaction* between the two charges can be obtained:

$$U(r) = \int_0^\infty -F(r)dr = \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon_r r} = \frac{z_1 z_2 e^2}{4\pi\varepsilon_0\varepsilon_r r}. \tag{3.3}$$

In solution, q_i is exchanged for the ionic valency, z_i, multiplied by the elementary charge, e. Charges can also interact with *electric dipoles*. Electric dipole moments are present in many molecules, implying that they have an unequal distribution of electrons, that is, a separation of charges. The energy of a charge–dipole interaction is given by:

$$U(r) = -\frac{zeu\cos\theta}{4\pi\varepsilon_0\varepsilon_r r^2}, \tag{3.4}$$

where $u = ql$ is the dipole moment of a polar molecule with distance l between its charges. The angle, θ, is the angle of the dipole moment. Molecules that are neither charged nor dipoles, can instead be polarized by the electric field from other molecules or particles, that is, they become *induced dipoles*. The energy of the charge–induced dipole interaction also has a separate equation associated to it:

$$U(r) = -\frac{(ze)^2}{2(4\pi\varepsilon_0\varepsilon_r)^2 r^4}\left(\alpha_0 + \frac{u^2}{3kT}\right), \tag{3.5}$$

where α_0 is the polarizability of a nonpolar molecule, k is the Boltzmann constant and T is the temperature. All atoms and molecules are polarizable to different degrees, which is important for some of the interactions presented in the following subsection.

3.3.2 Van der Waals interactions

The *van der Waals interactions* are a collection of three different types of interaction that all possess a $1/r^6$ distance dependence. These three interactions arise from (i) the dipole–induced dipole interaction known as the induction force or the *Debye interaction*, (ii) the orientation force, which is a dipole–dipole interaction known as the *Keesom interaction*, and finally (iii) the *London forces*, also known as *dispersion forces*. The total van der Waals interaction energy between two molecules is given by:

$$\begin{aligned} U(r) &= -[U_{Debye} + U_{Keesom} + U_{London}] \\ &= -\left[(u_1^2\alpha_{02} + u_2^2\alpha_{01}) + \frac{u_1^2u_2^2}{3kT} + \frac{3\alpha_{01}\alpha_{02}h\nu_1\nu_2}{2(\nu_1+\nu_2)}\right] \Big/ (4\pi\varepsilon_0)^2 r^6, \end{aligned} \tag{3.6}$$

where h is the Planck constant, and ν_i is the orbiting frequency of each electron. As mentioned in section 3.3.1, all atoms and molecules are polarizable. Although the time average of the dipole moment of a neutral molecule is zero, it is possible for it to possess an instantaneous dipole moment due to the movement of the electrons. This instantaneous dipole moment can give rise to induced dipoles in surrounding molecules, which causes interactions in the system. This is the basis of the dispersion forces, and although these forces usually are attractive, they can give rise to repulsion in many-body systems. The dispersion interactions are present between all atoms and molecules.

3.3.3 Hydrogen bonds

When a hydrogen atom is covalently bonded to an electronegative atom, it becomes depleted of electrons. Because the electron-depleted hydrogen is so small, it allows for other electronegative atoms, such as oxygen, nitrogen, or fluoride, to get close enough to form a relatively strong bond referred to as a *hydrogen bond*. The hydrogen bond is primarily of electrostatic character and can be compared to a strong dipole–dipole interaction. Additionally, the bond is directional and can orient surrounding molecules. Although listed here under “intermolecular interactions”, the hydrogen bond can also act as a non-bonding intramolecular interaction. Examples of hydrogen bonds are depicted in Figure 3.2.

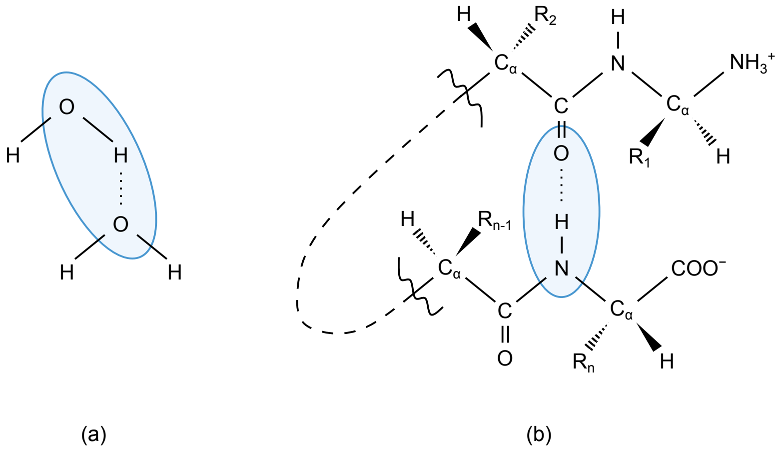


Figure 3.2: Illustrations of hydrogen bonds (dotted lines between hydrogen and oxygen atoms) as (a) an intermolecular interaction between two water molecules, and (b) as a non-bonding intramolecular interaction within a polypeptide.

3.3.4 The hydrophobic effect

The structure of water can be described as a network of hydrogen bonds. When introducing a hydrophobic molecule to an aqueous solution, it will disrupt the hydrogen bond network. To minimize this disruption, the water molecules have to rearrange themselves around the hydrophobic molecule. For small molecules, this can be done without breaking any of the hydrogen bonds. Nevertheless, the hydrogen bonds may need to deviate from their favored alignment, which is associated with a small energy increase. A rearrangement of the water molecules is entropically unfavorable, but it is generally more favorable than isolating water molecules completely from the hydrogen bond network. This tendency of water to maintain as much of the hydrogen bond network as possible is called the *hydrophobic effect*.

4. Simulation models

4.1 The atomistic model

As the name suggests, the atomistic model includes *all* atoms of the modeled molecules, including the solvent. An atomistic representation of histatin 5 is depicted in Figure 4.1. While the results from atomistic simulations can be of high detail, the computational cost of the simulations increases quickly as the size of the simulated systems increases. The model provided by the GROMACS package (version 4.6.7) [109–112] has been used as the atomistic simulation model for this work.

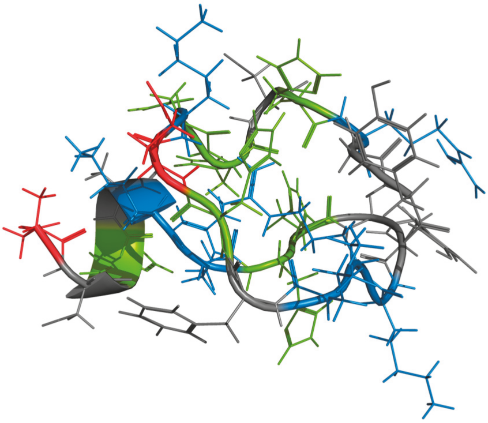


Figure 4.1: An atomistic representation of histatin 5. The backbone is also shown as a cartoon representation. Red color represents negative residues, blue represents positive residues, and green represents His residues. Surrounding atomistic water molecules are not included for visual simplicity.

The total potential energy in the atomistic model is constituted by intermolecular (nonbonded) and intramolecular (bonded) interaction potentials. The bonded and the non-bonded potentials include four and two types of interactions respectively:

$$\begin{aligned} U_{tot} &= U_{bonded} + U_{non\text{-}bonded} \\ &= U_{bond} + U_{angle} + U_d + U_{id} + U_{LJ} + U_{el}. \end{aligned} \tag{4.1}$$

All of the interaction potentials are summed over all groups of atoms involved in the respective interaction. The first bonded potential is the bond stretching, that is, a two-body harmonic stretching potential between covalently bonded atoms:

$$U_{bond} = \sum \frac{1}{2} k_b (r_{ij} - r_0)^2, \tag{4.2}$$

where r_{ij} is the distance between two bonded atoms with an equilibrium bond length r_0, and k_b is the force constant. A bond-angle vibration is included as a three-body harmonic potential between the atoms:

$$U_{angle} = \sum \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2, \tag{4.3}$$

where k_θ is a force constant, θ_{ijk} is the bond angle, and θ_0 is the equilibrium bond angle. The last two bonded potentials are torsion potentials with dihedral angles. A dihedral angle is an angle between two intersecting planes. In this model, the *proper dihedral angle* is defined according to the IUPAC/IUB convention [113]. For a proper dihedral, the angle, ϕ_{ijkl}, spans between the *ijk* and *jkl* planes formed by four linearly bonded atoms. The proper dihedral potential displays periodic behavior with the periodicity n and phase ϕ_s:

$$U_d = \sum k_\phi [1 + \cos(n\phi_{ijkl} - \phi_s)], \tag{4.4}$$

where k_ϕ is a force constant. To preserve the planar structure of planar groups, and to maintain chirality, the *improper dihedral angle* is introduced. Similar to the proper dihedral, the improper dihedral interaction is based on an angle, ξ_{ijkl}, between two planes. The difference is that the four interacting atoms do not need to be bonded linearly, and that this potential is harmonic in nature according to:

$$U_{id} = \sum \frac{1}{2} k_\xi (\xi_{ijkl} - \xi_0)^2, \tag{4.5}$$

with k_ξ as the force constant. All bonded interactions are illustrated in Figure 4.2.

The non-bonded interaction potentials are described by (i) a Lennard-Jones interaction, and (ii) a Coulomb interaction. Both potentials are pairwise additive and centro-symmetric. The Lennard-Jones interaction contains both a repulsion and a dispersion

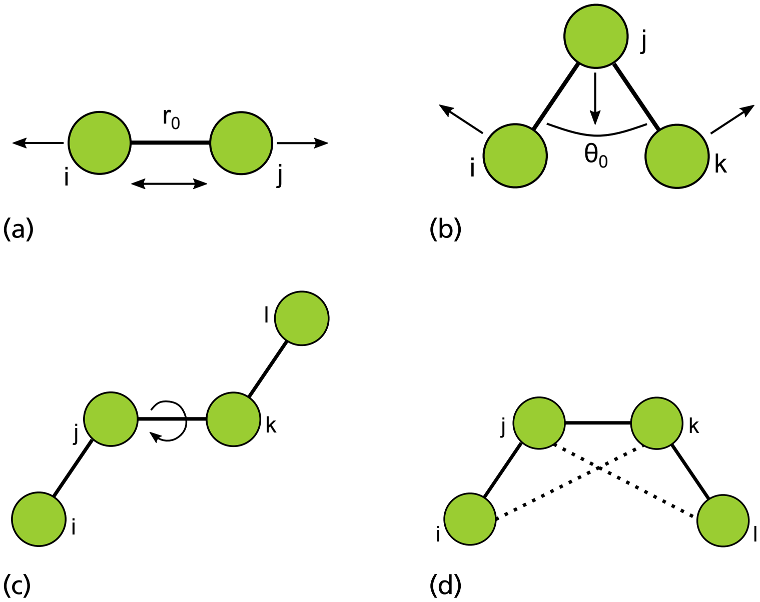


Figure 4.2: Examples of the bonded interactions in the MD simulations: (a) harmonic bond stretching, (b) bond-angle vibration, (c) proper dihedral, and (d) improper dihedral.

term:

$$U_{LJ} = \sum_{i<j} 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right), \tag{4.6}$$

where ϵ_{ij} is the potential well depth, σ_{ij} is the distance at which the potential becomes zero. The Coulomb potential represents the electrostatic interactions that act between two particles with the charges q_i and q_j:

$$U_{el} = \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}. \tag{4.7}$$

4.2 The coarse-grained model

The computational cost can easily become high when performing atomistic simulations of proteins, mainly because of the large amount of solvent that is needed. Two ways of reducing this cost are (i) to *coarse-grain* the protein and (ii) to treat the solvent implicitly through the dielectric constant. Coarse-graining is a way to simplify the simulation model by representing a collection of atoms by larger structures. For the model used in this work [114], each amino acid is represented by a single hard sphere (bead), see Figure 4.3. The N- and C-termini are also explicitly represented as a single sphere each. Each sphere can be neutral or assigned a point charge, which is set based on the amino acid sequence at physiological pH. The spheres are connected to each other, according to the amino acid sequence, by harmonic bonds in a so-called bead-necklace configuration. Any counter-ions in the system are treated explicitly as charged spheres, whereas the salt concentration is treated implicitly by adjusting the Debye screening length.

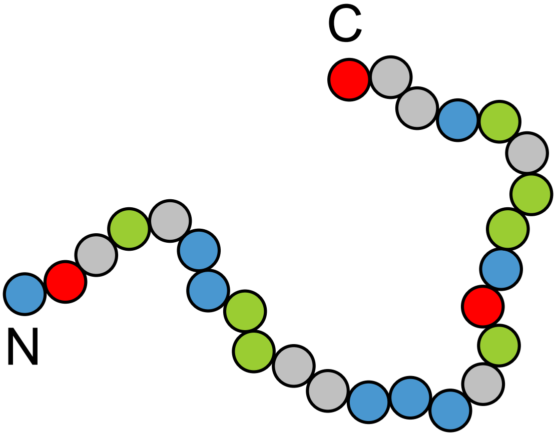


Figure 4.3: A coarse-grained representation of histatin 5—compare to the atomistic representation in Figure 4.1. Red beads have a charge of -1e, blue beads have a charge of $+1$e, and green beads represent His residues.

The total potential energy of the coarse-grained model is also separated into non-bonded and bonded interaction potentials [114, 115]. These potentials can be described as a sum of several different contributions:

$$\begin{aligned} U_{tot} &= U_{bonded} + U_{non\text{-}bonded} \\ &= U_{bond} + U_{hs} + U_{el} + U_{short}. \end{aligned} \tag{4.8}$$

For this model, the bonded potential was also comprised by a harmonic bond potential similar to that of the atomistic model:

$$U_{bond} = \sum_{i=1}^{N-1} \frac{1}{2} k_{bond} (r_{i,i+1} - r_0)^2, \tag{4.9}$$

where N is the total number of beads, $r_{i,i+1}$ is the distance between two bonded beads with an equilibrium bond length r_0, and k_{bond} is the force constant. An angular component can also be included in the bonded potential to impose restrictions on the chain flexibility but was not applied for this work.

The non-bonded potential comprises the three latter terms of the expression in Equation 4.8. A hard sphere potential, U_{hs}, is used to include the excluded volume of the beads:

$$U_{hs} = \sum_{i<j} u_{ij}^{hs}(r_{ij}), \tag{4.10}$$

where the hard sphere potential between the beads, $u_{ij}^{hs}(r_{ij})$ is given by:

$$u_{ij}^{hs}(r_{ij}) = \begin{cases} 0, & r_{ij} \geq R_i + R_j, \\ \infty, & r_{ij} < R_i + R_j, \end{cases} \tag{4.11}$$

where R is the radius of a bead. An extended Debye-Hückel potential is used to model the electrostatic potential, U_{el}:

$$U_{el} = \sum_{i<j} u_{ij}^{el}(r_{ij}) = \sum_{i<j} \frac{Z_i Z_j e^2}{4\varepsilon_0 \varepsilon_r} \frac{\exp[-\kappa(r_{ij} - (R_i + R_j))]}{(1 + \kappa R_i)(1 + \kappa R_j)} \frac{1}{r_{ij}}, \tag{4.12}$$

where Z is the bead charge, and κ is the inverse Debye screening length. The third and final nonbonded potential is a short-ranged attraction that acts between all beads through the following sum:

$$U_{short} = -\sum_{i<j} \frac{\varepsilon}{r_{ij}^6}, \tag{4.13}$$

where ε is a force constant.

5. Simulation methods

5.1 The simulation box

When performing computer simulations of solutions in a box, there will undoubtedly be particles interacting with the surface of the box. However, since most simulations of solutions are trying to capture realistic bulk behavior, such surface effects are suboptimal and need to be minimized. One way to solve this problem could be by simply simulating an extremely large box with a huge number of particles in it. Unfortunately, this is not feasible since such a simulation would use up an unreasonable amount of time and computer resources. Thus, another approach is needed to deal with this problem. The most common alternative is to introduce *periodic boundary conditions* [105]. In this method, the simulation box is replicated in a surrounding lattice in such a way so that when the particles in the original box move, the replicated particles will move in exactly the same directions. Thus, if a particle were to move out from the original box on one side, one of its replicated images will enter on the opposite side. By using periodic boundary conditions, the system has been converted from a simulation box of finite size, to a system of infinite size. An illustrative example of periodic boundary conditions is presented in Figure 5.1. Although using periodic boundary conditions might seem great for simulating solutions in bulk conditions, it is not entirely problem free. A bulk solution is non-periodic, and by using the periodic boundary conditions, a periodicity is introduced to the system, which can give rise to erroneous anisotropic correlations. The smaller the system is, the risk of this error becomes larger. This error is however expected to be smaller than the error that would arise from not using any periodic boundary conditions. To minimize the anisotropies in a periodic system, it is important to test and compare different box sizes.

The simplest simulation box has a cubic shape, but just because it is simple does not mean that it necessarily is the best alternative. One inconvenience with the cubic shape is that more solvent molecules are needed to fill box because of the corners of the cube. To minimize the solvent volume, it would be ideal to have a sphere. Unfortunately,

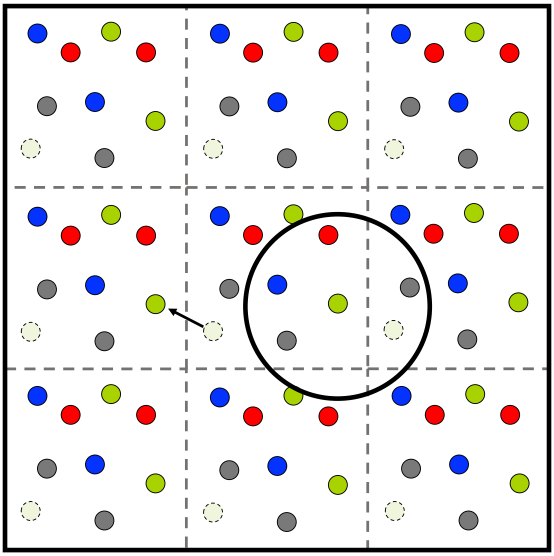


Figure 5.1: An illustration of periodic boundary conditions in two dimensions. The large circle represents a possible cutoff around the green particle that conforms to the minimum convention.

spheres cannot be organized in the lattice structure used for periodic boundary conditions. It is however possible to use other structures that are closer to a sphere in shape, but still available for lattice formation. Examples of such shapes are the truncated octahedron and the rhombic dodecahedron.

Periodic boundary conditions are normally used in combination with the *minimum image convention*, which states that each particle may only interact with the closest image of the remaining particles in the periodic lattice [105,116]. This is usually applied to the short-ranged nonbonded interactions by the use of a cubic or a spherical cutoff. The cutoff distance should be large enough to include all the important interactions but must not exceed half of the box length for a cubic box (this value depends on the box shape) to conform to the minimum image convention. This method is not always applicable for long-ranged interactions, which instead requires other corrections to the interaction potentials.

5.2 Molecular dynamics simulations

Molecular dynamics (MD) is a method that utilizes classical mechanics to describe particle displacement in computer simulations of molecular ensembles. All particles in the system are given initial coordinates and velocities. The initial conditions, together with the interaction potentials, enable calculations of forces in the system, and the particles will move accordingly. A basic MD algorithm [117] is shown in Example 5.1 below.

Example 5.1: A basic molecular dynamics algorithm

1. Generate the starting configuration with positions, velocities, and potential interactions of all atoms in the system.
2. Compute forces.
3. Update configuration.
4. Write output.
5. Repeat from step 2.

According to Allen and Tildesley (1989) [105] there are a few requirements that need to be fulfilled for an optimal simulation algorithm. These include that the algorithm should (i) be fast, (ii) allow for the use of a long time step, (iii) accurately reproduce the classical trajectory, (iv) follow the laws of conservation of energy and momentum, (v) be time-reversible, and finally, (vi) be simple. Some of these requirements are more important than others, depending on what type of simulation method that is used. For the MD method, it is more important that the time step is long than the algorithm being quick. The importance of the time step and how to increase it will be discussed in section 5.2.2. Since the method samples ensemble averages, exact classical trajectories are not necessary, and it is of greater importance to ensure the conservation of energy.

5.2.1 Equations of motions

The movement of the particles in an MD simulation is derived from the forces acting on each particle. The forces, $\mathbf{F}_i$, are given by the gradient of the potential energy at each particle position, $\mathbf{r}_i$:

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i}. \tag{5.1}$$

According to *Newton's equations of motion* in classical mechanics, the force is given by:

$$\mathbf{F}_i = m_i \cdot \mathbf{a} = m_i \cdot \frac{\partial^2 \mathbf{r}_i}{\partial t^2}, \tag{5.2}$$

where m_i is the particle mass, and t is the time. The new positions are obtained by integrating Newton's equations of motion. The integration methods are usually based on finite differences and solved at each simulation time step [118]. There are several integration methods available for MD simulations, but the integrator used in this work is called the *leap-frog algorithm* [118]. The name originates from that the positions and the velocities, $\mathbf{v}_i$, of the system are calculated alternately in time according to the following equations:

$$\mathbf{v}_i\left(t+\frac{1}{2}\Delta t\right) = \mathbf{v}_i\left(t-\frac{1}{2}\Delta t\right) + \frac{\Delta t}{m_i}\mathbf{F}_i(t), \tag{5.3}$$

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i\left(t+\frac{1}{2}\Delta t\right). \tag{5.4}$$

An illustration of the leap-frog algorithm is depicted in Figure 5.2. If the equations of motions are solved correctly in the simulations, the energy conservation law will apply. In addition, these equations are reversible in time.

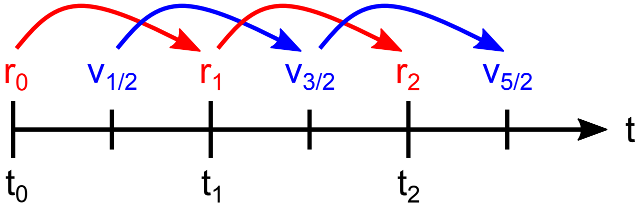


Figure 5.2: A schematic illustration of the leap-frog algorithm. The calculations of the coordinates (red) and the velocities (blue) are leaping past each other in time.

5.2.2 Constraints

How far the particles in a simulation move depends on the time step used in the simulations. The time step is usually of the same magnitude as molecular vibrations, that is, on the femtosecond scale. When studying proteins, some processes are comparatively slow, and it is therefore necessary to perform simulations on the microsecond scale. Performing long simulations with short time steps result in a large amount of simulation steps, which can become problematic considering available computer resources. Thus, it may be desirable to increase the time step. One way of achieving this is by imposing constraints on the relative coordinates in the simulated system. This is usually done by first solving the equations of motion without any constraints, and then performing a correction to the particle coordinates. One of the most common methods for applying bond constraints is the *LINCS algorithm* [119], which projects the new bond onto the old bond, followed by a correction for the change in bond length that arises due to rotation of the bond, see illustration in Figure 5.3.

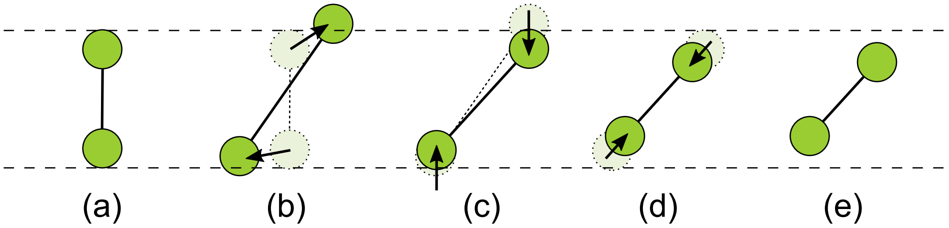


Figure 5.3: A schematic example of how bond constraints are applied by using the LINCS algorithm. State (a) shows the initial bond configuration. In (b) the particles are moved, followed by a correction by projection onto the original bond (c). The bond length is then adjusted accordingly (d) before reaching the final bond configuration (e).

5.2.3 Long-range interaction corrections

It is possible for the cutoff distance used for the non-bonded interactions in the MD simulations to interfere with the long-ranged interactions of the system. For those particular cases, it is necessary to apply long-range corrections. The correction of the Lennard-Jones potential is done assuming that the cutoff is long enough for ignoring the repulsion term, and that the system is homogeneous beyond the cutoff distance. The long-ranged electrostatic interactions are corrected by using the *particle-mesh Ewald* (PME) method [120], which is an improved version of the original Ewald summation method. In the Ewald summation, the potential energy is converted into two quickly converging terms: (i) a direct space sum, and (ii) a reciprocal space sum. Originally, the reciprocal sum was associated by a high computational cost, making the Ewald summation unfeasible for large systems. Instead, the PME method utilizes interpolation and fast Fourier transform to speed up the calculation of the reciprocal sum significantly for larger systems.

5.2.4 Temperature and pressure

The MD simulations of this work have been performed in an isobaric-isothermal ensemble, which requires both the temperature and the pressure to be constant throughout the simulations. This is done by applying temperature and pressure couplings. There are several different options for couplings available for MD simulations, but for this work, the *velocity-rescaling temperature coupling* [121] and the *Parinello-Rahman pressure coupling* [122] have been used. The velocity-rescaling coupling is based on another coupling method called the Berendsen temperature coupling [123]. When using the Berendsen algorithm, any deviations from the reference temperature, T_0, is corrected by an exponential decay,

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \tag{5.5}$$

where τ is a time constant. The problem with this algorithm is that it suppresses kinetic energy fluctuations, which can lead to incorrect ensemble averages. By instead using the velocity-rescaling temperature coupling, the problem is solved by the addition of a kinetic energy correction term. When using the Parinello-Rahman barostat, an additional term containing the simulation box vectors is included in the equations of motion to represent changes in the pressure.

5.3 Monte Carlo simulations

Contrary to the MD method is the *Metropolis Monte Carlo* (MC) method [105], which is based on random displacements instead of physical moves. Due to the unphysical nature of the particle displacement in the MC algorithm, it is necessary to first perform a trial move to check if the move is viable. This is done by calculating the energy of the new configuration and compare it to that of the old configuration. The trial move is always accepted if the new energy is lower than the old energy. However, to ensure that the MC simulation satisfies the ergodic hypothesis, it is also necessary to accept some moves that will lead to a configuration of higher energy. The probability of such a move is determined by a Boltzmann distribution. A basic MC algorithm is shown in Example 5.2 below.

Example 5.2: A basic Monte Carlo algorithm

1. Generate a random starting configuration.
2. Select a particle at random and calculate its energy, $U(o)$.
3. Displace the particle by performing a trial move.
4. Calculate the new energy, $U(n)$.
    (a) If $U(n) \leq U(o)$, accept the move.
    (b) If $U(n) > U(o)$, generate a random number, $X \in [0, 1]$.
    (c) If $X < e^{-[U(n)-U(o)]/kT}$, accept the move.
    (d) Otherwise reject the move and restore the previous configuration.
5. Repeat from step 2.

5.3.1 Trial moves

A large variety of moves can be applied in the MC algorithm, and they can be either single-particle moves, or multi-particle moves. All moves have a probability associated with them. In this section, only the moves used in this work will be summarized.

Single particle translation is the only single-particle move that has been used in the MC simulations of this work. For this move, a single particle is displaced a short distance, see Figure 5.4a. Usually, this distance is random in both magnitude and direction, although the magnitude cannot exceed a set maximum value.

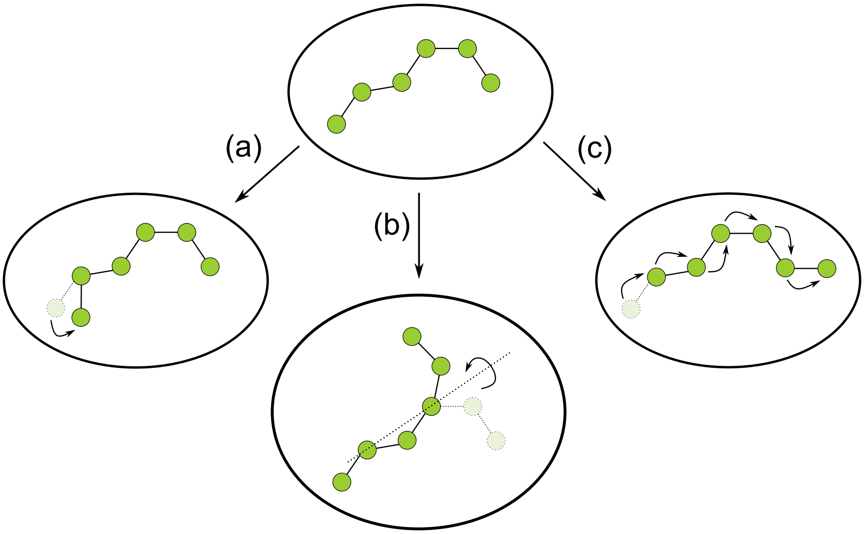


Figure 5.4: A schematic illustration of a few trial moves used in the MC simulations: (a) single particle translation, (b) Pivot rotation, and (c) slithering chain move.

The most basic group move used in the simulations is the *chain translation*. When using this move, the entire chain is displaced in the simulation box. Although this move does not alter the chain conformation and is thus not particularly important in simulations of single chains with implicit salt and solvent, it has significant impact on simulations with several proteins or other particles, and while using explicit salt and solvent.

Another group move that has been used is the *Pivot rotation*. When performing a Pivot rotation, one end of the simulated chain is rotated around an axis created by two randomly selected particles in the chain, see Figure 5.4b.

The fourth and last trial move is the *slithering chain move*, also known as a reptation move. Both names are related to reptiles, as the motion of a moving snake has been used as inspiration for the move. The move is initiated by the displacement of one of the particles at the end of the chain. The remaining particles in the chain will follow the movement in the trail of the previous configuration, analogous to the slithering move of a snake, see Figure 5.4c.

6. Simulation analyses

6.1 Sampling convergence

When performing simulations, it is of uttermost importance to make sure that the results are reliable for making quantitative predictions. Grossfield and Zuckerman (2009) [124] claimed that there are two main reasons for erroneous molecular simulations: (i) model inaccuracy, and (ii) insufficient sampling. While the accuracy of the model should be relatively easy to evaluate by comparing with experimental references, evaluating the sampling of a simulation is not as straightforward. A very common method of assessing the sampling of a simulation is to investigate the convergence of a single observable, such as the radius of gyration. Using this method alone may however be unreliable, especially for dynamical systems that possess one or more smaller states [124]. An illustration of single observable convergence of a globular protein compared to a flexible protein is given in Example 6.1.

Example 6.1: Convergence of globular vs. flexible proteins

A simulation of a protein is initiated in the conformation of an extended straight chain. The natural state of the protein is a globular conformation and thus, as the simulation progresses, the radius of gyration decreases until it converges to a constant value, see example in Figure 6.1a. For a system like this, the initial non-converged part is removed before performing further analyses of the system. This does however not apply for flexible proteins because being fully extended is one of the natural conformations, and the radius of gyration will vary a lot throughout the entire simulation, see Figure 6.1b. Therefore, it is impossible to determine convergence of flexible proteins, such as IDPs, by using this method only.

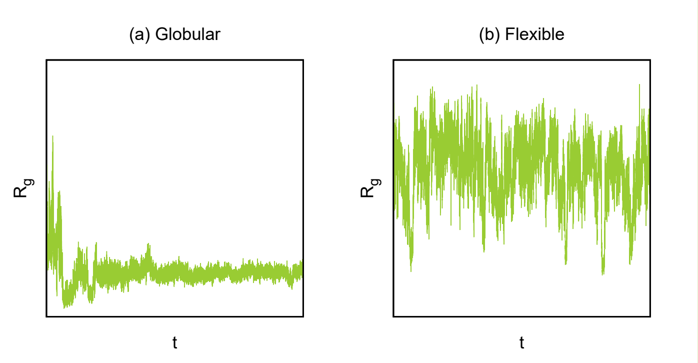


Figure 6.1: The radius of gyration as a function of simulation time for (a) a globular protein, and (b) a flexible protein. Both plots are shown on the same time scale.

Although the convergence of a single observable might be unreliable in dynamic simulations, it is possible to estimate the correlation and uncertainty of the observable. This can be done in two ways: (i) time correlation analysis, and (ii) block averaging. Time correlation is a statistical measure of correlation between different observables, A and B, at different time steps. Since the correlation is observed as a function of time, t, it is also known as a time correlation function, $C_{AB}(t)$:

$$C_{AB}(t) = \langle A(t' + t)B(t) \rangle. \tag{6.1}$$

If the correlated observables are identical ($A = B$), this measure is instead referred to as an *autocorrelation function* (ACF), $C(t)$. The integral of the ACF is called the correlation time, τ:

$$\tau = \int_0^\infty C(t)dt, \tag{6.2}$$

which describes the length of the simulation time required for an observable to no longer have any correlation with its previous values. The correlation time may be used to estimate the quality of the sampling of a single observable by $N = t_{tot}/\tau$, where t_{tot} is the total simulation time, and $N \gg 1$ indicates good sampling [124].

By using the block averaging method, it is possible to obtain error estimates of a single observable. This method is especially useful for estimating single observable convergence since it is based on information from the entire simulated trajectory [124]. To obtain the error estimate, the simulated trajectory is divided into n segments referred to as "blocks", and an average value of the observable, B_i, is calculated for each block. The *block-averaged standard error* (BSE), $\varepsilon(n)$, is then calculated for each block according to the following equation [117, 124]:

$$\varepsilon(n) = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (B_i - \langle B \rangle)^2}, \tag{6.3}$$

where $\langle B \rangle$ is the total average. These errors are then plotted as a function of the block size, and the resulting curve increases with the block size until it converges to a final BSE. A converged BSE is usually a good indicator of a converged observable. More information regarding this type of error estimation can be found in the following references and references within: [124–126]. Example 6.2 below includes a discussion and illustrations of how ACF and BSE can be applied when analyzing convergence of simulations of flexible proteins.

Example 6.2: ACF and BSE

Although the radius of gyration may vary a lot for flexible proteins (see Example 6.1), it can still be more or less converged. Figure 6.2 shows examples where the radius of gyration of three different simulations of flexible proteins have been analyzed through their respective ACF and BSE. Already by visual inspection of both plots it can be observed that the radius of gyration of Sim 1 (dark green) seems to be considerably less converged compared to Sim 2 (light green) and Sim 3 (blue). This is also evident from the correlation time, which is only 9 ps for Sim 3 and 1607 ps for Sim 2, but as long as 23824 ps for Sim 1. For the latter case, this corresponds to slightly more than 2 % of the total simulation time. The BSE of Sim 1 converges at 0.058 nm, Sim 2 at 0.013 nm, and Sim 3 at 0.009 nm.

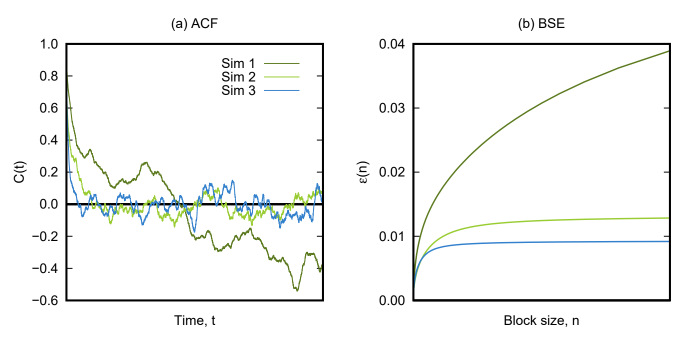


Figure 6.2: (a) Examples of ACFs with different correlation times, and (b) examples of BSEs of different levels of convergence. All curves are obtained from simulations with a total simulation time of 1 μs each.

6.2 Distances

Because of the flexible nature of IDPs, it is only reasonable to characterize them by their average properties. The *radius of gyration*, R_g, is an average measure of the compactness of a structure. In a macromolecule, the radius of gyration is defined as the root mean square distance between the mass elements and the center of mass [127]. The radius of gyration is calculated using the following equation:

$$R_g = \sqrt{\frac{\sum_{i=1}^{n} m_i \|\mathbf{r}_i\|^2}{\sum_{i=1}^{n} m_i}}, \tag{6.4}$$

where n is the total number of mass elements, m_i is the mass of each mass element, and $\mathbf{r}_i$ is the position of the mass elements relative to the center of mass of the macromolecule. For polymers it is common to treat each monomer as individual mass elements. This also extends to proteins, for which the amino acids are considered to be the individual mass elements.

Another property that is useful to study is the *end-to-end distance*, R_{ee}. This distance is calculated as the root mean square of the distance between the two ends of a linear chain:

$$R_{ee} = \sqrt{\|\mathbf{r}_1 - \mathbf{r}_n\|^2}. \tag{6.5}$$

The ends of a chain are defined differently depending on the model used. In the atomistic model, the distance is calculated between the α-carbons in the terminal amino acid residues of the protein. In the coarse-grained model, the distance is instead calculated between the center of mass of the terminal beads.

To ensure that the protein is not interacting with itself during a simulation, it is also important to keep track of the minimum distance between the protein and its periodic images. In the MD simulations, the minimum distance should never be larger than the cutoff distance of the nonbonded interactions. For the MC simulations, the box size is always at least double the theoretical maximum extension of the modeled chain.

6.3 Principal component analysis

A more extensive study of conformational variations can be done using *principal component analysis* (PCA) based on the method proposed by Campos and Baptista (2009) [128]. PCA is a mathematical method used to reduce the dimensionality of a data set without losing information about variations [129]. Applied to our protein system, the aim is to map the conformational space onto a low-dimensional representation, and to study the conformations in a multidimensional free energy landscape. All of the conformational information could technically be found in a complete energy landscape, but this is generally undesired due to two reasons. Firstly, a complete energy landscape provides much more information than is actually needed. Secondly, the use of a complete energy landscape includes a vast number of dimensions, scaling as $(3N - 6)$ with the number of atoms, N, in the system, which is unfeasible for most systems. The PCA analysis transforms a set of correlated variables into a set of uncorrelated variables, or coordinates, called principal components (PCs). Most of the variation is captured by the first PC, and most of the remaining variation is accounted for in the second PC, and so on. To obtain an analysis of the conformational space of high accuracy, it is necessary to use several PCs. However, since the strongest correlations should be included in the first PCs, it should be sufficient to only use the first two to get an adequate estimation of the most significant conformational information of the system. Furthermore, when performing a PCA on a protein system, a few dimensions can be excluded by ignoring all atoms that are not part of the peptide backbone. When performing this type of PCA on proteins, structures from each time step in a simulation are aligned and compared to a reference structure. Deciding the reference structure is difficult for IDPs because of their intrinsic flexibility. Thus, it should be noted that PCA might not be the most optimal analysis for IDPs. In our case, the reference structure has been chosen from a *root mean square deviation* (RMSD) analysis that selects the most "average" structure in the simulated trajectory, that is, the structure that differs the least from all other structures in the simulation. Because of this, PCA of IDPs is still useful for comparing replicates of the same protein simulations (and can thus help in assessing sampling convergence) and can be used to identify different conformational classes.

6.4 DSSP analysis

In 1983 W. Kabsch and C. Sander presented the *Define Secondary Structure of Proteins* (DSSP) program for objective determination of the secondary structure in proteins [130]. For the program to work properly it needs a 3D structure input, from which the most likely secondary structure is computed. Instead of using backbone dihedrals, the algorithm utilizes structure recognition of hydrogen bonding patterns. The program defines the hydrogen bond by a generous energy cutoff that is estimated by assuming the hydrogen bond to be a purely electrostatic interaction. Although the definition of the hydrogen bond can be disputed, Kabsch and Sander claimed that their approximation is a suitable compromise for the intended purpose of the program. The secondary structures that can be identified by the algorithm are the 3_{10}-helix, the α-helix, the π-helix, the isolated β-bridge, the extended strand (β-sheet), as well as hydrogen bonded turns and bends. Unfortunately, the classic DSSP program is unable to identify the PPII helix. Thus, Mansiaux et al. (2011) developed a new method called *DSSP-PPII* [131], which is able to perform PPII assignment by utilizing the DSSP program.

6.5 The form factor in coarse-grained MC simulations

To obtain a scattering profile from the coarse-grained MC simulations used in this work, the individual particles (beads) are considered to be *point scatterers*. A point scatterer scatter equally at all angles and thus has a constant form factor. With a constant form factor, the total scattering intensity is directly proportional to the structure factor. In a system containing N identical beads, the structure factor is given by:

$$S(q) = \left\langle \frac{1}{N} \left| \sum_{j=1}^{N} \exp(i\mathbf{q} \cdot \mathbf{r}_j) \right|^2 \right\rangle. \tag{6.6}$$

However, in a system using different bead types, i and j, partial structure factors can be obtained by using the following equation:

$$S(q)_{ij} = \left\langle \frac{1}{(N_i N_j)^{1/2}} \left[\sum_{i=1}^{N_i} \exp(i\mathbf{q} \cdot \mathbf{r}_i) \right] \left[\sum_{j=1}^{N_j} \exp(-i\mathbf{q} \cdot \mathbf{r}_j) \right] \right\rangle, \tag{6.7}$$

where N_i and N_j are the total number of beads i and j, respectively. The total structure factor can then be calculated from the partial structure factors by using the relation given by:

$$S(q) = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \frac{(N_i N_j)^{1/2}}{N} S(q)_{ij}. \tag{6.8}$$

In a system with a single coarse-grained peptide chain, all scattering arises from intraparticle interference, which means that the calculated scattering profile is, in fact, the form factor. Because the beads are considered to be point scatterers, system differences are only observed at short distances, that is, at high q-values.

The method presented in this section has previously been described in a paper by Cragnell et al. (2018) [114]. Because the method does not differentiate between beads that are within a chain, and beads that belong to different chains, it is impossible to obtain the peptide form factor in a system containing more than one peptide chain. Further scattering theory is presented in Section 7.1.

7. Experimental background

Since the intrinsic flexibility of IDPs causes them to assume a wide range of conformations, their structure cannot be captured experimentally by, for example, crystallographic techniques. Fortunately, there are other options available to experimentally study the structure of IDPs. A few of the most common techniques are SAXS, NMR, and single molecule FRET, which all provide ensemble averages and/or distance distributions of IDP structures in solution. In the following subsections, some of the most important experimental techniques used for this thesis work will be presented.

7.1 Small-angle X-ray scattering

Small-angle X-ray scattering is an experimental technique used to obtain the average structure, shape, and flexibility of particles in the size range of 1-100 nm. It can be applied to samples in both solid and liquid states and is thus also suitable for IDPs in solution. The results obtained from SAXS are representative, but only give the ensemble average of the sample. For this reason, it is usually advantageous to combine SAXS with other complementary techniques.

7.1.1 The scattering vector

The fundamental principle of SAXS is, as the name suggests, the scattering of X-rays. When a sample is irradiated by X-rays, some of the radiation will be scattered by all the atoms in the sample. If the scattering is elastic, that is, it has the same energy as the incident radiation, the scattered waves from the atoms will interfere with each other. When the interference reaches a detector, it is interpreted as a 2D interference pattern, from which the internal structure of the sample can be deduced.

The momentum transfer, also known as the *scattering vector*, $\mathbf{q}$, is defined as the difference between the incident wave vector, $\mathbf{k}_0$, and the scattered wave vector, $\mathbf{k}_\mathrm{s}$, see Figure

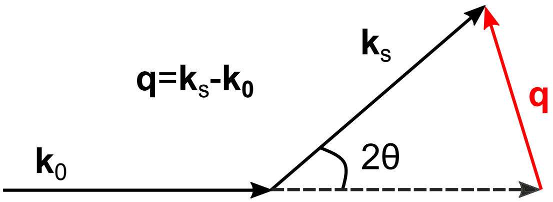


Figure 7.1: A schematic representation of the wave vectors and the resulting scattering vector (red).

7.1. The angle between the two wave vectors is defined as 2θ. Since the scattering is elastic, the magnitude of the two wave vectors is the same, that is, $|\mathbf{k}_0| = |\mathbf{k}_\mathrm{s}| = 2\pi/\lambda$, where λ is the wavelength of the incident wave. Thus, the magnitude of the scattering vector, q, can be obtained accordingly:

$$q = |\mathbf{q}| = \frac{4\pi \sin(\theta)}{\lambda}. \tag{7.1}$$

7.1.2 The form factor

The *scattering intensity* is normally expressed as a function of the magnitude of the scattering vector: $I(q)$. The scattering intensity is dependent on the particle volume and the electron density, which means that larger particles give rise to larger scattering intensities. When considering scattering from particles, the scattering intensity can also be expressed by the following relation:

$$I(q) = P(q)S(q), \tag{7.2}$$

where $P(q)$ is known as the *form factor*, and $S(q)$ is the so-called *structure factor*. The form factor represents the *intra*-particle distances, which are derived from the interference pattern that arises due to scattering from individual particles. Similarly, the structure factor represents the *inter*-particle distances. Properties such as average shape and size distributions can be obtained from the form factor. To obtain the pure form factor, it is necessary to have very dilute samples. However, dilute samples are extremely sensitive to contaminations, since the scattering from small particles can be completely dominated by the scattering from larger particles. Thus, it is important to keep the samples pure, and to perform several measurements for averaging when trying to obtain the form factor. Examples of different IDP form factors are shown in Figure 7.2a.

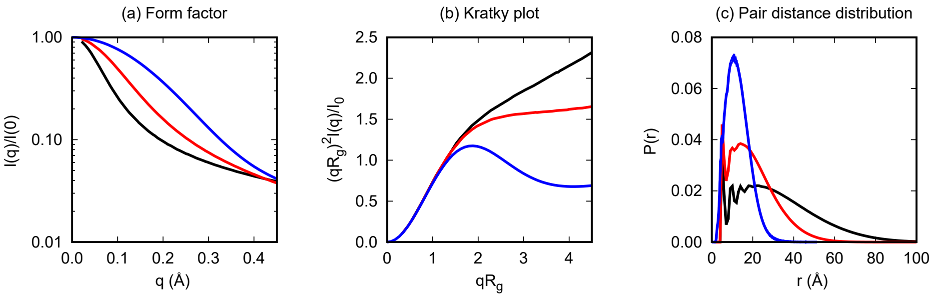


Figure 7.2: Examples of different plots obtainable from SAXS. (a) The form factor, (b) the Kratky plot, and (c) the pair distance distribution function. The blue curve represents a more globular shape, the red curve represents a flexible random chain, and the black curve represents a more rigid and extended random chain.

7.1.3 The Kratky plot

The shape and flexibility of a protein in solution can be determined by transforming the SAXS spectra into a *Kratky plot*. A Kratky plot is done by plotting $q^2 I(q)$ vs. q [22]. To make the plot dimensionless, $I(q)$ is normalized by the *forward scattering*, $I(0)$, and q is normalized by the radius of gyration. The shape of the resulting curve gives a good indication of the shape and flexibility of the protein. A well-defined globular protein will show as a bell-like shape in the Kratky plot, while a flexible, random chain will increase and reach a plateau. Further increase of the plateau at high q-values indicates a more rigid and extended structure. Examples of these different shapes are illustrated in the Kratky plot in Figure 7.2b.

7.1.4 The pair distance distribution function

By studying the form factor, it is possible to obtain the shape and the intra-particle distances expressed as a *pair distance distribution function*, $P(r)$. This is done by applying a Fourier transformation on the entire scattering profile [132]:

$$P(q) = 4\pi \int_0^\infty P(r) \frac{\sin(qr)}{qr} dr. \tag{7.3}$$

The shape of the pair distance distribution function also gives the shape of the particles in the studied solution. It is possible to obtain the radius of gyration and the forward scattering from the pair distance distribution function as well [22, 133]:

$$R_g^2 = \frac{\int P(r) r^2 dr}{2 \int P(r) dr}, \tag{7.4}$$

$$I(0) = 4\pi \int_0^{D_{max}} P(r) dr. \tag{7.5}$$

The pair distance distribution function is assumed to be zero at $r = 0$ and at D_{max}, which is the maximum extension of the protein chain. Examples of the pair distance distribution for different types of peptides are depicted in Figure 7.2c. Other shapes, as for example disks, hollow spheres, and dumbbells, would result in a very different appearance of the pair distance distribution function, which is nicely illustrated in Figure 5 in the paper by Svergun and Koch (2003) [106].

7.1.5 The Guinier approximation

Another way of obtaining the radius of gyration from the form factor is by using the *Guinier approximation* [22,106,133], which states that at sufficiently low values of q, the scattering profile can be approximated to:

$$I(q) = I(0) \exp\left(-\frac{q^2 R_g^2}{3} \right), \tag{7.6}$$

or rather:

$$\ln\big(I(q)\big) = \ln\big(I(0)\big) - \frac{R_g^2}{3} q^2. \tag{7.7}$$

By making a straight-line fit in the so-called *Guinier region* the forward scattering and the radius of gyration can be obtained as the y-axis intercept and the slope, respectively, see example in Figure 7.3. The approximation is restricted to $qR_g < 1.3$ for well-folded, globular proteins, but the Guinier region is even smaller for extended molecules, including many IDPs. For the analysis of IDPs, a value of $qR_g < 0.8$ has been recommended [22], although this value may vary depending on the nature of the IDPs.

In contrast to the pair distance distribution method, which utilizes the entire scattering profile, using such a restricted part of the curve may result in only a few data points, which reduces the reliability of the resulting radius of gyration. For this reason, utilizing the pair distance distribution function is considered to be more precise for determining the radius of gyration [133]. The pair distance distribution method is, however, very sensitive and consequently more difficult to reproduce. Therefore, it is advantageous to compare and report the radius of gyration from both methods.

7.2 Circular dichroism spectroscopy

By utilizing *circular dichroism* (CD) spectroscopy, it is possible to detect the secondary structure of proteins in solution. Although IDPs often lack secondary structure, it is still an important tool for detecting changes in the conformational ensemble as an

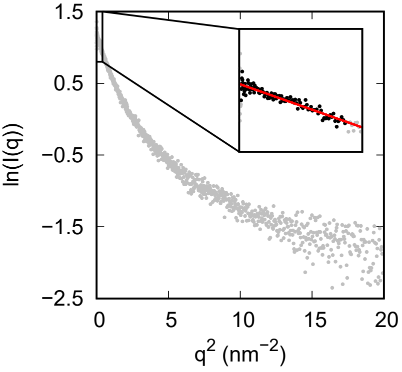


Figure 7.3: An example of a Guinier plot of an IDP. To obtain the forward scattering and the radius of gyration, linear regression (red line) is done on the black data points, which belongs to the Guinier region ($qR_g < 0.8$). The remaining scattering profile is shown in gray.

effect of changes in the solution environment. The basic principle behind the CD technique is the difference in absorption of left- and right-handed circularly polarized light. Differently polarized light is absorbed to different extents by optically active molecules in solution. The *difference in absorption*, ΔA, is often expressed in terms of *ellipticity*, θ, in degrees (deg) instead, and the two quantities are related through the following equation: $\theta = 32.98 \cdot \Delta A$ [134].

Studies of the secondary structure of proteins on a conventional CD instrument are done in a wavelength range of 180-260 nm. To obtain more reliable results it is necessary to measure in the far-UV region (170 nm and below), but this requires more elaborate techniques and synchrotron radiation. It is the absorbance of the peptide bond that makes it possible to determine the secondary structure of proteins. The characteristic spectra arise due to a strong $\pi \rightarrow \pi*$ transition around 190 nm, and a weaker $n \rightarrow \pi*$ transition around 220 nm [134]. Depending on the secondary structure of the studied protein, a characteristic CD spectrum will be observed. Different secondary structures give rise to different shapes in the spectra, see Figure 7.4. For example, an α-helix is characterized by negative bands around 210 and 220 nm, as well as a positive band between 190-195 nm. A negative band close to 220 nm and a positive band at 195 nm are characteristic for antiparallel β-sheets. IDPs that lack secondary structure are attributed a random coil structure that is recognized by a single negative band at low wavelengths in the CD spectrum. Less known is the extended PPII helix [13] that can be identified for some IDPs. This structure is defined by a weak positive band close to 220 nm and, a strong negative band around 190 nm [135].

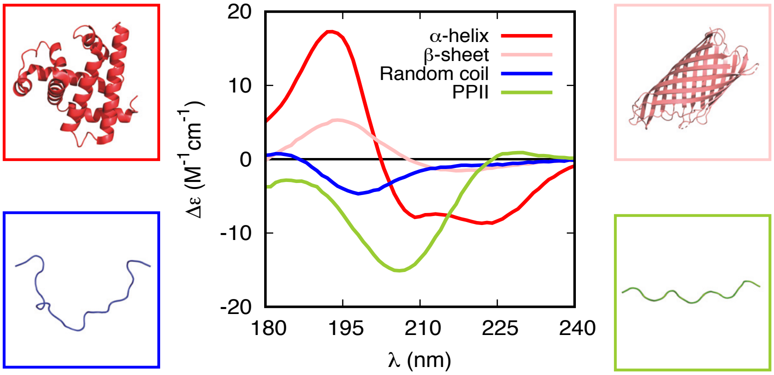


Figure 7.4: Examples of selected secondary structures that can be identified in a CD spectrum. All the plotted data was obtained from the Protein Circular Dichroism Data Bank (PCDDB) [136]. The protein structures in the corresponding boxes are examples; they are not necessarily the same proteins that are shown in the graph, but their secondary structure content is of the same type. The typical spectrum of the α-helix and its corresponding example are shown in red[1]. Similarly, the β-sheet examples are shown in pink[2], the random coil examples in blue[3], and the PPII helix examples in green[4].

To compare CD spectra, it is advantageous to normalize the spectra with respect to molar concentrations. This is done by converting the ellipticity to *molar ellipticity* or *mean residue ellipticity*. For proteins, the mean residue ellipticity is calculated from the mean residue weight, MRW, of the peptide bonds:

$$MRW = \frac{M}{N-1}, \tag{7.8}$$

where M is the molecular weight of the protein, and N is the number of amino acid residues in the protein chain. The mean residue ellipticity, $[\theta]$, at wavelength λ is then given by:

$$[\theta] = \frac{MRW \cdot \theta_\lambda}{10 \cdot d \cdot c}, \tag{7.9}$$

where θ_λ is the observed ellipticity at wavelength λ, d is the pathlength in cm, and c is the concentration in g/mL. The mean residue ellipticity is usually given in deg cm^2 dmol^{-1}. If the CD data is expressed as the difference in absorption, it is instead normalized by converting ΔA to the *molar differential extinction coefficient*, $\Delta\varepsilon$:

$$\Delta\varepsilon = \frac{\Delta A}{C \cdot d}. \tag{7.10}$$

---

[1] Spectrum CD0000048000 Lees et al. (2006) [137], and structure 1MNK; Krzywda and Wilkinson (1995) Interactions among residues CD3, E7, E10 and E11 in myoglobins: attempts to simulate the O2 and CO binding properties of *Aplysia* myoglobin; doi: 10.2210/pdb1mnk/pdb [138] from RCSB PDB [3].

[2] Spectrum CD0000116000, Abdul-Gader et al. (2011) [139], and structure 1UYO; Oomen et al. (2004) Translocator domain of autotransporter NalP from *Neisseria meningitidis*; doi: 10.2210/pdb1UYO/pdb [140] from RCSB PDB [3].

[3] Spectrum CD0006121000, Tolchard et al. (2018) [141] and in-house generated structure example.

[4] Spectrum CD0004553000, Lopes et al. (2014) [142] and in-house generated structure example.

The variable C is the molar concentration, which gives $\Delta\varepsilon$ the unit of M^{-1} cm^{-1}. Because of the previously mentioned relation between the ellipticity and the difference in absorption, conversion between the mean residue ellipticity and the molar differential extinction coefficient is done according to the following equation: $[\theta] = 3298 \cdot \Delta\varepsilon$.

The resolution obtained from CD measurements is quite low as it only provides general information about structural averages. The technique is also sensitive to the choice of solution components and contaminants. Several buffers and ions have shown to possess high absorption in the far-UV range. To minimize the absorbance from other species than the protein of interest, it is recommended to for example use low buffer concentrations, and to exchange chloride ions with sulphate or fluoride ions [134]. It is also possible to use a shorter cell pathlength to reduce the effect of disturbing species.

8. Research

This chapter summarizes the research and results that have been procured during my PhD studies. Most of the studies in this thesis concern histatin 5, which has been thoroughly studied by using several different theoretical and experimental techniques. Here, I first present histatin 5 while discussing its conformational properties as a single chain in bulk conditions. The first section deals with the effect of charge regulation in histatin 5, followed by a section investigating its temperature dependence. This section also concerns investigation and evaluation of different simulation methods. Further research concerning different MD force fields is presented in the subsequent section, where the simulated secondary structure of histatin 5 (and a few other peptides) is investigated. Toward the end of this chapter, the focus is shifted back to the conformational ensemble, but for a more potent variant of histatin 5, that is, the shorter P-113 peptide conjugated to the polyamine spermidine. Histatin 5 and its variants are of interest because of their antimicrobial properties, but we have also studied the bacterial KEIF peptide, which is summarized in a section of its own. Apart from characterizing the conformational ensemble of KEIF, the study also covered investigation of the interactions between the peptide and model membranes. The chapter is concluded with research implications and a discussion about future research connected to this work.

8.1 The effect of charge regulation in histatin 5

Because histatin 5 is comprised by such a large portion of His residues, it is important to consider charge regulation. It is known that histatin 5 interacts with negatively charged surfaces [143–145] and membranes (see Section 2.6), and it would be naive to assume that the His residues of histatin 5 are completely neutral in the vicinity of such structures under physiological conditions. Thus, histatin 5 has been investigated by simulations of its two extreme cases: (i) with all His residues neutral, and (ii) with all His residues protonated, that is, with a charge of $+1e$. In the latter case, histatin 5 has a net charge of $+12e$ instead of $+5e$. By considering the charge distribution of histatin 5, the conformational properties can be predicted. With FCR = 0.38 and NCPR = 0.21, histatin 5 with neutral His residues has been found to belong to the conformational class R3 and is thus predicted to be a polyampholytic IDP (see Figure 2.4, Section 2.2 for reference). When all the His residues are protonated, the peptide instead belongs to the R4 class (FCR = 0.67 and NCPR = 0.50) and thus behaves as a polyelectrolyte.

By performing MD simulations of histatin 5 with all His residues neutral (Hst5_HIE) and with all His residues protonated (Hst5_HIP), the conformational properties were investigated and compared. The radius of gyration was found to be larger for Hst5_HIP, and it was observed to sample more extended conformations than Hst5_HIE, see Figure 8.1a. Both distance distributions had similar shapes, with a FWHM of 0.54 nm and 0.51 nm for Hst5_HIE and Hst5_HIP respectively. Representative structures of the two histatin 5 variants are displayed in Figure 8.1b.

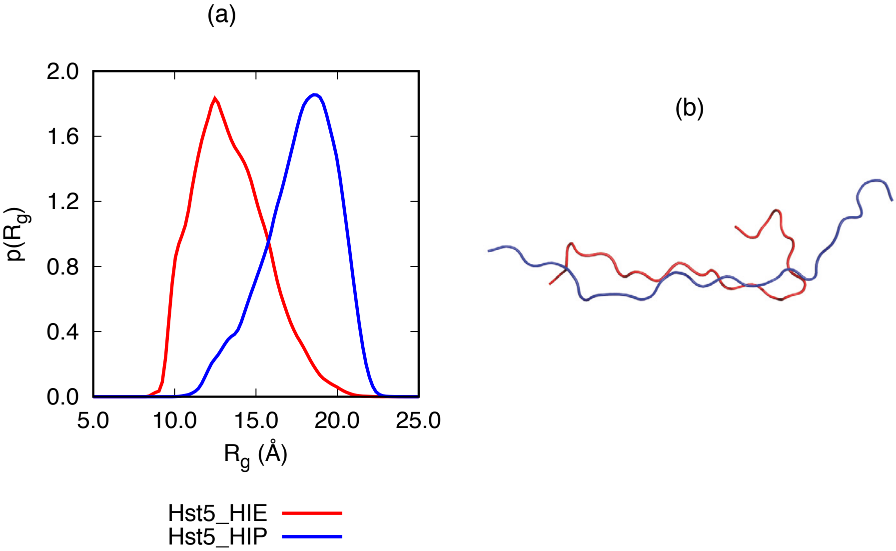


Figure 8.1: (a) Distance distribution of the radius of gyration, and (b) cartoon representations of the largest cluster conformer of the two Hst5 variants. Hst5_HIE is shown in red and Hst5_HIP in blue.

The secondary structure was estimated based on the backbone dihedral angles of each amino acid residue, and their relative populations in different regions of a Ramachandran map. Both variants of the peptide had large population of the PPII-helical and β-sheet regions of the Ramachandran map, and both of these regions were more populated for Hst5_HIP, see Figure 8.2. While Hst5_HIE displayed some α-helical structure elements, this was almost absent in Hst5_HIP. Thus, it seems like protonation of the HIS residues caused the PPII and β-sheet content of histatin 5 to increase at the expense of α-helical and random coil structures.

These results are extremely important for other studies of histatin 5 at for example lower

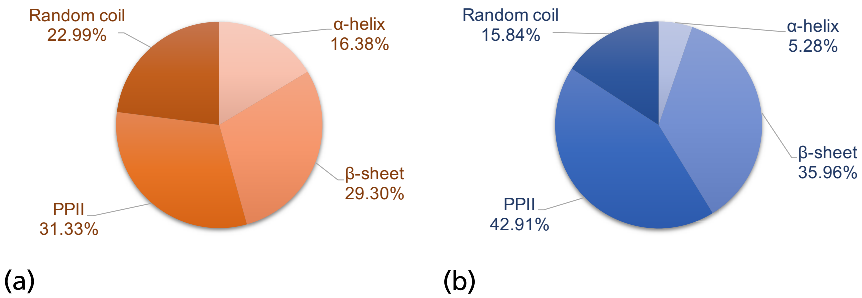


Figure 8.2: The secondary structure content of (a) Hst5_HIE and (b) Hst5_HIP from MD simulations.

pH values or in systems containing other charged chemical species, such as surfaces or membranes. Thus, further investigation was performed, based on the paper by Hyltegren et al. (2020) [146], in which surface adsorption of histatin 5 was studied using a combination of MC simulations and a library of structures obtained from MD simulations. In their paper, only histatin 5 with neutral HIS residues were investigated. Therefore, we decided to perform similar simulations but with both Hst5_HIE and Hst5_HIP, where the previously presented MD simulations were used to build the libraries of MD structures.

The MC simulations of Hst5_HIE and Hst5_HIP in the vicinity of negatively charged surfaces yielded similar results to what was obtained in the MD simulations, both for the radius of gyration and its distance distribution, as well as for the average secondary structure content per amino acid. It was also discovered that these properties were nearly unaffected by changes in the salt concentration. When it came to surface adsorption, clear differences were found between the two histatin 5 variants, and these results were affected by the salt concentration. Three regions were defined for determining adsorption: (A) bulk region ($A > 30\,\text{Å}$ from the surface), (B) intermediate region ($6\,\text{Å} < B \leq 30\,\text{Å}$ from the surface), and (C) adsorbed region ($C \leq 6\,\text{Å}$ from the surface). A peptide was considered to belong to a region if its center of mass was found within the limits of that region. At low salt concentration, the majority of the simulated frames were located in region B for both Hst5_HIE and Hst5_HIP, whereas at high salt concentration there were most often located in the bulk (region A). Although higher salt concentrations seemed to screen the electrostatic attraction between the peptides and the surface, a larger fraction of Hst5_HIP was always found close to the surface compared to Hst5_HIE. There also seemed to be a small increase in the average radius of gyration of Hst5_HIP when going across the three regions from A to C. However, further investigation is needed to determine if this trend is of statistical significance.

8.2 Temperature dependence in simulations of histatin 5

During 2016 trial measurements and short simulations of histatin 5 at different temperatures had been performed and were indicating contradictory results for different methods. Thus, longer simulations using different techniques, as well as several different experimental techniques were performed to further investigate these discrepancies. SAXS, as well as atomistic MD simulations and coarse-grained MC simulations were used to study the change in the radius of gyration with temperature. However, because atomistic MD simulations are computationally expensive and SAXS beam time is precious, only a few carefully chosen temperatures were selected for these two techniques: 10 °C, 20 °C, 37 °C, and 50 °C. The selection was limited by the temperature range of the SAXS equipment (4-60 °C), and the same temperatures were selected for both techniques for easier comparison.

Atomistic MD simulations of histatin 5 were performed with three different combinations of force fields and water models: (i) Amber ff99SB-ILDN [89] and TIP3P [147] (MD(A3)), (ii) Amber ff99SB-ILDN and TIP4P-D [99] (MD(A4)), and (iii) CHARMM36m [93] with TIP3P (MD(C3)). Although the first combination, MD(A3), was known to give too compact structures for IPDs [86, 99], it was unknown if the temperature-induced conformational trends would be the same as for MD(A4). Thus, MD(A3) was still considered to be interesting for the study and was included, nevertheless. The combination of MD(A4) was previously known to work well for histatin 5 at room temperature and was therefore an obvious choice to use in this study. Simulations with the MD(C3) setup have previously been proven to work well for IDPs, which made it a good alternative for expanding the study to include a "new", different force field. The coarse-grained MC simulations of histatin 5 (MC(IW)) were performed using an implicit water model and were thus computationally cheap to perform compared to the MD simulations. This made it possible to perform the MC(IW) simulations at a wider selection of temperatures, although the selected temperatures were still chosen within the temperature range of the previous techniques (10-50 °C).

Two additional experimental techniques were also included in the study: CD spectroscopy and NMR spectroscopy. The CD spectroscopy was performed at the same four selected temperature that were used in the SAXS measurements and the MD simulations, and the aim of the measurements was to detect any temperature-induced changes to the secondary structure of histatin 5. The NMR spectroscopy of histatin 5 was performed in a temperature range of 5-50 °C. The aim of the NMR measurements was to obtain hydrodynamic radii of histatin 5 at the different temperatures, which were then converted to radii of gyration by a compaction factor of 1.1 that was obtained using the method presented by Marsh and Forman-Kay (2010) [148]. Re-evaluation of the paper after publication, and by using the method by Nygaard et al. (2017) [149]

as intended, revealed that a conversion factor of $\sim 0.91 \pm 0.05$ would have been a more reasonable choice. Although this would change the obtained values, it would not significantly affect the trend that is discussed in the paper, and thus not affect the conclusions either.

The resulting average radii of gyration, $\langle R_g \rangle$, at different temperatures are depicted in Figure 8.3. Interestingly enough, all the different methods used gave different temperature trends. No trend could be found from the SAXS measurements, and although some correlations were found from the MD simulations, they were not statistically significant enough to draw any conclusion. Only the NMR and the MC(IW) simulations displayed strong and significant temperature correlations, although in the opposite direction of each other. Further investigation of the SAXS data revealed that no obvious differences between the results at different temperatures could be discerned for any of the systems.

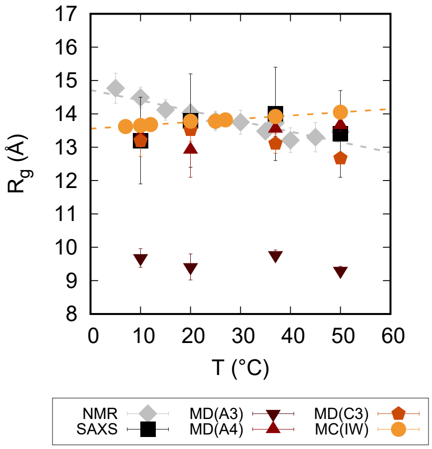


Figure 8.3: The radius of gyration, $\langle R_g \rangle$, of histatin 5 as a function of temperature of the six different methods. The dashed lines are linear regression trend lines. All data points include error bars, although the errors are too small to be visible for some data points.

A more obvious temperature-induced conformational change was observed in the CD spectra, which are depicted in Figure 8.4. Negative bands around 191 and 240 nm, as well as a positive band around 222 nm, were observed for all temperatures, although the absolute intensites of the bands were found to decrease with increased temperature. The shape of the observed spectra was a clear indication of PPII structure. By studying the change in ellipticity at 191 and 222 nm, a significant linear correlation with temperature was found, which indicated a destabilization of the PPII structure.

Secondary structure analysis was performed on representative structures obtained from conformation clustering of each simulated MD trajectory by utilizing a DSSP-PPII

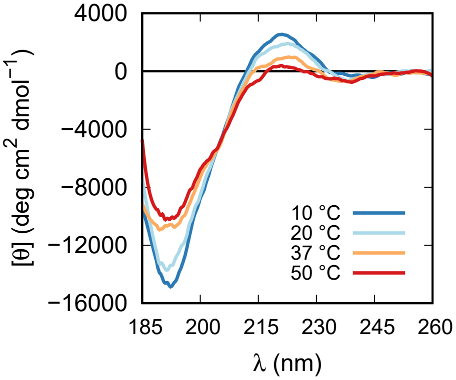


Figure 8.4: Normalized CD spectra of histatin 5 at the four different temperatures.

program similar to the program developed by Mansiaux et al. (2011) [131] and to the software used in the PolyPrOnline database [150]. To our surprise, the analysis revealed no PPII structure for any of the simulated structures. With the exception of the MD(A3) simulation, which possessed some degree of helical structures, the MD simulations were dominated by bends, turns and random coils instead. Initial speculation of why no PPII structure was identified mainly revolved around the problem of sampling; maybe enhanced sampling or longer simulations would be needed to capture the correct secondary structure of IDPs in MD simulations. However, by comparing MD(A4) after 5 µs to after 7 µs provided identical representative structures from the conformation cluster analysis, which strongly indicated that the sampling was sufficient for this simulation.

Because the temperature trends so far had been all over the place in the MD simulations, the temperature dependence of the two MD water models were investigated. MD simulations of pure water (TIP3P and TIP4P-D), were performed at different temperatures using the Amber ff99SB-ILDN force field. The dielectric constant was then computed compared to reference values [151–154]. The dielectric constant of both water models was found to decrease linearly with increasing temperature, although the actual magnitudes of the values did not match the reference values; TIP3P generated values larger than the reference, and TIP4P-D generated values smaller than the reference. However, since both water models displayed a correct temperature dependence in the temperature range of interest, it was concluded that the deviating behavior of histatin 5 at different temperatures in the MD simulations is highly unlikely to be caused by the water models.

The dielectric constant was also investigated in the coarse-grained MC model by performing additional simulations of histatin 5 but by using the dielectric constants obtained from the MD simulations in place of the standard reference values. Interestingly enough, this did barely affect the resulting radius of gyration at all. The results were initially surprising because the largest difference between the simulated dielectric constants at a single temperature can be compared to the difference between the dielectric constant of water and the dielectric constant of for example ethanol. Nevertheless, the results indicated that the temperature-induced conformational changes of histatin 5 in the MC simulations are completely independent of electrostatic interactions in the system. To further investigate this result, coarse-grained MC simulations were also performed of two other chains with equal length as histatin 5 but consisting completely of (i) neutral beads and (ii) beads with a charge of $+0.2e$ each (net change $+5e$). Although this led to different values of the radius of gyration, the temperature trend was revealed to be more or less the same for all three chains. From these results it was clear that the temperature dependence was not electrostatically driven in the coarse-grained MC model.

8.3 PPII structure in simulations of histatin 5

It was unexpected that the previous analysis of the simulated histatin 5 trajectories did not display any PPII structure. Thus, further investigation was performed using a different approach. By simply analyzing the backbone dihedral angles of the trajectories, Ramachandran plots were made, which clearly showed that the simulated histatin 5 in fact possessed substantial PPII content. Subsequently, this also revealed problems with the DSSP-PPII program, which was immediately updated to properly assign PPII structure. It was also discovered that the different force fields that had been used (Amber ff99SB-ILDN and CHARMM36m) had propensities for different secondary structure elements, although both provided PPII structure. This sparked an interest for further comparing different force fields and their ability to sample PPII structure in histatin 5 and in other small peptides that are known to possess PPII structure.

Four different force field that are known work for simulating IDPs were chosen: (A) Amber ff99SB-disp [96], (B) Amber ff99SB-ILDN [89,99], (C) CHARMM36IDPSFF [97, 98], and (D) CHARMM36m [93]. The force fields were used in combination with their recommended water models. Histatin 5 and four other peptides, as well as four variants of one of the peptides, were simulated for at least a total of 5 µs each. All the chosen force fields provided similar average values of the radius of gyration, although the values of CHARMM36IDPSFF were slightly smaller compared to the rest. By using PCA, force fields were shown to sample the same conformational spaces for each individual peptide. This was however done with different probabilities, and

CHARMM36IDPSFF seemed to be most adept at exploring all of the conformational space at the chosen temperature. The secondary structure estimation of this study was performed in the same way as it was in the study of Section 8.1. All force field sampled PPII structure but to different extents. Amber ff99SB-disp had the highest population in the PPII region for most of the peptides, whereas Amber ff99SB-ILDN and CHARMM36m were more prone to sample β-sheet structure, and CHARMM36IDP often sampled more "random coils". The force fields were more directly compared by investigating the conformation clustering in the simulations of histatin 5. For this analysis, all simulations with the different force fields were concatenated to a single trajectory on which the conformation clustering was performed. The clusters were then analyzed to obtain the relative population of the individual force fields, see Figure 8.5. These results revealed that the force fields populated the different clusters unevenly, which further indicates that there are slight biases toward different conformations for the different force fields.

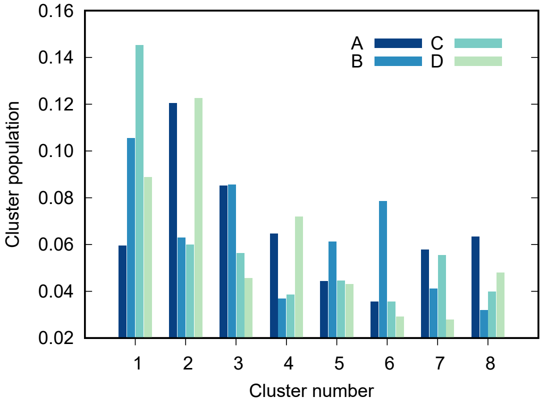


Figure 8.5: Weighted cluster population from the individual force fields in the top eight clusters of histatin 5. A = Amber ff99SB-disp, B = Amber ff99SB-ILDN, C = CHARMM36IDP, and D = CHARMM36m.

An effort was made to determine which of the force fields were more accurate. This was done by making CD predictions using SESCA [155]. Unfortunately, the algorithm was unable to provide any simulated spectrum that could be matched with any of the experimental spectra.

As a final part of this study, the correlation between the number of Pro residues and the PPII content was investigated for P_{13} and four variants containing only Ala and Pro, see Figure 8.6. A linear regression was performed, which revealed that all force fields provided linear correlations of statistical significance ($p < 0.05$). This correlation does however need to be investigated further to see how it is affected by for example chain length and the relative position of the Pro residues within the sequences.

(a)

Name	Sequence	#P
P_{13}	PPPPPPPPPPPPP	13
V1: P_6AP_6	PPPPPPAPPPPPP	12
V2: $(PA)_6P$	PAPAPAPAPAPAP	7
V3: $(P_3A_2)_2P_3$	PPPAAPPPAAPPP	9
V4: $(A_3P_2)_2A_3$	AAAPPAAAPPAAA	4

(b)

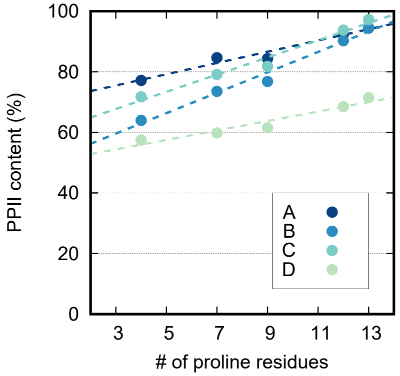


Figure 8.6: (a) The amino acid sequences and the total number of Pro residues (#P) in each sequence of P_{13} and its four variants. (b) PPII content as a function of the total number of proline residues for the four different force fields. A = Amber ff99SB-disp, B = Amber ff99SB-ILDN, C = CHARMM36IDP, and D = CHARMM36m.

8.4 Conformational properties of Histatin 5-spermidine conjugates

Because the candidacidal effect was shown to be increased when the shorter histatin 5 fragment, P-113, was conjugated to the polyamine spermidine, we decided to investigate its conformational ensemble to see how it contributes to this increased candidacidal effect. By considering the charge distribution of histatin 5 and its variants, the conformational properties were predicted, see Table 8.1. The predictions are based on the amino acid charges at physiological pH (pH 7.4), assuming positively charged N-terminal residues and negatively charged C-terminal residues, as well as completely neutral histidine residues.

Table 8.1: Net charge, FCR, NCPR, and the corresponding conformational class for the protein sequences in aqueous solution at physiological pH.

Name	q/e	FCR	NCPR	Class
Histatin 5	+5	0.38	0.21	R3
P-113	+5	0.58	0.42	R4
CSpd	+6	0.40	0.40	R4
NSpd	+4	0.40	0.27	R3

The shorter histatin 5 variant, P-113 (also known as Hst5$_{4-15}$ in Paper I), was predicted to belong to the R4 class, which is reasonable since it has the same net charge as histatin 5 but distributed over half the number of amino acid residues. When conjugated to spermidine, a glycine linker consisting of three glycine residues in a row has been added to the P-113 chain, which makes the peptide part of the conjugates increase from twelve to fifteen amino acid residues. Since one terminus each of NSpd and CSpd is connected to spermidine, they only contain one charged terminal residue each, see Figure 8.7, which was accounted for in the predictions. Conjugation of spermidine at the C-terminus of P-113 did not seem to impose a large difference for the prediction, as the protein part of CSpd also was predicted to belong to the R4 class. Conjugation at the N-terminus, however, seems to predict slightly more flexibility as NSpd instead was predicted to belong to the R3 class. In summary, based on the charge distribution only, histatin 5 and the peptide part of NSpd are expected to form polyampholytic coils or hairpins, whereas P-113 and the peptide part of CSpd are expected to form polyelectrolytic rods/coils (see section 2.2 for reference) in aqueous solution at physiological pH.

Figure 8.7: A schematic illustration of the charge distribution of (from top to bottom) histatin 5, P-113, CSpd, and NSpd. The N-termini and the C-termini are explicitly included as NT and CT respectively. Succinic acid and spermidine are denoted by SAC and SPD respectively. Gray residues are neutral, blue residues are positive ($+1e$), red residues are negative ($-1e$), and the green spermidine has a charge of $+2e$.

SAXS curves representing the histatin 5-spermidine conjugates were obtained from both experimental measurements and from MD simulations. Initial comparisons of the experimental and the computational data revealed a large discrepancy between the experimental and computational data, as the SAXS curves did not overlap properly. Further investigation proceeded, and after reevaluating the system and the methods it was hypothesized that the spermidine could not be detected properly in the SAXS measurements. Calculation of the scattering length density of spermidine was performed and revealed to be close to that of the bulk used in the experiments (20 mM Tris and 140 mM NaCl in aqueous solution), a case which is known for giving a very low contrast in the measurements. Hence, the experimental SAXS curve of the histatin 5-spermidine conjugate was compared to that of the P-113 fragment only. The comparison revealed nearly identical SAXS curves, thus confirming the hypothesis. To further support the hypothesis, SAXS curves representing only the peptide part of the simulated conjugates were also obtained and compared to the experimental SAXS curve of the histatin 5-spermidine conjugate (see Figure 8.8). Finally, the experimental and the simulated results were in reasonable agreement and the previously stated hypothesis was again confirmed.

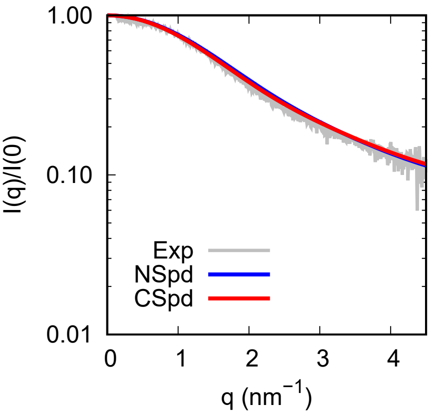


Figure 8.8: SAXS spectrum of the histatin 5-spermidine conjugate from experiment (gray) compared to the SAXS spectra representing only the peptide part of the simulated conjugates (red and blue).

The conclusion that the spermidine is basically "invisible" in the SAXS measurements provided an additional important piece of information: the conformational properties of P-113 are unaffected by conjugation to spermidine. This was also observed by investigating the distance distributions of the radius of gyration (and the end-to-end distance) of the simulated P-113 and the histatin 5-spermidine conjugates, where the peptide part of the histatin 5-spermidine conjugates were seen to sample the same distributions as P-113, see Figure 8.9. Similarly, PCA of the protein backbone of both P-113 and the histatin 5-spermidine conjugates proved that the same conformational space was sampled in all different systems.

Evaluation of the SAXS curves representing the histatin 5-spermidine conjugates showed that the protein part of the conjugates possessed extended flexible conformations, identified by the characteristic shape of the Kratky plots. For the conjugates in their entirety, a slightly more extended and rigid average conformation was indicated by the shape of the Kratky plot and the $P(r)$ of the simulated data, see Figure 8.10.

When comparing the results of the two types of conjugates from the simulations, CSpd was found to be more rigid and extended relative to NSpd, despite the latter having a longer contour length because of its succinic acid linker. In fact, the presence of the succinic linker could be an important factor causing the observed differences. Conjugation at the N-terminus decreases the charge density of the entire NSpd conjugate, both by increasing the contour length of the molecule, and also by effectively removing one positive charge by not having any N-terminus. In contrast, CSpd has a shorter contour length as well as no negative C-terminal charge, which leads to a higher charge density and increased intramolecular electrostatic repulsion. This could also be related

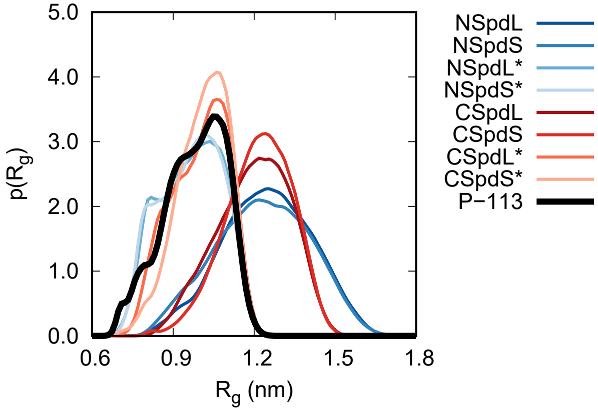


Figure 8.9: Distance distribution functions of the radius of gyration of the simulated conjugates and the P-113 fragment. The conjugates have been analyzed in their entirety (NSpdL, NSpdS, CSpdL and CSpdS) and by considering their peptide part only (NSpdL*, NSpdS*, CSpdL* and CSpdS*).

to the fact that CSpd has been shown to have slightly superior fungicidal activity compared to NSpd [76]. It was hypothesized that the more rigid and extended structure of CSpd exposes the spermidine further to the solvent, which makes it more accessible for recognition by the polyamine transporters.

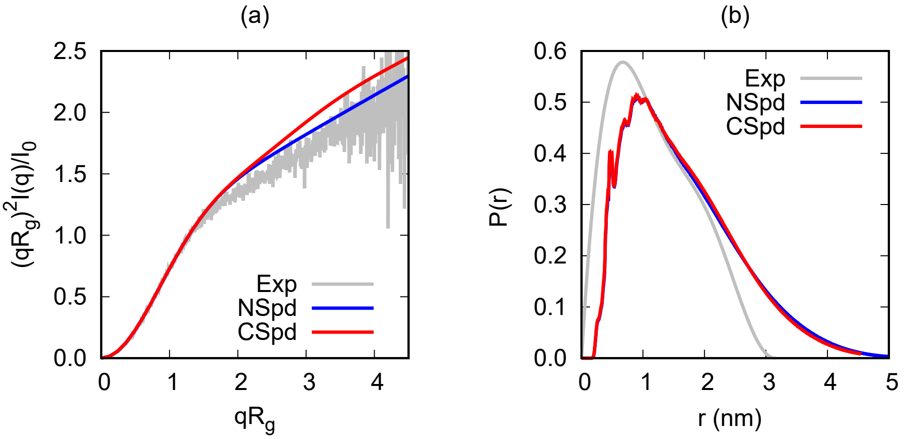


Figure 8.10: (a) Kratky plot and (b) P(r) of the histatin 5-spermidine conjugate from experiments (gray), as well as from the simulated conjugates (blue and red).

The secondary structure of histatin 5 and the histatin 5-spermidine conjugates were investigated by performing CD measurements. The resulting spectra (see Figure 8.11) showed expected α-helical structure for histatin 5, but no characteristic features associated with established secondary structure could be seen for the conjugates, which were thus considered to be completely disordered in this medium. The experiments were performed in trifluoroethanol (TFE) because, at the time of the study in Paper I, histatin 5 was considered to possess no secondary structure in aqueous solution. Later, we learned that histatin 5 actually possess some degree PPII structure in aqueous solution (see Section 8.2). Therefore, it does not seem unreasonable for the histatin 5-spermidine conjugates to also have the ability to assume PPII structure. In fact, since PPII structure is quite extended, the histatin 5-spermidine conjugates might even have a higher tendency to form PPII structure compared to histatin 5.

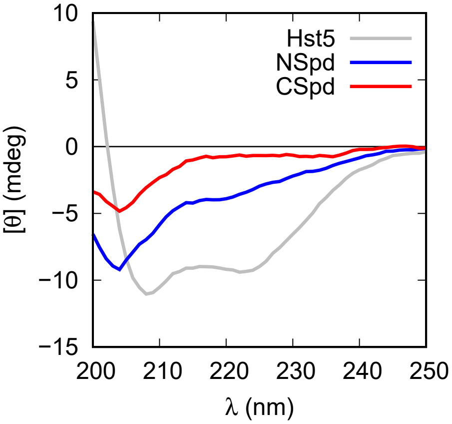


Figure 8.11: CD spectra of histatin 5 (gray) and the two conjugates (blue and red) in TFE.

8.5 Examining KEIF

The interest of KEIF sparked from the suggestion that it might be intrinsically disordered, and from the fact that its function in MgtA was unknown. Because of our previous experience with small IDPs, and because of the limited computer resources, it was decided that we would study KEIF only, that is, without the rest of the MgtA protein. Initial investigation was done based on the primary structure of the peptide, which predicted that KEIF would assume globular conformations in aqueous solution and that it was unlikely to be able to traverse a cell membrane on its own because of its low hydrophobicity.

Bulk properties of KEIF were investigated using SAXS, CD spectroscopy, and atomistic MD simulations. Both the SAXS results and the simulations agreed on the conformational properties of KEIF, which were found to be a fully flexible and slightly extended peptide, despite being predicted to form globules in aqueous solution. Additionally, *ensemble optimization method* (EOM) [156, 157] was used to fit theoretical scattering intensities to the experimental SAXS data. There was good correspondence between the scattering curves from the MD simulations and the EOM. The SAXS curves are depicted in Figure 8.12, where both the from factor (a) and the Kratky plot (b) is shown.

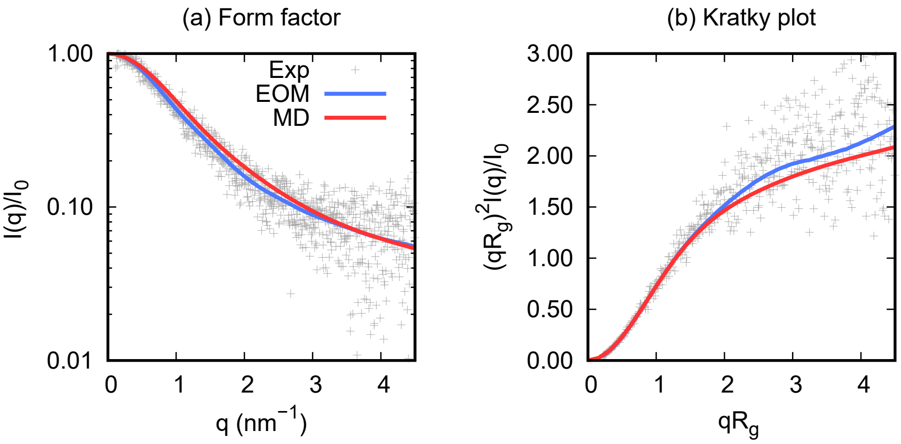


Figure 8.12: (a) KEIF form factor, and (b) Kratky plot of the experimental SAXS results (grey) compared to EOM (blue) and MD simulations (red).

The CD spectroscopy showed that KEIF was dominated by β-sheet structure and "random coils" in aqueous solution. Analysis of the spectra was done using BeStSel [158, 159], which does not include the PPII helix, although the shape of the spectra indicates the presence of PPII structure. Similarly to histatin 5, KEIF also adopt more α-helical conformations in TFE. By analyzing the backbone dihedral angles from the simulated trajectory, the simulations confirmed that KEIF possess a significant amount of PPII helical structure under bulk conditions, see Figure 8.13. The simulations also revealed some β-structure but only little α-helical structure.

Because KEIF is part of a membrane protein, it was also of interest to study its interactions with model membranes. This was investigated exclusively by experimental methods where interactions between KEIF and neutral and anionic vesicles were studied using *dynamic light scattering* (DLS), *laser Doppler velocimetry* (LDV), *cryogenic transmission electron microscopy* (cryo-TEM), and CD spectroscopy. The results suggested that KEIF does not possess any means for significant interactions with neutral model membranes. Interaction with anionic vesicles were however observed by all methods, which together suggested that KEIF adsorbs to the surface of anionic membranes. Additionally, the CD spectroscopy revealed a conformational change in KEIF in the presence of the anionic vesicles; the α-helical content increased to 10 % from previously being completely absent. Thus, there is a possibility that KEIF actually is significant

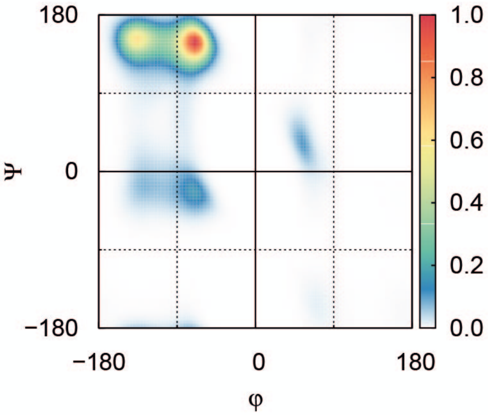


Figure 8.13: A Ramachandran plot obtained from analysis of the backbone dihedral angles in the trajectory from MD simulations of KEIF. The plot is normalized for a maximum intensity of 1.

for the function of MgtA in the bacterial cells. For example, it was suggested that KEIF could be part of the mechanism that anchors the MgtA protein in the bilayer as illustrated in Figure 8.14, although further research is needed to confirm this hypothesis.

8.6 Outlook

The research of this work has mainly focused on studying the conformational properties of single chain IDPs, in particularly histatin 5 and KEIF. Because of its role as a saliva protein, and because of its antimicrobial properties, this research concerning histatin 5 might, in the long run, aid the development of new saliva substitutes and new antimicrobial drugs. On a similar theme, but in a different way, an understanding of how bacterial proteins like MgtA work will also aid the development of new antibiotics.

Although the single chain properties are important, it is also important to know how the conformational properties are affected by higher concentrations and how they might change in the presence of surfaces and membranes. The effect of changing the conditions accordingly may be observed by applying different experimental techniques. However, because of the flexible nature of the IDPs, computational studies will still be vital to elucidate the conformational details of the IDPs, which in turn are needed to gain a holistic understanding of how the IDPs perform their biological functions. New force fields for MD simulations of IDPs are continuously being developed and it is important to remain updated on this field to ensure the chosen methods work well for the specific systems.

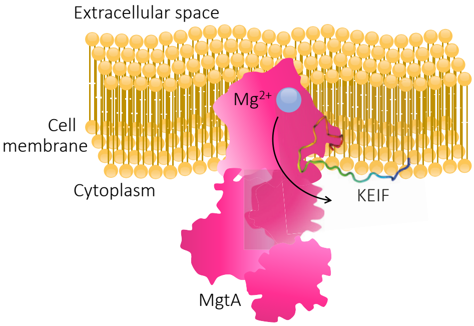


Figure 8.14: A schematic cross-section of MgtA (pink) in a cell membrane (yellow), illustrating how KEIF (rainbow cartoon) could aid in anchoring MgtA to the cell membrane.

Histatin 5 is a well-studied peptide that is frequently involved in new research studies. Nevertheless, its complete mechanism still needs to be defined for its full potential to be unlocked. Apart from being a useful antimicrobial agent, it has also been shown to possess other interesting properties, such as being able to spontaneously forming cushioned model membranes [160]. Therefore, further studies should be performed investigating surface and membrane interactions of histatin 5 and its derivatives. One presently ongoing computational project in our group concerns the interaction of the P-113 peptide with model membranes.

In our study of KEIF, we hypothesized that this IDR might be part of a mechanism that anchors the MgtA protein to the bacterial cell membrane. This hypothesis should be further explored by for example performing experiments of MgtA with and without the KEIF region, in the vicinity of model membranes.

9. References

[1] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of molecular biology*, vol. 157, no. 1, pp. 105–132, 1982.

[2] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*. W. H. Freeman and Company, 7 ed., 2012.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.

[4] D. Ma, N. R. Brandon, T. Cui, V. Bondarenko, C. Canlas, J. S. Johansson, P. Tang, and Y. Xu, "Four-α-helix bundle with designed anesthetic binding pockets. part I: structural and dynamical analyses," *Biophysical journal*, vol. 94, no. 11, pp. 4454–4463, 2008.

[5] Y. Hori, H. Fujiwara, W. Fujiwara, and K. Makabe, "Grafting a short chameleon sequence from αb crystallin into a β-sheet scaffold protein," *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 5, pp. 416–424, 2019.

[6] J. R. Tame and B. Vallone, "The structures of deoxy human haemoglobin and the mutant Hb Tyrα42His at 120 K," *Acta Crystallographica Section D: Biological Crystallography*, vol. 56, no. 7, pp. 805–811, 2000.

[7] V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic, and A. K. Dunker, "Unfoldomics of human diseases: linking protein intrinsic disorder with diseases," *BMC genomics*, vol. 10, no. 1, pp. 1–17, 2009.

[8] V. N. Uversky, J. R. Gillespie, and A. L. Fink, "Why are "natively unfolded" proteins unstructured under physiologic conditions?," *Proteins: structure, function, and bioinformatics*, vol. 41, no. 3, pp. 415–427, 2000.

[9] P. Tompa, "Intrinsically unstructured proteins," *Trends in biochemical sciences*, vol. 27, no. 10, pp. 527–533, 2002.

[10] S. Krimm and M. L. Tiffany, "The circular dichroism spectrum and structure of unordered polypeptides and proteins," *Israel Journal of Chemistry*, vol. 12, no. 1-2, pp. 189–200, 1974.

[11] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions," *Nature reviews Molecular cell biology*, vol. 6, no. 3, pp. 197–208, 2005.

[12] F. Zhu, J. Kapitan, G. E. Tranter, P. D. Pudney, N. W. Isaacs, L. Hecht, and L. D. Barron, "Residual structure in disordered peptides and unfolded proteins from multivariate analysis and ab initio simulation of Raman optical activity data," *Proteins: Structure, Function, and Bioinformatics*, vol. 70, no. 3, pp. 823–833, 2008.

[13] A. A. Adzhubei, M. J. Sternberg, and A. A. Makarov, "Polyproline-II helix in proteins: structure and function," *Journal of molecular biology*, vol. 425, no. 12, pp. 2100–2132, 2013.

[14] S. Kakinoki, Y. Hirano, and M. Oka, "On the stability of polyproline-I and II structures of proline oligopeptides," *Polymer Bulletin*, vol. 53, no. 2, pp. 109–115, 2005.

[15] Z. Shi, K. Chen, Z. Liu, and N. R. Kallenbach, "Conformation of the backbone in unfolded proteins," *Chemical reviews*, vol. 106, no. 5, pp. 1877–1897, 2006.

[16] V. N. Uversky, "Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding," *The protein journal*, vol. 28, no. 7-8, pp. 305–325, 2009.

[17] R. K. Das and R. V. Pappu, "Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 33, pp. 13392–13397, 2013.

[18] R. K. Das, K. M. Ruff, and R. V. Pappu, "Relating sequence encoded information to form and function of intrinsically disordered proteins," *Curr. Opin. Struct. Biol.*, vol. 32, pp. 102–112, 2015.

[19] K. M. Ruff, "Predicting conformational properties of intrinsically disordered proteins from sequence," in *Intrinsically Disordered Proteins*, pp. 347–389, Springer, 2020.

[20] P. Bernadó and D. I. Svergun, "Structural insights into intrinsically disordered proteins by small-angle X-ray scattering," in *Instrumental analysis of intrinsically disordered proteins: Assessing structure and conformation*, pp. 451–476, John Wiley and Sons Hoboken, New Jersey, 2010.

[21] E. V. Kuzmenkina, C. D. Heyes, and G. U. Nienhaus, "Single-molecule FRET study of denaturant induced unfolding of RNase H," *Journal of molecular biology*, vol. 357, no. 1, pp. 313–324, 2006.

[22] V. Receveur-Bréchot and D. Durand, "How random are intrinsically disordered proteins? A small angle scattering perspective," *Current Protein and Peptide Science*, vol. 13, no. 1, pp. 55–75, 2012.

[23] M. R. Jensen, R. W. Ruigrok, and M. Blackledge, "Describing intrinsically disordered proteins at atomic resolution by NMR," *Current opinion in structural biology*, vol. 23, no. 3, pp. 426–435, 2013.

[24] V. N. Uversky, V. Dave, L. M. Iakoucheva, P. Malaney, S. J. Metallo, R. R. Pathak, and A. C. Joerger, "Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases," *Chemical reviews*, vol. 114, no. 13, pp. 6844–6879, 2014.

[25] S. P. Humphrey and R. T. Williamson, "A review of saliva: normal composition, flow, and function," *The Journal of prosthetic dentistry*, vol. 85, no. 2, pp. 162–169, 2001.

[26] L. C. Schenkels, E. C. Veerman, and A. V. Nieuw Amerongen, "Biochemical composition of human saliva in relation to other mucosal fluids," *Critical reviews in oral biology & medicine*, vol. 6, no. 2, pp. 161–175, 1995.

[27] J. Tenovuo, "Clinical applications of antimicrobial host proteins lactoperoxidase, lysozyme and lactoferrin in xerostomia: efficacy and safety," *Oral diseases*, vol. 8, no. 1, pp. 23–29, 2002.

[28] H. Välimaa, M. Waris, V. Hukkanen, M. Blankenvoorde, A. Nieuw Amerongen, and J. Tenovuo, "Salivary defense factors in herpes simplex virus infection," *Journal of dental research*, vol. 81, no. 6, pp. 416–421, 2002.

[29] Y. Yamaguchi, M. Semmel, L. Stanislawski, A. Strosberg, and M. Stanislawski, "Virucidal effects of glucose oxidase and peroxidase or their protein conjugates on human immunodeficiency virus type 1.," *Antimicrobial agents and chemotherapy*, vol. 37, no. 1, pp. 26–31, 1993.

[30] M. J. Levine, "Development of artificial salivas," *Critical Reviews in Oral Biology & Medicine*, vol. 4, no. 3, pp. 279–286, 1993.

[31] J. W. Millsop, E. A. Wang, and N. Fazel, "Etiology, evaluation, and management of xerostomia," *Clinics in Dermatology*, vol. 35, no. 5, pp. 468–476, 2017.

[32] J. Guggenheimer and P. A. Moore, "Xerostomia: etiology, recognition and treatment," *The Journal of the American Dental Association*, vol. 134, no. 1, pp. 61–69, 2003.

[33] H. Mese and R. Matsuo, "Salivary secretion, taste and hyposalivation," *Journal of oral rehabilitation*, vol. 34, no. 10, pp. 711–723, 2007.

[34] P. C. Fox, "Xerostomia: recognition and management," *Dental Assistant*, vol. 77, no. 5, p. 18, 2008.

[35] J. M. Plemons, I. Al-Hashimi, and C. L. Marek, "Managing xerostomia and salivary gland hypofunction: executive summary of a report from the American Dental Association Council on Scientific Affairs," *The Journal of the American Dental Association*, vol. 145, no. 8, pp. 867–873, 2014.

[36] S. S. Anttila, M. L. Knuuttila, and T. K. Sakki, "Depressive symptoms as an underlying factor of the sensation of dry mouth," *Psychosomatic medicine*, vol. 60, no. 2, pp. 215–218, 1998.

[37] M. Bergdahl and J. Bergdahl, "Low unstimulated salivary flow and subjective oral dryness: association with medication, anxiety, depression, and stress," *Journal of dental research*, vol. 79, no. 9, pp. 1652–1658, 2000.

[38] L. M. Sreebny and S. S. Schwartz, "A reference guide to drugs and dry mouth–2nd edition," *Gerodontology*, vol. 14, no. 1, pp. 33–47, 1997.

[39] C. Scully Cbe, "Drug effects on salivary glands: dry mouth," *Oral diseases*, vol. 9, no. 4, pp. 165–176, 2003.

[40] J. Karbach, C. Walter, and B. Al-Nawas, "Evaluation of saliva flow rates, Candida colonization and susceptibility of Candida strains after head and neck radiation," *Clinical oral investigations*, vol. 16, no. 4, pp. 1305–1312, 2012.

[41] P. A. Moore, J. Guggenheimer, K. R. Etzel, R. J. Weyant, and T. Orchard, "Type 1 diabetes mellitus, xerostomia, and salivary flow rates," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 92, no. 3, pp. 281–291, 2001.

[42] M. Tanasiewicz, T. Hildebrandt, and I. Obersztyn, "Xerostomia of various etiologies: A review of the literature.," *Advances in clinical and experimental medicine: official organ Wroclaw Medical University*, vol. 25, no. 1, p. 199, 2016.

[43] R. M. López-Pintor, E. Casañas, J. González-Serrano, J. Serrano, L. Ramírez, L. de Arriba, and G. Hernández, "Xerostomia, hyposalivation, and salivary flow in diabetes patients," *Journal of diabetes research*, vol. 2016, 2016.

[44] R. Billings, H. Proskin, and M. Moss, "Xerostomia and associated factors in a community-dwelling adult population," *Community dentistry and oral epidemiology*, vol. 24, no. 5, pp. 312–316, 1996.

[45] D. Greenspan, "Xerostomia: diagnosis and management.," *Oncology (Williston Park, NY)*, vol. 10, no. 3 Suppl, p. 7, 1996.

[46] M. Rad, S. Kakoie, F. N. Brojeni, and N. Pourdamghan, "Effect of long-term smoking on whole-mouth salivary flow rate and oral health," *Journal of dental research, dental clinics, dental prospects*, vol. 4, no. 4, p. 110, 2010.

[47] S. F. Cassolato and R. S. Turnbull, "Xerostomia: clinical aspects and treatment," *Gerodontology*, vol. 20, no. 2, pp. 64–77, 2003.

[48] R. Cohen, F. Roth, E. Delgado, D. Ahearn, and M. Kalser, "Fungal flora of the normal human small and large intestine," *New England Journal of Medicine*, vol. 280, no. 12, pp. 638–641, 1969.

[49] J. A. Vazquez and J. D. Sobel, "Mucosal candidiasis.," *Infectious disease clinics of North America*, vol. 16, no. 4, p. 793, 2002.

[50] P. Eggimann, J. Garbino, and D. Pittet, "Epidemiology of Candida species infections in critically ill non-immunosuppressed patients," *The Lancet infectious diseases*, vol. 3, no. 11, pp. 685–702, 2003.

[51] V. J. Fraser, M. Jones, J. Dunkel, S. Storfer, G. Medoff, and W. C. Dunagan, "Candidemia in a tertiary care hospital: epidemiology, risk factors, and predictors of mortality," *Clinical Infectious Diseases*, vol. 15, no. 3, pp. 414–421, 1992.

[52] D. P. Lew and F. A. Waldvogel, "Osteomyelitis," *The Lancet*, vol. 364, no. 9431, pp. 369–379, 2004.

[53] P. Vergidis, C. J. Clancy, R. K. Shields, S. Y. Park, B. N. Wildfeuer, R. L. Simmons, and M. H. Nguyen, "Intra-abdominal candidiasis: the importance of early source control and antifungal treatment," *PLoS One*, vol. 11, no. 4, p. e0153247, 2016.

[54] S. I. Blot, K. H. Vandewoude, and J. J. De Waele, "Candida peritonitis," *Current opinion in critical care*, vol. 13, no. 2, pp. 195–199, 2007.

[55] B. J. Kullberg and M. C. Arendrup, "Invasive candidiasis," *New England Journal of Medicine*, vol. 373, no. 15, pp. 1445–1456, 2015.

[56] P. G. Pappas, M. S. Lionakis, M. C. Arendrup, L. Ostrosky-Zeichner, and B. J. Kullberg, "Invasive candidiasis," *Nature Reviews Disease Primers*, vol. 4, no. 1, pp. 1–20, 2018.

[57] A. A. Cleveland, L. H. Harrison, M. M. Farley, R. Hollick, B. Stein, T. M. Chiller, S. R. Lockhart, and B. J. Park, "Declining incidence of candidemia and the shifting epidemiology of candida resistance in two US metropolitan areas, 2008–2013: results from population-based surveillance," *PLoS one*, vol. 10, no. 3, p. e0120452, 2015.

[58] W. Siqueira, H. Margolis, E. Helmerhorst, F. Mendes, and F. Oppenheim, "Evidence of intact histatins in the in vivo acquired enamel pellicle," *Journal of dental research*, vol. 89, no. 6, pp. 626–630, 2010.

[59] F. Oppenheim, T. Xu, F. McMillian, S. Levitz, R. Diamond, G. Offner, and R. Troxler, "Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on Candida albicans.," *J. Biol. Chem.*, vol. 263, no. 16, pp. 7472–7477, 1988.

[60] P. A. Raj, M. Edgerton, and M. Levine, "Salivary histatin 5: dependence of sequence, chain length, and helical conformation for candidacidal activity," *J. Biol. Chem.*, vol. 265, no. 7, pp. 3898–3905, 1990.

[61] H. Tsai, P. A. Raj, and L. A. Bobek, "Candidacidal activity of recombinant human salivary histatin-5 and variants.," *Infect. Immun.*, vol. 64, no. 12, pp. 5000–5007, 1996.

[62] H. Tsai and L. Bobek, "Human salivary histatins: promising anti-fungal therapeutic agents," *Crit. Rev. Oral Biol. Med.*, vol. 9, no. 4, pp. 480–497, 1998.

[63] A. Kurut, J. Henriques, J. Forsman, M. Skepö, and M. Lund, "Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk," *Proteins: Struct., Funct., Bioinf.*, vol. 82, no. 4, pp. 657–667, 2014.

[64] D. M. Rothstein, P. Spacciapoli, L. T. Tran, T. Xu, F. D. Roberts, M. Dalla Serra, D. K. Buxton, F. G. Oppenheim, and P. Friden, "Anticandida activity is retained in P-113, a 12-amino-acid fragment of histatin 5," *Antimicrob. Agents Chemother.*, vol. 45, no. 5, pp. 1367–1373, 2001.

[65] E. J. Helmerhorst, W. Van't Hof, P. Breeuwer, E. C. Veerman, T. Abee, R. F. Troxler, A. V. Nieuw Amerongen, and F. G. Oppenheim, "Characterization of histatin 5 with respect to amphipathicity, hydrophobicity, and effects on cell and mitochondrial membrane integrity excludes a candidacidal mechanism of pore formation," *Journal of Biological Chemistry*, vol. 276, no. 8, pp. 5643–5649, 2001.

[66] M. Edgerton, S. E. Koshlukova, T. E. Lo, B. G. Chrzan, R. M. Straubinger, and P. A. Raj, "Candidacidal activity of salivary histatins identification of a histatin 5-binding protein on candida albicans," *Journal of Biological Chemistry*, vol. 273, no. 32, pp. 20438–20447, 1998.

[67] S. E. Koshlukova, T. L. Lloyd, M. W. Araujo, and M. Edgerton, "Salivary histatin 5 induces non-lytic release of ATP from Candida albicans leading to cell death," *Journal of Biological Chemistry*, vol. 274, no. 27, pp. 18872–18879, 1999.

[68] A. B. Mochon and H. Liu, "The antimicrobial peptide histatin-5 causes a spatially restricted disruption on the Candida albicans surface, allowing rapid entry of the peptide into the cytoplasm," *PLoS Pathog*, vol. 4, no. 10, p. e1000190, 2008.

[69] H. L. Norris, R. Kumar, C. Y. Ong, D. Xu, and M. Edgerton, "Zinc binding by histatin 5 promotes fungicidal membrane disruption in C. albicans and C. glabrata," *Journal of Fungi*, vol. 6, no. 3, p. 124, 2020.

[70] E. J. Helmerhorst, P. Breeuwer, W. Van't Hof, E. Walgreen-Weterings, L. C. Oomen, E. C. Veerman, A. V. Nieuw Amerongen, and T. Abee, "The cellular target of histatin 5 on Candida albicans is the energized mitochondrion," *Journal of Biological Chemistry*, vol. 274, no. 11, pp. 7286–7291, 1999.

[71] A. L. den Hertog, H. W. W. F. Sang, R. Kraayenhof, J. G. Bolscher, W. Van't Hof, E. C. Veerman, and A. V. Nieuw Amerongen, "Interactions of histatin 5 and histatin 5-derived peptides with liposome membranes: surface effects, translocation and permeabilization," *Biochemical Journal*, vol. 379, no. 3, pp. 665–672, 2004.

[72] A. L. den Hertog, J. van Marle, H. A. van Veen, W. Van't Hof, J. G. Bolscher, E. C. Veerman, and A. V. Niew Amerongen, "Candidacidal effects of two antimicrobial peptides: histatin 5 causes small membrane defects, but LL-37 causes massive disruption of the cell membrane," *Biochemical Journal*, vol. 388, no. 2, pp. 689–695, 2005.

[73] A. Ruissen, J. Groenink, W. Van't Hof, E. Walgreen-Weterings, J. Van Marle, H. Van Veen, W. Voorhout, E. Veerman, and A. Nieuw Amerongen, "Histatin 5 and derivatives: Their localization and effects on the ultra-structural level," *Peptides*, vol. 23, no. 8, pp. 1391–1399, 2002.

[74] R. Kumar, S. Chadha, D. Saraswat, J. S. Bajwa, R. A. Li, H. R. Conti, and M. Edgerton, "Histatin 5 uptake by Candida albicans utilizes polyamine transporters Dur3 and Dur31 proteins," *J. Biol. Chem.*, vol. 286, no. 51, pp. 43748–43758, 2011.

[75] S. Tati, W. S. Jang, R. Li, R. Kumar, S. Puri, and M. Edgerton, "Histatin 5 resistance of candida glabrata can be reversed by insertion of candida albicans polyamine transporter-encoding genes DUR3 and DUR31," *PloS one*, vol. 8, no. 4, p. e61480, 2013.

[76] S. Tati, R. Li, S. Puri, R. Kumar, P. Davidow, and M. Edgerton, "Histatin 5-spermidine conjugates have enhanced fungicidal activity and efficacy as a topical therapeutic for oral candidiasis," *Antimicrobial agents and chemotherapy*, vol. 58, no. 2, pp. 756–766, 2014.

[77] D. G. Kehres and M. E. Maguire, "Structure, properties and regulation of magnesium transport proteins," *Biometals*, vol. 15, no. 3, pp. 261–270, 2002.

[78] M. E. Maguire, "Magnesium transporters: properties, regulation and structure," *Front Biosci*, vol. 11, pp. 3149–3163, 2006.

[79] V. V. Lunin, E. Dobrovetsky, G. Khutoreskaya, R. Zhang, A. Joachimiak, D. A. Doyle, A. Bochkarev, M. E. Maguire, A. M. Edwards, and C. M. Koth, "Crystal structure of the corA Mg 2+ transporter," *Nature*, vol. 440, no. 7085, pp. 833–837, 2006.

[80] M. Hattori, Y. Tanaka, S. Fukai, R. Ishitani, and O. Nureki, "Crystal structure of the MgtE Mg 2+ transporter," *Nature*, vol. 448, no. 7157, pp. 1072–1075, 2007.

[81] W. Kühlbrandt, "Biology, structure and mechanism of P-type ATPases," *Nature reviews Molecular cell biology*, vol. 5, no. 4, pp. 282–295, 2004.

[82] S. Subramani, H. Perdreau-Dahl, and J. P. Morth, "The magnesium transporter A is activated by cardiolipin and is highly sensitive to free magnesium in vitro," *Elife*, vol. 5, p. e11407, 2016.

[83] S. Tzu and M. Mushashi, *The Art of War & The Book of Five Rings*. BN Publishing, 2007. Translated from Chinese by Lionel Giles and D.W.

[84] R. B. Best, N.-V. Buchete, and G. Hummer, "Are current molecular dynamics force fields too helical?," *Biophysical journal*, vol. 95, no. 1, pp. L07–L09, 2008.

[85] S. Piana, J. L. Klepeis, and D. E. Shaw, "Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations," *Current opinion in structural biology*, vol. 24, pp. 98–105, 2014.

[86] J. Henriques, C. Cragnell, and M. Skepö, "Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment," *J. Chem. Theory Comput.*, vol. 11, no. 7, pp. 3420–3431, 2015.

[87] R. B. Best and G. Hummer, "Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides," *The journal of physical chemistry B*, vol. 113, no. 26, pp. 9004–9015, 2009.

[88] R. B. Best and J. Mittal, "Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse," *The Journal of Physical Chemistry B*, vol. 114, no. 46, pp. 14916–14923, 2010.

[89] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins: Struct., Funct., Bioinf.*, vol. 78, no. 8, pp. 1950–1958, 2010.

[90] R. B. Best, W. Zheng, and J. Mittal, "Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association," *J. Chem. Theory Comput.*, vol. 10, no. 11, pp. 5113–5124, 2014.

[91] F. Jiang, C.-Y. Zhou, and Y.-D. Wu, "Residue-specific force field based on the protein coil library. RSFF1: modification of OPLS-AA/L," *The journal of physical chemistry B*, vol. 118, no. 25, pp. 6983–6998, 2014.

[92] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB," *Journal of chemical theory and computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

[93] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: an improved force field for folded and intrinsically disordered proteins," *Nature methods*, vol. 14, no. 1, pp. 71–73, 2017.

[94] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, *et al.*, "OPLS3: a force field providing broad coverage of drug-like small molecules and proteins," *Journal of chemical theory and computation*, vol. 12, no. 1, pp. 281–296, 2016.

[95] D. Song, W. Wang, W. Ye, D. Ji, R. Luo, and H.-F. Chen, "ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins," *Chemical biology & drug design*, vol. 89, no. 1, pp. 5–15, 2017.

[96] P. Robustelli, S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," *Proceedings of the National Academy of Sciences*, vol. 115, no. 21, pp. E4758–E4766, 2018.

[97] H. Liu, D. Song, H. Lu, R. Luo, and H.-F. Chen, "Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins," *Chemical biology & drug design*, vol. 92, no. 4, pp. 1722–1735, 2018.

[98] H. Liu, D. Song, Y. Zhang, S. Yang, R. Luo, and H.-F. Chen, "Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins," *Physical Chemistry Chemical Physics*, vol. 21, no. 39, pp. 21918–21931, 2019.

[99] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, "Water dispersion interactions strongly influence simulated structural properties of disordered protein states," *J. Phys. Chem. B*, vol. 119, no. 16, pp. 5113–5123, 2015.

[100] D. Mercadante, S. Milles, G. Fuertes, D. I. Svergun, E. A. Lemke, and F. Gräter, "Kirkwood–Buff approach rescues overcollapse of a disordered protein in canonical protein force fields," *J. Phys. Chem. B*, vol. 119, no. 25, pp. 7975–7984, 2015.

[101] S.-H. Chong, P. Chatterjee, and S. Ham, "Computer simulations of intrinsically disordered proteins," *Annual Review of Physical Chemistry*, vol. 68, pp. 117–134, 2017.

[102] R. B. Best, "Computational and theoretical advances in studies of intrinsically disordered proteins," *Current Opinion in Structural Biology*, vol. 42, pp. 147–154, 2017.

[103] P. S. Nerenberg and T. Head-Gordon, "New developments in force fields for biomolecular simulations," *Current opinion in structural biology*, vol. 49, pp. 129–138, 2018.

[104] T. L. Hill, *An introduction to statistical thermodynamics*. Courier Corporation, 1986.

[105] M. Allen and D. Tildesley, *Computer simulation of liquids*, vol. 18. Oxford university press, 1989.

[106] D. I. Svergun and M. H. Koch, "Small-angle scattering studies of biological macromolecules in solution," *Reports on Progress in Physics*, vol. 66, no. 10, p. 1735, 2003.

[107] D. F. Evans and H. Wennerström, *The colloidal domain: where physics, chemistry, biology, and technology meet*. Wiley-Vch New York, 2 ed., 1999.

[108] J. N. Israelachvili, *Intermolecular and surface forces*. Academic press, 2011.

[109] H. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, vol. 91, no. 1, pp. 43–56, 1995.

[110] E. Lindahl, B. Hess, and D. van der Spoel, “GROMACS 3.0: a package for molecular simulation and trajectory analysis,” *J. Mol. Model.*, vol. 7, no. 8, pp. 306–317, 2001.

[111] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, and H. Berendsen, “GROMACS: fast, flexible, and free,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, 2005.

[112] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation,” *J. Chem. Theory Comput.*, vol. 4, no. 3, pp. 435–447, 2008.

[113] G. Moss, “Basic terminology of stereochemistry (IUPAC Recommendations 1996),” *Pure and applied chemistry*, vol. 68, no. 12, pp. 2193–2222, 1996.

[114] C. Cragnell, E. Rieloff, and M. Skepö, “Utilizing coarse-grained modeling and monte carlo simulations to evaluate the conformational ensemble of intrinsically disordered proteins and regions,” *Journal of molecular biology*, 2018.

[115] J. Reščič and P. Linse, “MOLSIM: A modular molecular simulation software,” *Journal of computational chemistry*, vol. 36, no. 16, pp. 1259–1274, 2015.

[116] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[117] D. van Der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 4.6.7*. www.gromacs.org, 2014.

[118] H. Berendsen and W. Van Gunsteren, “Practical algorithms for dynamic simulations,” *Molecular-dynamics simulation of statistical-mechanical systems*, pp. 43–65, 1986.

[119] B. Hess, H. Bekker, H. Berendsen, and J. Fraaije, “LINCS: a linear constraint solver for molecular simulations,” *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, 1997.

[120] T. Darden, D. York, and L. Pedersen, “Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems,” *J. Chem. Phys.*, vol. 98, p. 10089, 1993.

[121] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J. Chem. Phys.*, vol. 126, p. 014101, 2007.

[122] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys. (Melville, NY, U. S.)*, vol. 52, p. 7182, 1981.

[123] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[124] A. Grossfield and D. M. Zuckerman, "Chapter 2 Quantifying uncertainty and sampling quality in biomolecular simulations," in *Annual reports in computational chemistry*, vol. 5, pp. 23–48, Elsevier, 2009.

[125] H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data," *The Journal of Chemical Physics*, vol. 91, no. 1, pp. 461–466, 1989.

[126] B. Hess, "Determining the shear viscosity of model liquids from molecular dynamics simulations," *The Journal of chemical physics*, vol. 116, no. 1, pp. 209–217, 2002.

[127] R. Stepto, T. Chang, P. Kratochvíl, M. Hess, K. Horie, T. Sato, and J. Vohlídal, "Definitions of terms relating to individual macromolecules, macromolecular assemblies, polymer solutions, and amorphous bulk polymers (IUPAC Recommendations 2014)," *Pure and Applied Chemistry*, vol. 87, no. 1, pp. 71–120, 2015.

[128] S. R. Campos and A. M. Baptista, "Conformational analysis in a multidimensional energy landscape: study of an arginylglutamate repeat," *J. Phys. Chem. B*, vol. 113, no. 49, pp. 15989–16001, 2009.

[129] I. Jolliffe, "Principal component analysis," *Technometrics*, vol. 45, no. 3, p. 276, 2003.

[130] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983.

[131] Y. Mansiaux, A. P. Joseph, J.-C. Gelly, and A. G. de Brevern, "Assignment of polyproline II conformation and analysis of sequence–structure relationship," *PloS one*, vol. 6, no. 3, p. e18401, 2011.

[132] H. Schnablegger and Y. Singh, "The SAXS guide: getting acquainted with the principles," *Austria: Anton Paar GmbH*, 2011.

[133] D. A. Jacques and J. Trewhella, "Small-angle scattering for structural biology? Expanding the frontier while avoiding the pitfalls," *Protein Science*, vol. 19, no. 4, pp. 642–657, 2010.

[134] S. M. Kelly, T. J. Jess, and N. C. Price, "How to study proteins by circular dichroism," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1751, no. 2, pp. 119–139, 2005.

[135] V. N. Uversky and A. K. Dunker, *Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools*. Springer, 2012.

[136] L. Whitmore, A. J. Miles, L. Mavridis, R. W. Janes, and B. A. Wallace, "PCDDB: new developments at the protein circular dichroism data bank," *Nucleic acids research*, vol. 45, no. D1, pp. D303–D307, 2017.

[137] J. G. Lees, A. J. Miles, F. Wien, and B. Wallace, "A reference database for circular dichroism spectroscopy covering fold and secondary structure space," *Bioinformatics*, vol. 22, no. 16, pp. 1955–1962, 2006.

[138] S. J. Smerdon, S. Krzywda, A. M. Brzozowski, G. J. Davies, A. J. Wilkinson, A. Brancaccio, F. Cutruzzola, C. T. Allocatelli, M. Brunori, R. E. J. Brantley, *et al.*, "Interactions among residues CD3, E7, E10, and E11 in myoglobins: attempts to simulate the ligand-binding properties of Aplysia myoglobin," *Biochemistry*, vol. 34, no. 27, pp. 8715–8725, 1995.

[139] A. Abdul-Gader, A. J. Miles, and B. A. Wallace, "A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy," *Bioinformatics*, vol. 27, no. 12, pp. 1630–1636, 2011.

[140] C. J. Oomen, P. van Ulsen, P. Van Gelder, M. Feijen, J. Tommassen, and P. Gros, "Structure of the translocator domain of a bacterial autotransporter," *The EMBO journal*, vol. 23, no. 6, pp. 1257–1266, 2004.

[141] J. Tolchard, S. J. Walpole, A. J. Miles, R. Maytum, L. A. Eaglen, T. Hackstadt, B. A. Wallace, and T. M. Blumenschein, "The intrinsically disordered Tarp protein from chlamydia binds actin with a partially preformed helix," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.

[142] J. L. Lopes, A. J. Miles, L. Whitmore, and B. A. Wallace, "Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses," *Protein Science*, vol. 23, no. 12, pp. 1765–1772, 2014.

[143] C. Richardson, M. Johnsson, P. Raj, M. Levine, and G. Nancollas, "The influence of histatin-5 fragments on the mineralization of hydroxyapatite," *Archives of oral biology*, vol. 38, no. 11, pp. 997–1002, 1993.

[144] A. Yin, H. Margolis, J. Grogan, Y. Yao, R. Troxler, and F. Oppenheim, “Physical parameters of hydroxyapatite adsorption and effect on candidacidal activity of histatins,” *Archives of oral biology*, vol. 48, no. 5, pp. 361–368, 2003.

[145] M. Yoshinari, T. Kato, K. Matsuzaka, T. Hayakawa, T. Inoue, Y. Oda, K. Okuda, and M. Shimono, “Adsorption behavior of antimicrobial peptide histatin 5 on PMMA,” *Journal of Biomedical Materials Research Part B: Applied Biomaterials: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials*, vol. 77, no. 1, pp. 47–54, 2006.

[146] K. Hyltegren, M. Polimeni, M. Skep⊠, and M. Lund, “Integrating all-atom and coarse-grained simulations–toward understanding of idps at surfaces,” *Journal of chemical theory and computation*, vol. 16, no. 3, pp. 1843–1853, 2020.

[147] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, p. 926, 1983.

[148] J. A. Marsh and J. D. Forman-Kay, “Sequence determinants of compaction in intrinsically disordered proteins,” *Biophysical journal*, vol. 98, no. 10, pp. 2383–2390, 2010.

[149] M. Nygaard, B. B. Kragelund, E. Papaleo, and K. Lindorff-Larsen, “An efficient method for estimating the hydrodynamic radius of disordered protein conformations,” *Biophysical journal*, vol. 113, no. 3, pp. 550–557, 2017.

[150] R. Chebrek, S. Leonard, A. G. de Brevern, and J.-C. Gelly, “PolyprOnline: polyproline helix II and secondary structure assignment database,” *Database*, vol. 2014, 2014.

[151] C. Malmberg and A. Maryott, “Dielectric constant of water from 00 to 1000 C,” *J. Res. Nat. Bur. Stand*, vol. 56, pp. 369131–8, 1956.

[152] B. B. Owen, R. C. Miller, C. E. Milner, and H. L. Cogan, “The dielectric constant of water as a function of temperature and pressure,” *The Journal of Physical Chemistry*, vol. 65, no. 11, pp. 2065–2070, 1961.

[153] M. Uematsu and E. Frank, “Static dielectric constant of water and steam,” *Journal of Physical and Chemical Reference Data*, vol. 9, no. 4, pp. 1291–1306, 1980.

[154] D. P. Fernández, A. Goodwin, E. W. Lemmon, J. Levelt Sengers, and R. Williams, “A formulation for the static permittivity of water and steam at temperatures from 238 K to 873 K at pressures up to 1200 MPa, including derivatives and

Debye–Hückel coefficients," *Journal of Physical and Chemical Reference Data*, vol. 26, no. 4, pp. 1125–1166, 1997.

[155] G. Nagy, M. Igaev, N. C. Jones, S. V. Hoffmann, and H. Grubmüller, "SESCA: predicting circular dichroism spectra from protein molecular structures," *Journal of chemical theory and computation*, vol. 15, no. 9, pp. 5087–5102, 2019.

[156] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, "Structural characterization of flexible proteins using small-angle X-ray scattering," *J. Am. Chem. Soc.*, vol. 129, no. 17, pp. 5656–5664, 2007.

[157] G. Tria, H. D. Mertens, M. Kachala, and D. I. Svergun, "Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering," *IUCrJ*, vol. 2, no. 2, pp. 207–217, 2015.

[158] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, and J. Kardos, "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 24, pp. E3095–E3103, 2015.

[159] A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.-H. Lee, Y. Goto, M. Réfrégiers, and J. Kardos, "BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W315–W322, 2018.

[160] Y. Gerelli, A. Eriksson Skog, S. Jephthah, R. J. Welbourn, A. Klechikov, and M. Skepö, "Spontaneous formation of cushioned model membranes promoted by an intrinsically disordered protein," *Langmuir*, vol. 36, no. 15, pp. 3997–4004, 2020.

10. Scientific publications

Paper 1

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Article
pubs.acs.org/jcim

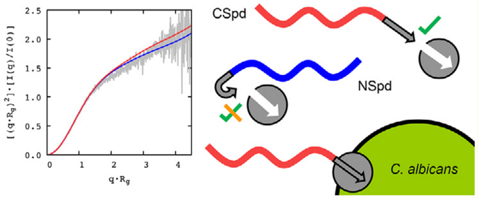
Structural Characterization of Histatin 5−Spermidine Conjugates: A Combined Experimental and Theoretical Study

Stephanie Jephthah,[†] João Henriques,[†] Carolina Cragnell,[†] Sumant Puri,[‡] Mira Edgerton,[§] and Marie Skepö*[,†]

[†]Division of Theoretical Chemistry, Lund University, Post Office Box 124, S−221 00 Lund, Sweden
[‡]Pediatrics and Community Oral Health Sciences, Temple University, Philadelphia, Pennsylvania 19147, United States
[§]Department of Oral Biology, University at Buffalo, Buffalo, New York 14214−3092, United States

S *Supporting Information*

ABSTRACT: Histatin 5 (Hst5) is a naturally occurring antimicrobial peptide that acts as the first line of defense against oral candidiasis. It has been shown that conjugation of the active Hst5 fragment, Hst5_{4-15}, and the polyamine spermidine (Spd) improves the candidacidal effect. Knowledge about the structure of these conjugates is, however, very limited. Thus, the aim of this study was to characterize the structural properties of the Hst5_{4-15}−Spd conjugates by performing atomistic molecular dynamics simulations in combination with small-angle X-ray scattering. It was shown that the Hst5_{4-15}−Spd conjugates adopt extended and slightly rigid random coil conformations without any secondary structure in aqueous solution. It is hypothesized that the increased fungal killing potential of Hst5_{4-15}−Spd, in comparison with the Spd−Hst5_{4-15} conjugate, is due to the more extended conformations of the former, which cause the bonded Spd molecule to be more accessible for recognition by polyamine transporters in the cell.



INTRODUCTION

Invasive fungal infections are known problems for critically ill patients, and the occurrence of these type of infections has been increasing over the years.[1] The most frequently isolated yeast species causing such infections are the *Candida* spp., mainly *Candida albicans*, but other non-*albicans* species have also been identified as the cause of fungal infections.[1−3] Many effective antifungal agents are available for the treatment of invasive candidiasis, such as polyenes, azoles, and echinocandins.[1,4] However, despite the access to these drugs, failures in treatment of invasive candidiasis due to drug-resistant *Candida* species still occur.[1,4,5] There are also other problems related to the drugs, such as frequent side effects, inconvenient drug administration, and unfavorable interactions with other drugs.[1,4] Thus, the need to develop new therapeutic agents is imperative.[5]

Naturally occurring antimicrobial peptides have been proposed as possible therapeutic agents against fungal infections.[5,6] One example of such a peptide is histatin 5 (Hst5), which is a histidine-rich intrinsically disordered protein that can be found in human parotid and submandibular−sublingual saliva.[7−9] Hst5 has various properties that contribute to oral health, and one of the most important is its antifungal action. Because of its effectiveness against fungal infections, especially against the blastospore and the germinated form of *C. albicans*,[10] several studies have focused on investigating the possibility of using enhanced Hst5 variants in therapeutic contexts.[5,11−13]

A study by Rothstein et al.[12] showed that the candidacidal activity of the 12 amino acid Hst5 fragment Hst5_{4-15} is equivalent to that of the full-length protein. Later, Kumar et al.[13] observed that Hst5 utilizes polyamine transporters for intracellular uptake and transport in fungal cells. It was also shown that Hst5 operates like an analogue of the polyamine spermidine (Spd), which uses the same means of transportation into fungal cells. Tati et al.[5] suggested that the uptake and activity of Hst5 into fungal cells might be increased by conjugation with Spd. Two different Hst5_{4-15}−Spd conjugates were studied, consisting of the active Hst5_{4-15} fragment with a Spd molecule conjugated either to the N-terminus (NSpd) or C-terminus (CSpd) of the fragment. It was observed that the conjugates have a greater candidacidal effect against *C. albicans* and *C. glabrata* and are more resistant to protease degradation compared with the unaltered full-length Hst5.[5]

Although their fungal killing potential has been established, very little is known about the structure of the Hst5_{4-15}−Spd conjugates. In fact, until now no direct measurements or simulations have been carried out in order to assess it. Thus, the aim of this study was to characterize the Hst5_{4-15}−Spd conjugates by comparing the results of atomistic molecular dynamics (MD) simulations with small-angle X-ray scattering (SAXS) data.

Received: March 15, 2017
Published: June 6, 2017

DOI: 10.1021/acs.jcim.7b00150
J. Chem. Inf. Model. 2017, 57, 1330−1341

■ MATERIALS AND METHODS

Chemicals. The Hst5_{4-15}−Spd conjugate (CSpd) and the Hst5_{4-15} fragment were obtained from Genemed Synthesis Inc. (San Antonio, TX, USA) as white to off-white powders with 97.46% and 96.18% purity, respectively. The counterion was trifluoroacetate (TFA). A 20 mM Tris (Saveen Werner AB, >99.9% purity, CAS Registry no. 72-2-25) buffer was prepared in Milli-Q water and acidified to pH 7 with HCl. The ionic strength of the buffer solution was adjusted with NaCl (Scharlau, >99.5% purity, CAS Registry no. 7647-14-5). The buffer was filtered through a 0.2 μm hydrophilic polypropylene membrane (Pall Corporation) before any peptide was added. After the peptide was dissolved, a concentration cell (Vivaspin 20, 2 kDa molecular weight cutoff (MWCO), product no. VS02H92, Sartorius Stedim Biotech GmbH, Göttingen, Germany) was used to remove low-molecular-weight impurities. The samples were rinsed with buffer corresponding to at least 10 times the sample volume by centrifugations at 3000 rpm and 18 °C. The samples were also dialyzed (Slide-A-Lyzer MINI, 2 kDa MWCO, product no. 69580, Thermo Scientific, United States) for 12 h to ensure exact background for the SAXS measurements. The studied peptide concentrations ranged from 0.8 to 5.12 mg mL^{-1} in order to overlap the overall physiological protein concentration range in saliva (1−3.5 mg mL^{-1}). The ionic strength of the samples was set to 140 mM to exclude any electrostatic repulsion effects and to facilitate comparison with simulations, where the protein does not interact with its periodic images.

Small-Angle X-ray Scattering Measurements. SAXS measurements were performed using beamline BM29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The incident-beam wavelength was 0.99 Å. The distance between the sample and the PILATUS 1 M detector was 2867 mm, giving a scattering vector range of 0.036−4.95 nm^{-1}. The scattering vector is defined as $q = 4\pi \sin \theta/\lambda$, where 2θ is the scattering angle. For each sample and pure solvent, several successive 1 s frames (typically 10−25) were recorded and analyzed. Special attention was paid to radiation damage by comparing the successive frames prior to further processing of the data. The background (pure solvent) was subtracted from the corresponding sample solutions. All measurements were performed at 20 °C. $I(0)$ was converted to an absolute scale by measuring the scattering of water. Before SAXS measurements, the samples were further centrifuged with an ultracentrifuge (TLA55 rotor) at 13 000 rpm for at least 30 min in order to remove unspecific aggregates. Protein concentrations were measured after preparation and again immediately before the SAXS measurements using a Nanodrop spectrophotometer (ε = 1490 M^{-1} cm^{-1}, λ = 280 nm).

The scattering of a macromolecule in solution depends on the contrast ("scattering power" relative to the solvent), given by the square of the difference between the scattering length densities of the solute and solution. The scattering length density differences (determined from the electron densities) of Hst5_{4-15} and Spd are 3.099×10^{10} and 0.564×10^{10} cm^{-2} respectively, in pure water. Furthermore, the scattering length density differences were calculated with MULCh[14] to be 3.013×10^{10} cm^{-2} for Hst5_{4-15} and 1.030×10^{10} cm^{-2} for Spd in 20 mM Tris and 140 mM NaCl. From these numbers we assume that the SAXS spectra give information about the peptide part of the conjugate only, and not the bonded Spd. The above values could, for example, be compared with the scattering length density difference for a 10-monomer poly(ethylene glycol) molecule (i.e., ~1.5×10^{10} cm^{-2}), which is known to give very low contrast. Experimental SAXS spectra of the Hst5_{4-15} fragment and the conjugate were thus compared (see Figure 1). The origin of the small deviation between the scattering curves in the high-q region is most probably due to the background subtraction. For both Hst5_{4-15} and the Hst5_{4-15}−Spd conjugate, the apparent molecular mass, determined from $I(q = 0)$, was found to be within a 10% deviation from the molecular mass calculated from the amino acid sequence.

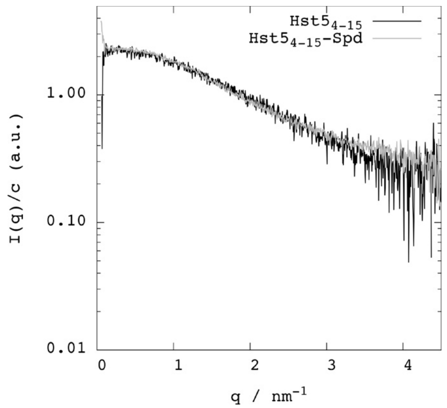


Figure 1. Comparison of the scattering intensities of the Hst5_{4-15} peptide and the Hst5_{4-15}−Spd conjugate. The concentrations were 1.50 and 1.29 mg/mL for conjugate and peptide sample, respectively.

Circular Dichroism. Hst5 (60 μm) and Hst5_{4-15}−Spd conjugates (60 μM) were dissolved in 95% trifluoroethanol (TFE) (Sigma-Aldrich, St. Louis, MO, USA) and analyzed in a fluorimeter cell (Starna Cells, Inc., Atascadero, CA, USA) with a 1 cm path length at 25 °C. Far-UV spectra (190−280 nm) were collected with a circular dichroism (CD) spectrophotometer (model 815, Jasco, Easton, MD, USA) and a thermostatic water bath. The data were processed using Jasco software and analyzed using Prism 5.0.

Nano LC−MS/MS. Mass spectrometry was employed to ensure the purity of the Hst5_{4-15}−Spd samples. The experiments were performed with an EasyLC nanoflow HPLC interfaced with a nanoEasy spray ion source (Proxeon Biosystems, Odense, Denmark) connected to a Fusion Orbitrap mass spectrometer (Thermo Fisher Scientific). The sample was loaded on a 2 cm PepMap column (75 μm inner diameter packed with 3 μm resin), and the chromatographic separation was performed at 35 °C on a 25 cm (75 μm inner diameter) EASY-Spray column packed with 2 μm of resin (Proxeon Biosystems). The nanoHPLC was operated at 300 nL/min with a gradient of 5−22% solvent B (0.1% (v/v) formic acid in acetonitrile) in solvent A (0.1% (v/v) formic acid in water) over 20 min, 22−32% over 2 min, and then an increase to 98% B over 2 min. An MS scan (m/z 350−1500) was recorded in the Orbitrap mass analyzer, which was set at a resolution of 60 000 at m/z 400, 1 × 10 automatic gain control target, and a maximum ion injection time of 500 ms. The mass spectrometric conditions were as follows: spray voltage, 1.9 kV; no sheath or auxiliary gas flow; S-lens, 60%; ion transfer tube temperature, 275 °C. The molecular mass was determined to be 1862.084 Da (theoretical weight 1863.0 Da), and very few ions were detected apart from the conjugate, indicating a purity of >98% (Figures 2 and 3).

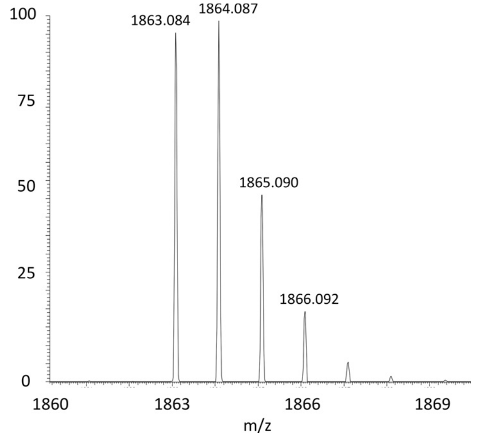


Figure 2. Mass determination of Hst5$_{4-15}$−Spd using a MALDI Orbitrap instrument. The *y* axis displays the relative intensity and the *x* axis the mass range.

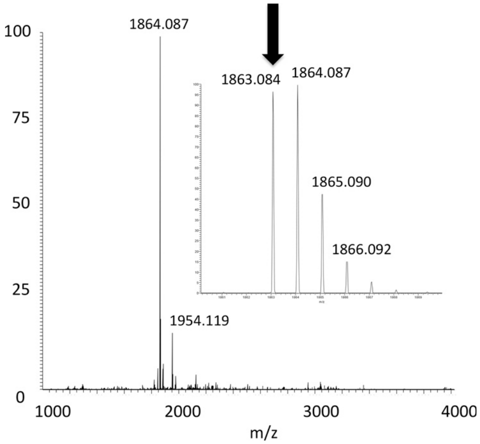


Figure 3. Mass determination of Hst5$_{4-15}$−Spd using a MALDI Orbitrap instrument. The inset shows the high-resolution separation (as shown in Figure 2), which enables the accurate mass determination of the different isotopes. The monoisotopic peak is indicated by the arrow. The *y* axis displays the relative intensity and the *x* axis the mass range.

■ COMPUTATIONAL SECTION

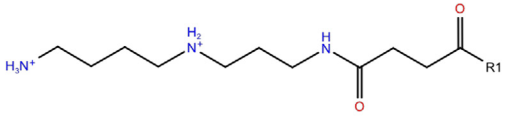
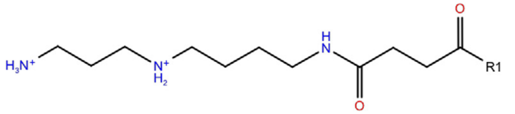
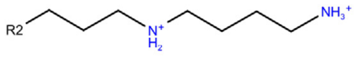
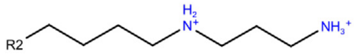
Models. Hst5$_{4-15}$ (Ala$_1$−Lys−Arg−His$_4$−His−Gly−Tyr−Lys$_8$−Arg−Lys−Phe−His$_{12}$) and its Spd conjugates (see Table 1) were built with PyMOL.[15] The initial structures were assumed to be linear to avoid subsequent biasing of the conformational sampling. All four possible isomers were built and simulated. The nomenclature (as shown in the first column of Table 1) is XSpdY, where X is either N or C, depending on whether it is an N- or C-terminal conjugate, and Y is either S or L (which stand for short side and long side, respectively), depending on the extremity by which Spd is bonded.

Charge Parametrization. Spd and succinic acid geometry optimizations (HF/6-31G*) were performed using the Gaussian 09 program.[16] The same program, method, and basis set were also used to derive the electrostatic potentials from which the restrained electrostatic potential (RESP) charges (AmberTools 14 suite[17]) were calculated. A two-stage RESP fit was executed for each model shown in Table 2. In the first step, the charges of all noncapping atoms were allowed to vary. The charges of the capping residues, i.e., acetyl (ACE) or methylamine (NME), were taken from Cornell et al.[18] Hydrogen charge symmetry was enforced for Spd amine groups with more than one hydrogen. In the second step, only the charges of methylene bridge groups were allowed to vary. Charge symmetry was enforced for the respective hydrogen atom charges. All of the other charges were fixed to the output of the first step. The final charges (presented in Table S1) were then used to build the new residue topology entries in the AMBER ff99SB-ILDN force field.[19] The remaining bonded and nonbonded parameters were taken from the aforementioned force field, using the appropriate "atom type" for each atom in the models presented in Table 2.

Molecular Dynamics Simulations. All of the molecular dynamics simulations were performed with the GROMACS package (version 4.6.7)[20−23] using the AMBER ff99SB-ILDN force field[19] and the TIP4P-D water model[24] in the isobaric−isothermal ensemble. All of the simulation boxes were constructed as rhombic dodecahedrons with periodic boundary conditions and a minimum distance of 1 nm between the solute and the box. All of the systems were neutralized with Cl^- ions (five for Hst5$_{4-15}$, six for the different N-terminal Hst5$_{4-15}$−Spd conjugates, and eight for their C-terminal counterparts). Simulations with an additional 140 mM NaCl were also performed for the Hst5$_{4-15}$ fragment and the CSpdS conjugate in order to mimic experimental conditions and to ensure that the addition of salt to the simulations had no effect on the estimation of the structural properties studied in this work (see Figure S3 and Table S2). The equations of motion were numerically integrated using the Verlet leapfrog algorithm with a time step of 2 fs. The nonbonded interactions were treated with a Verlet list cutoff scheme in order to make use of the program's native GPU acceleration. The short-range interactions were calculated using a nonbonded pair list with a single cutoff of 1 nm, updated every 100 fs. Long-range dispersion corrections were applied to the system's energy and pressure. Particle mesh Ewald[25] was used to handle the long-range electrostatic interactions with cubic interpolation and a grid spacing of 0.16 nm. The solvent and solute were separately coupled to temperature baths at 300 K with the velocity rescaling thermostat[26] and a relaxation time of 0.1 ps. A Parrinello−Rahman pressure coupling[27] was used at 1 bar with a relaxation time of 2 ps and an isothermal compressibility of 4.5×10^{-5} bar^{-1}. All of the bond lengths were constrained using the LINCS algorithm.[28]

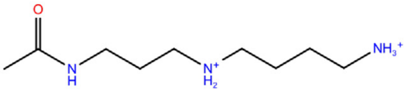
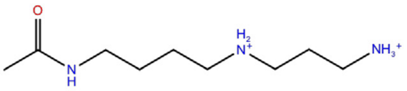
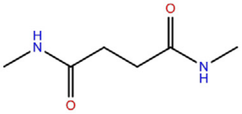
The minimization procedure was performed using the steepest-descent algorithm[29] without setting a maximum number of steps in order to achieve convergence within the available machine precision. Initiations were performed in a two-step scheme of 500 ps and 1 ns using position restraints of 1000 kJ mol^{-1} nm^{-2} on all protein heavy atoms. The first step was performed under the canonical ensemble to stabilize the temperature, and the second step was performed under the isobaric−isothermal ensemble.

Table 1. $Hst5_{4-15}$−Spd Conjugate Models

NSpdS	Spermidine (short side) + succinic acid + R1*	
NSpdL	Spermidine (long side) + succinic acid + R1*	
CSpdS	R2† + spermidine (short side)	
CSpdL	R2† + spermidine (long side)	

* R1 = GGG linker + AKRHHGYKRKFH
† R2 = AKRHHGYKRKFH + GGG linker

Table 2. Models Used for the Charge Derivation of Spd and Succinic Acid

ACE capped spermidine (short side)	
ACE capped spermidine (long side)	
NME capped succinic acid (both sides)	

Each conjugate system was simulated for a total of 5 μs (five replicates of 1 μs each). The protein fragment alone (i.e., $Hst5_{4-15}$) was simulated for twice as long (10 replicates of 1 μs each) because of slower convergence. Coordinates and energies were saved every 10 ps. Residue charges were set to be representative of those at pH 7, with histidines being modeled in the deprotonated form.

The MD trajectories were used in their entirety for all of the analyses because of the inherent difficulty in defining initial equilibration times for such highly dynamic systems.

Analyses. As noticed before by our group when we studied Hst5, the simulation time needed to equilibrate properties such as the radius of gyration (R_g) and the end-to-end distance (R_{ee}) around their average values is extremely short,[30,31] so short in fact that it is negligible given the very long simulation times of the replicates of each system (see Figure S1). Figure S2 shows the individual distributions of the radius of gyration and end-to-end distance per replicate for each simulated system. It is clearly visible that satisfactory convergence was achieved for all of the conjugates. However, some replicates of the system consisting of the protein fragment alone appeared to be "stuck" in metastable, compact conformations at times. Hence the doubled number of simulations performed for this system.

Information about all of the other analyses performed in this paper can be found in the previous studies by Henriques et al.[30] and Henriques and Skepö.[31]

Figures. All of the graphs were created using GNUPLOT.[32] Protein snapshot figures were rendered using the PyMOL ray tracer. Two-dimensional chemical structure representations were drawn with Maestro.[33]

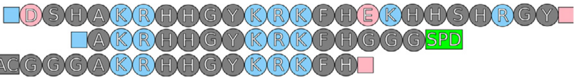


Figure 4. Schematic picture of the charge distributions for Hst5 (top row) and the conjugates with Spd attached either to the C-terminus (middle row) or the N-terminus (bottom row). The color code for the charges (in units of *e*) is as follows: blue, +1; pink, −1; gray, 0; green, +2. The total net charges are +5, +8, and +6, respectively. SPD stands for spermidine, and SAC stands for succinic acid.

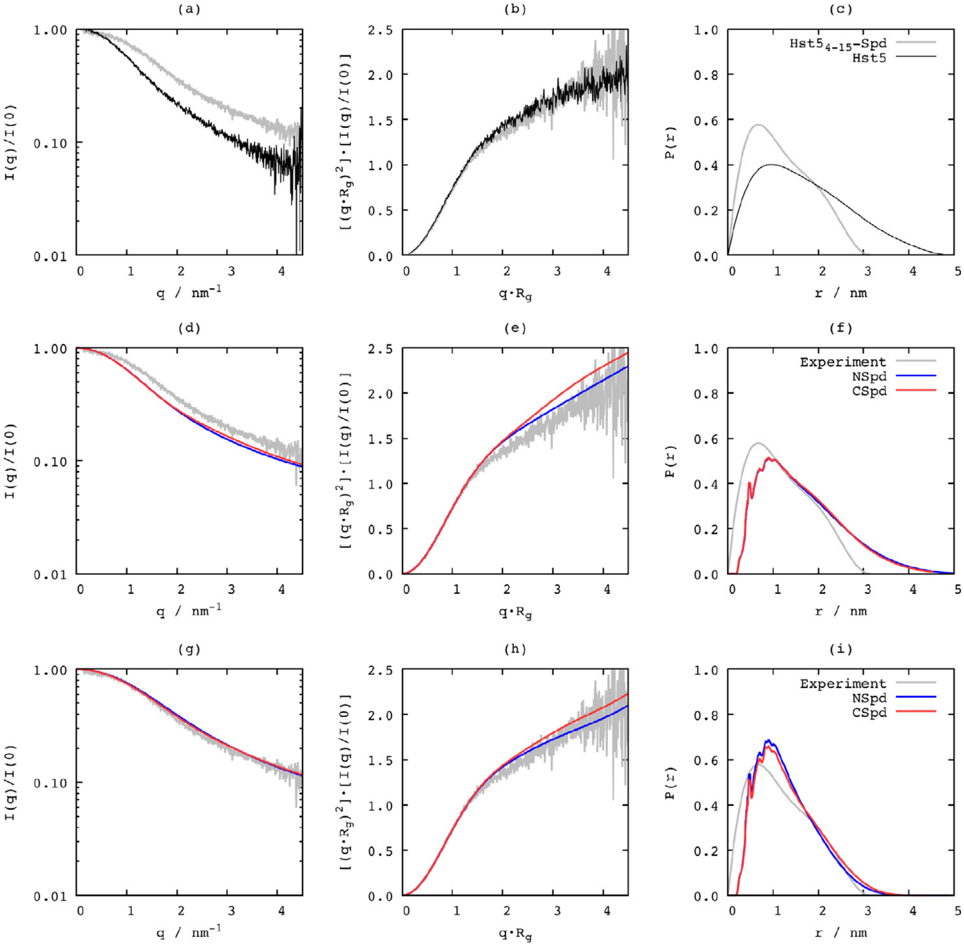


Figure 5. Figure sets depicting (left to right) the form factor, the Kratky plot, and $P(r)$ determined by SAXS measurements and simulations. The upper row (a−c) shows a comparison between the experimental data for Hst5 and the Hst5$_{4-15}$−Spd conjugate. The middle row (d−f) compares the experimental conjugate results to the simulated results for the N- and C-terminal conjugates. The bottom row (g−i) compares the experimental conjugate results to the simulated results for the Hst5$_{4-15}$ part of the conjugates only. All of the Hst5 data used in this figure can be found in previous studies by Cragnell et al.[7]

■ RESULTS AND DISCUSSION

General Introduction. Figure 1a in the paper by Tati et al.[5] shows the killing percentage as a function of the peptide concentration for *C. albicans*. It is shown that the conjugates have a higher killing efficiency than Hst5, with a difference of almost a factor of 2 at the highest evaluated peptide concentration. Only at lower concentrations is there a deviance between the N- and C-terminal conjugates, where the latter have the higher fungicidal activity.

Figure 4 in this paper shows a schematic picture of the primary structures of the peptides with respect to both length and charge.

DOI: 10.1021/acs.jcim.7b00150
J. Chem. Inf. Model. 2017, 57, 1330−1341

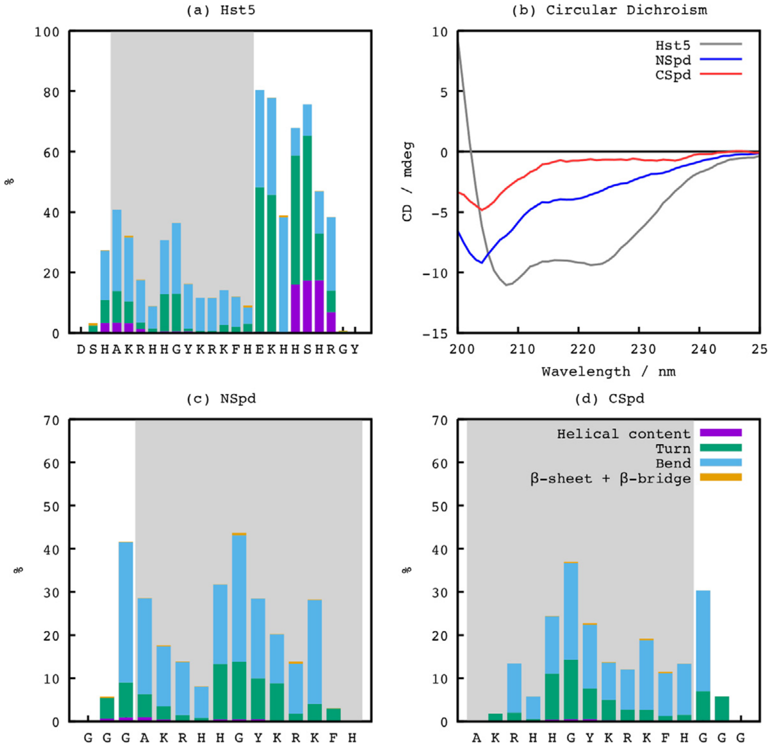


Figure 6. (a, c, d) Stacked secondary structure histograms for the simulated Hst5 and the peptide part of the conjugates. All of the histograms sum up to 100%, but the coil percentages are not shown for visual simplicity. The shaded areas indicate the conjoint amino acid sequence that can be found in all the peptides of this study. (b) CD spectra for Hst5 and the two conjugates in trifluoroethanol.

As can be seen, Hst5 has a longer sequence and contains two negatively charged residues (excluding the C-terminal carboxyl group), whereas the conjugates contain solely positively charged and neutral amino acids. Hence, the average charges per amino acid are 0.21*e* for Hst5 and 0.27*e* and 0.40*e* for the N- and C-terminal conjugates respectively, i.e., the conjugates have a higher charge than the native peptide.

As explained in the Computational Section, Spd can be attached at either the N- or the C-terminus, and at each end there are two alternatives, either short or long (see Table 1). Since it is not possible to elucidate whether the molecular structures of the conjugates from the supplier are long or short, comparisons of simulation results with experiments are done using a mixture of the two. In the second part of this paper, a computational analysis of the conjugates will be performed, where the individual molecular structures will be taken into consideration.

Histatin 5 versus the Conjugates. To be able to compare the SAXS measurements with the molecular dynamics simulations, samples with a low peptide concentration and an ionic strength of 140 mM were used. Under these conditions, the peptide conformations and locations are uncorrelated, and the solution scattering effectively gives the form factor. Four CSpd concentrations were measured to ensure concentration independence of the radius of gyration and the scattering function (see Figure S4). A CSpd concentration of 2.4 mg mL^{-1} was used as the form factor to give the appropriate signal-to-noise ratio. The Guinier approximation, restricted to $qR_g < 0.8$, and the pair distance distribution function $P(r)$ (GNOM[34]) gave consistent values of $I(q = 0)$ (0.00193 and 0.00196 cm^{-1}, respectively) and R_g (0.92 ± 0.05 and 0.94 ± 0.009 nm, respectively). The apparent molecular mass, which provides an indication of monodispersity, was determined from the specific volume of CSpd (ν_p = 0.694−0.715 $\mathrm{cm}^3\ \mathrm{g}^{-1}$; obtained using Sednterp[35]) and $I(q = 0)$ to be within the range 1.747−2.066 kDa. This is in good agreement with the molecular mass calculated from the amino acid sequence (1.863 kDa), indicating that CSpd is indeed monomeric.

Figure 5 shows comparisons of the form factors, Kratky plots, and distance distribution functions, $P(r)$, for the experimental data for Hst5 and CSpd (upper panel), the experimental CSpd data and the simulated CSpd and NSpd results (middle panel), and the experimental CSpd data and the simulated results for the $\mathrm{Hst5}_{4-15}$ fragments in NSpd and CSpd (bottom panel). The Hst5 form factor used in this figure can be found in the previous study by Cragnell et al.[7] The last two panels show results for both SAXS and molecular dynamics simulations, whereas the first is for experiments solely. First of all, it is clearly shown that the simulations and the experiments are in reasonable agreement for the $\mathrm{Hst5}_{4-15}$ fragment part of NSpd and CSpd with the SAXS scattering of the conjugate (bottom panel), i.e., the applied force

Table 3. Averages and Standard Deviations of the Radius of Gyration (R_g), the End-to-End Distance (R_{ee}), and the Persistence Length (L_p) for Each Simulated System; The Upper Values in Each Row Represent the Complete Conjugate, Whereas the Lower Values Represent the Fragment Part of the Conjugate Only

	Hst5$_{4-15}$	NSpdS	NSpdL	CSpdS	CSpdL
$\langle R_g \rangle \pm \sigma$ (nm)[a]	0.97 ± 0.11	1.23 ± 0.17 0.97 ± 0.11	1.25 ± 0.17 0.97 ± 0.12	1.22 ± 0.12 1.01 ± 0.09	1.20 ± 0.14 1.00 ± 0.10
$\langle R_g^{b} \rangle \pm \sigma^{b}$	0.76 ± 0.09	0.64 ± 0.09 0.71 ± 0.08	0.64 ± 0.08 0.71 ± 0.08	0.70 ± 0.07 0.75 ± 0.07	0.69 ± 0.08 0.74 ± 0.08
$\langle R_{ee} \rangle \pm \sigma$ (nm)	2.49 ± 0.65	3.36 ± 1.01 2.40 ± 0.69	3.47 ± 0.97 2.47 ± 0.64	3.69 ± 0.79 2.72 ± 0.51	3.62 ± 0.83 2.67 ± 0.54
$\langle R_{ee}^{c} \rangle \pm \sigma^{c}$	0.64 ± 0.17	0.50 ± 0.15 0.60 ± 0.17	0.51 ± 0.14 0.62 ± 0.16	0.59 ± 0.13 0.69 ± 0.13	0.58 ± 0.13 0.68 ± 0.14
$\langle L_p \rangle \pm \sigma$ (bonds)	7.05 ± 3.91	5.68 ± 2.87 7.02 ± 3.84	5.72 ± 2.94 7.01 ± 3.91	7.18 ± 3.71 7.99 ± 3.99	7.15 ± 3.79 7.93 ± 4.09

[a] Experimental conjugate R_g = 0.94 ± 0.009 nm (the latter is the error and not the standard deviation). [b] Normalized by the maximum radius of gyration. [c] Normalized by the approximate contour length.

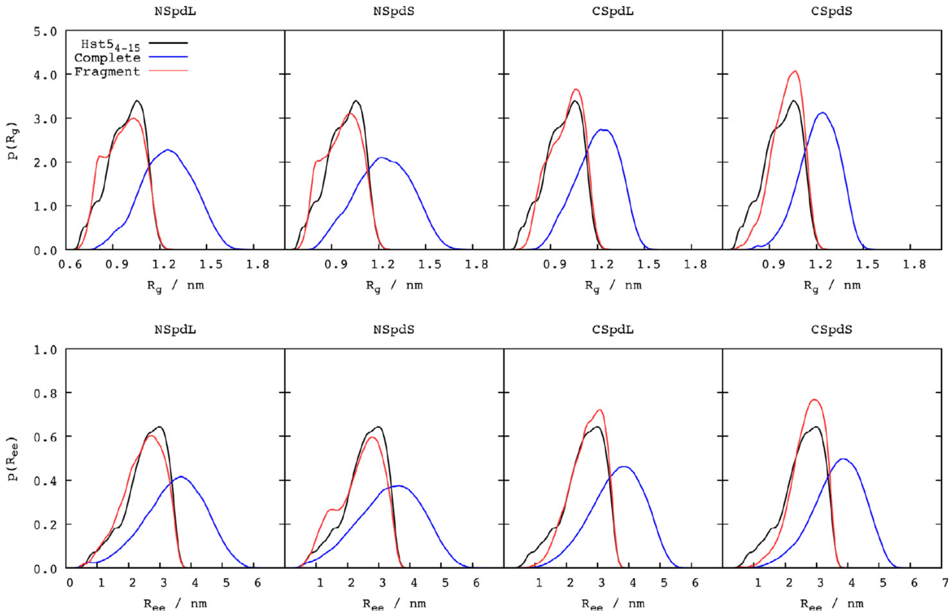


Figure 7. Density estimates of the radius of gyration (upper row) and the end-to-end distance (lower row) for all of the simulated conjugates. The blue lines show the distributions for the complete conjugates, whereas the red lines show the distributions for the peptide part of the conjugates only (excluding the triple-glycine linkers). The result for each conjugate is compared to that for the simulated Hst5$_{4-15}$ fragment, which is shown in black.

fields, the water models, and the parametrization of Spd are reliable and give accurate results. The assumption stated in Materials and Methods that the SAXS curve gives information mainly about the peptide part of the conjugate also seems to be valid. Taking the whole conjugate (with the Spd part included) into account when calculating the scattering function and $P(r)$ gives rise to too-large average dimensions. From the distance distribution functions it can be seen that, as expected, Hst5 is more extended than the fragments (D_{max} = 5 nm vs 3 nm) independent of whether Spd is attached to the N-terminus or the C-terminus. The Kratky plot indicates that Hst5 attains a random coil conformation, which is expected since Hst5 has been shown to behave as such.[36,37] On the other hand, as a result of their small size in combination with the higher charge density, the fragments become slightly stiffer. Hence, from a conformational point of view, the conjugates display similar trends, differing from that of Hst5, as shown in Figure 5 (upper panel). The radius of gyration of the conjugates obtained from SAXS is approximately 0.94 nm (as obtained using the pair distance distribution function $P(r)$[34]), whereas the average radius of gyration obtained from the simulations ranges from 0.97 to 1.01 nm.

Figure 6 shows the secondary structure predictions obtained from the molecular dynamics simulations as well as the CD spectra for Hst5 and the conjugates. It should be noted that the CD measurements were performed in TFE. From the experimental point of view, it is well-known that Hst5 does not possess any secondary structure in aqueous solutions, which has been shown both using NMR and CD.[36,37] This is in line with the MD predictions given in Figure 6a, although from a secondary structure analysis at the atomistic level it becomes clear that some part of the amino acid sequence possesses helical content, which can be attributed to the sequence of His-Ser-His at positions 19−21. Despite this, the amino acid sequence consists mainly of turns and bends. On the other hand, when Hst5 is dissolved in an organic solvent such as TFE, its secondary structure content is increased, adopting an α-helical structure, as shown in the CD spectra. The apparent bands at 208 and 222 nm are indicative of an α-helical structure. This is in line with previous studies of Hst5 in TFE.[36,38] In regard to the conjugates, which lack the His-Ser-His sequence, MD simulations indicate only turns and bends for CSpd. The α-helical content for NSpd is negligible. In TFE, the conjugates still do not form any secondary structure, as shown by the absence of spectral features in the respective CD spectra (Figure 6b). Thus, the conjugates are completely disordered in both water and TFE, and it is plausible that the His-Ser-His sequence contributes to forming the α-helical structure. The helix-forming sequence in Hst5 is also a part of a Zn-binding motif (HEXXH motif).[39] A loss of this α-helical secondary structure upon binding of Fe and Zn has been reported.[40]

Computational Analysis of the Conjugates. Table 3 presents the simple and normalized average values of the radius of gyration (R_g), the end-to-end distance (R_{ee}), and the persistence length (L_p) for each simulated system. Two different sets of numbers are reported for the different conjugates. This is the case because we are interested in studying the conjugates both in part (i.e., the Hst5$_{4-15}$ fragment) and as a unit (i.e., including Spd). Figure 7 displays the distributions from which these values were computed, and Figure 8 depicts the representative structure of each simulated system. From an analysis of the differences between the conjugate isomers (i.e., short vs long), it is noticeable that each pair of isomers shows very consistent results regardless of the end by which the Spd molecule is attached. Despite its asymmetry, attaching Spd at either end shifts the position of the central secondary amine by only one bond (recall Table 1).

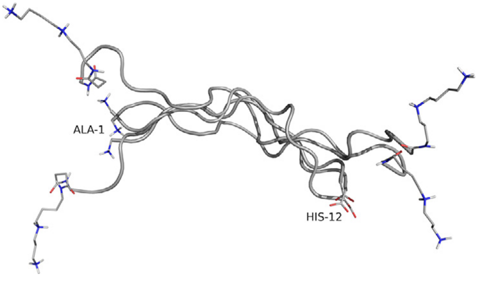


Figure 8. Aligned representative structures of the simulated systems.

From an experimental point of view, it is more interesting to analyze the differences depending on whether Spd is attached to the N- or C-terminus of the peptide. The N-terminal versions of the Hst5$_{4-15}$−Spd conjugate (NSpd) contain a short succinic acid linker (needed for covalent bonding of Spd and the GGG linker), which necessarily increases their maximum extension and degrees of freedom. Despite their longer contour length, the N-terminal conjugates appear to adopt more compact conformations than their C-terminal counterparts, as is visible from the normalized average radii of gyration and end-to-end distances presented in Table 3. These findings are also reflected by their average persistence lengths, which are significantly shorter than those for the C-terminal variants, indicating that the former are more flexible. One possible explanation for these observations is that the CSpd conjugates have a higher linear charge density (cf., net charge of +8e vs +6e) and shorter contour length, which necessarily results in increased intramolecular electrostatic repulsion in comparison with the NSpd counterparts.

One could hypothesize that there might be a correlation between the results from the microbiological studies performed by Tati et al.[5] and the less flexible and more elongated average structure of the CSpd conjugates, since CSpd has experimentally been shown to have slightly superior killing activity compared with its N-terminal counterpart. It seems plausible that a more elongated and rigid structure should better expose Spd to the solvent, facilitating its recognition by the polyamine transporters.

As expected, the system consisting of the Hst5$_{4-15}$ fragment alone presents a smaller radius of gyration and end-to-end distance than the different conjugates. However, when normalized, these values clearly indicate that for its size, this system adopts more expanded conformations than its conjugate counterparts. Interestingly, this peptide is also significantly more extended than its parent protein, Hst5, for which the normalized $\langle R_g \rangle$ and $\langle R_{ee} \rangle$ are 0.56 ± 0.09 and 0.35 ± 0.14 nm, respectively, under the same simulation conditions, force field, and water model.[31] The latter is likely due to the stronger intramolecular electrostatic repulsion arising from the higher net charge per residue of the peptide compared with full-length Hst5.

When the focus is on the fragment part only, the most pronounced differences between the simulations of the different conjugates and Hst5$_{4-15}$ alone become almost negligible, as if the fragment portion of the conjugates is oblivious to the other constituents that make up the whole conjugate molecule, behaving just as it would by itself in solution. This is particularly well illustrated by the significant overlap between the black and red curves in Figure 7, and one interesting question is whether conjugation with Spd affects the conformational space of Hst5$_{4-15}$ at all. For that purpose, a principal component analysis (PCA) of the peptide backbone for all five systems was performed, as devised by Campos and Baptista.[41] In general, we found that the first two principal components (PCs) comprise the majority of the variance in each system. Furthermore, visual inspection and root-mean-square calculations (not shown) prove that corresponding minima in the free energy landscapes enclose closely related conformations and that different minima relate to distinct conformational classes, whose dissimilarity increases with their distance in the landscape. This finding validates the analysis in itself and shows that higher dimensionality (i.e., a higher number of PCs) is not necessary for this specific purpose. Figure 9 shows that all of the systems sample very similar low-energy conformational regions, which means that on average, regardless of whether there is conjugation with Spd or not, the 12 amino acid active Hst5 segment adopts similar conformations. Hence, conjugation with Spd does not noticeably affect the conformational space of Hst5$_{4-15}$. This is also visible from the aligned representative structures in Figure 10. It should be noted, though, that while most of the sampled conformations occupy the upper right quadrant of each plot, some simulations venture all the way down to the lower left quadrant, on which compact, β-hairpin-like structures are aggregated (see Figure 11).

DOI: 10.1021/acs.jcim.7b00150
J. Chem. Inf. Model. 2017, 57, 1330−1341

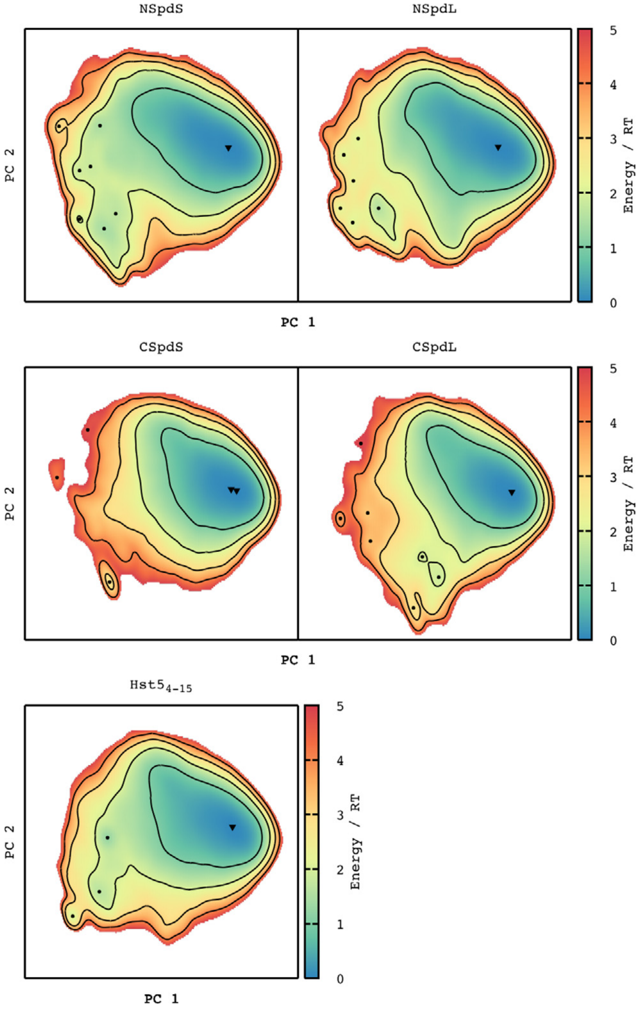


Figure 9. Free energy landscapes for the $Hst5_{4-15}$ part of each simulated system using the first two principal components. Contour lines are drawn for energies between 0 and 5RT with 1RT increments. Triangles mark the centers of the lowest-energy minima, and dots represent other local minima. Plots in the same row have axes of equal size.

The computational results illustrated in Figures 7−10 are in good agreement with the experimental SAXS results (see Figure 1), as neither computations nor experiments show any significant conformational change in the $Hst5_{4-15}$ fragment when it is conjugated to Spd compared to the fragment alone.

■ CONCLUSION

The first part of this work investigated the conformational and structural properties of the $Hst5_{4-15}$−Spd conjugates in comparison to Hst5. Reasonable agreement between experimental and simulated SAXS results were shown, suggesting that the applied force field, water model, and Spd parametrization used for the simulations are reliable. It is known from both SAXS measurements and simulations that Hst5 adopts a random coil conformation in aqueous solution.[7,30,31] The $Hst5_{4-15}$ fragment part of the conjugates was shown to behave similarly to Hst5, displaying slightly more rigid conformations, which is reasonable in view of the combination of shorter contour lengths and higher charge density. The remaining simulated SAXS curves were more or less identical for the two conjugates. When the secondary structure as obtained from the simulations was studied, it was observed that the conjugates do not possess any significant secondary structure in aqueous solution. This is in line with the previous results reported for Hst5, although for the latter there are some indications of helical content for the His-Ser-His (19−21) sequence. The CD spectra, however, showed that the

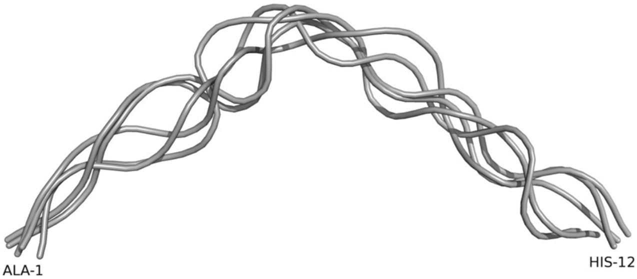


Figure 10. Same as Figure 8, but here only the $Hst5_{4-15}$ section of the conjugates was considered for the calculation of the representative structures.

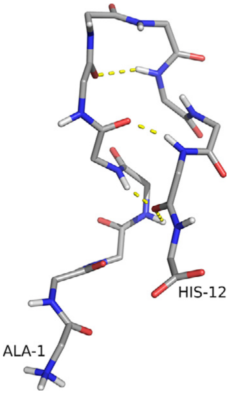


Figure 11. Representative structure of a local free energy minimum from the bottom left quadrants of Figure 9. Hydrogen bonds within the main chain are represented as dashed yellow lines.

conjugates seemed to retain their lack of secondary structure in the organic solvent. It is hypothesized that this might be due to the absence of the His-Ser-His sequence in the conjugates.

In the second part of this work, the different types of conjugates were compared on the basis of simulation results solely. It was observed that despite having a longer contour length, the NSpd conjugate still adopts a more compact conformation compared with the CSpd conjugate. This effect could arise for two reasons: (1) the intramolecular electrostatic repulsion is higher for CSpd, as it has a higher charge density, and/or (2) the succinic acid linker in NSpd gives rise to more flexibility in the molecule, which might allow the Spd to fold back toward the $Hst5_{4-15}$ fragment, hence reducing the radius of gyration of the entire conjugate. The simulations also showed that conjugation with Spd does not appear to affect the conformational ensemble of the $Hst5_{4-15}$ fragment. This was also observed in the SAXS measurements from a comparison of the form factors, assuming that the Spd part did not contribute to the scattering.

We previously observed that structural changes induced by introduction of Spd improved fungicidal abilities of both Hst5 conjugates, which were attributable to resistance of the conjugates to salivary proteases and to enhanced uptake into fungal cells. We hypothesize that the $Hst5_{4-15}$−Spd conjugate has better activity compared with the Spd−$Hst5_{4-15}$ conjugate since it assumes a structure that better conceals residues that are enzymatic cleavage sites and protects them from degradation while exposing those residues involved in cell surface and *C. albicans* Dur3 transporter binding. Indeed, although both conjugates remain unstructured, these results show differences in the rigidity and compactness of the two conjugates. The more rigid random coil conformation of $Hst5_{4-15}$−Spd may result in residues that are stabilized to allow better occlusion, translocation, or release of the substrate to the inner side of the Dur transporter channel.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.7b00150.

> Information regarding the charge parametrization, time series of the radius of gyration and end-to-end distances and the respective density estimates (per simulation replicate), results for supplementary simulations performed with 140 mM salt, and additional experimental SAXS scattering intensities at different conjugate concentrations as well as the form factors on a log−log scale (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: marie.skepo@teokem.lu.se. Phone: +46 462 223 366.

ORCID

Stephanie Jephthah: 0000-0003-4287-506X
Carolina Cragnell: 0000-0001-6751-0644
Sumant Puri: 0000-0002-5509-007X
Mira Edgerton: 0000-0002-9680-5908
Marie Skepö: 0000-0002-8639-9993

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge financial support from OMM Linnaeus Center, the Swedish Research Council, the Swedish Foundation for Laryngectomees, and the U.S. National Institutes of Health (Grant NIH R01DE010641 to M.E.). The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC). To finalize, we thank the European Synchrotron Radiation Facility

(Grenoble, France) for providing beamtime and Dr. Petra Pernot for assistance in using beamline BM29.

■ REFERENCES

(1) Enoch, D. A.; Ludlam, H. A.; Brown, N. M. Invasive Fungal Infections: A Review of Epidemiology and Management Options. *J. Med. Microbiol.* **2006**, *55*, 809−818.

(2) Eggimann, P.; Garbino, J.; Pittet, D. Epidemiology of Candida Species Infections in Critically Ill Non-immunosuppressed Patients. *Lancet Infect. Dis.* **2003**, *3*, 685−702.

(3) Pappas, P. G.; Rex, J. H.; Lee, J.; Hamill, R. J.; Larsen, R. A.; Powderly, W.; Kauffman, C. A.; Hyslop, N.; Mangino, J. E.; Chapman, S.; et al. A Prospective Observational Study of Candidemia: Epidemiology, Therapy, and Influences on Mortality in Hospitalized Adult and Pediatric Patients. *Clin. Infect. Dis.* **2003**, *37*, 634−643.

(4) Eggimann, P.; Garbino, J.; Pittet, D. Management of Candida Species Infections in Critically Ill Patients. *Lancet Infect. Dis.* **2003**, *3*, 772−785.

(5) Tati, S.; Li, R.; Puri, S.; Kumar, R.; Davidow, P.; Edgerton, M. Histatin 5-Spermidine Conjugates Have Enhanced Fungicidal Activity and Efficacy as a Topical Therapeutic for Oral Candidiasis. *Antimicrob. Agents Chemother.* **2014**, *58*, 756−766.

(6) Danesi, R.; Senesi, S.; Wout, J. W. v. t.; Dissel, J. T. v.; Lupetti, A.; Nibbering, P. H. Antimicrobial Peptides: Therapeutic Potential for the Treatment of Candida Infections. *Expert Opin. Invest. Drugs* **2002**, *11*, 309−318.

(7) Cragnell, C.; Durand, D.; Cabane, B.; Skepö, M. Coarse-grained Modelling of the Intrinsically Disordered Protein Histatin 5 in Solution: Monte Carlo Simulations in Combination with SAXS. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 777−791.

(8) Oppenheim, F. G.; Xu, T.; McMillian, F. M.; Levitz, S. M.; Diamond, R. D.; Offner, G. D.; Troxler, R. F. Histatins, a Novel Family of Histidine-rich Proteins in Human Parotid Secretion. Isolation, Characterization, Primary Structure, and Fungistatic Effects on Candida albicans. *J. Biol. Chem.* **1988**, *263*, 7472−7477.

(9) Raj, P. A.; Edgerton, M.; Levine, M. J. Salivary Histatin 5: Dependence of Sequence, Chain Length, and Helical Conformation for Candidacidal Activity. *J. Biol. Chem.* **1990**, *265*, 3898−3905.

(10) Tsai, H.; Raj, P. A.; Bobek, L. A. Candidacidal Activity of Recombinant Human Salivary Histatin-5 and Variants. *Infect. Immun.* **1996**, *64*, 5000−5007.

(11) Tsai, H.; Bobek, L. Human Salivary Histatins: Promising Anti-Fungal Therapeutic Agents. *Crit. Rev. Oral Biol. Med.* **1998**, *9*, 480−497.

(12) Rothstein, D. M.; Spacciapoli, P.; Tran, L. T.; Xu, T.; Roberts, F. D.; Dalla Serra, M.; Buxton, D. K.; Oppenheim, F. G.; Friden, P. Anticandida Activity is Retained in P-113, a 12-Amino-acid Fragment of Histatin 5. *Antimicrob. Agents Chemother.* **2001**, *45*, 1367−1373.

(13) Kumar, R.; Chadha, S.; Saraswat, D.; Bajwa, J. S.; Li, R. A.; Conti, H. R.; Edgerton, M. Histatin 5 Uptake by Candida albicans Utilizes Polyamine Transporters Dur3 and Dur31 Proteins. *J. Biol. Chem.* **2011**, *286*, 43748−43758.

(14) Whitten, A. E.; Cai, S.; Trewhella, J. *MULCh*: Modules for the Analysis of Small-angle Neutron Contrast Variation Data from Biomolecular Assemblies. *J. Appl. Crystallogr.* **2008**, *41*, 222−226.

(15) *The PyMOL Molecular Graphics System*, version 1.2r1; Schrödinger, LLC: New York, 2009.

(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.

(17) Case, D. A.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T. E., III; Darden, T.; Duke, R. E.; Gohlke, H.; Götz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *Amber 14*; University of California: San Francisco, 2014.

(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(19) 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., Bioinf.* **2010**, *78*, 1950−1958.

(20) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(21) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(22) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(23) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(24) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.

(25) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(26) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(27) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys. (Melville, NY, U. S.)* **1981**, *52*, 7182.

(28) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: a Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(29) Fletcher, R.; Powell, M. A Rapidly Convergent Descent Method for Minimization. *Comput. J.* **1963**, *6*, 163−168.

(30) Henriques, J.; Cragnell, C.; Skepö, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 3420−3431.

(31) Henriques, J.; Skepö, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: on the Accuracy of the TIP4P-D Water Model and the Representativeness of Protein Disorder Models. *J. Chem. Theory Comput.* **2016**, *12*, 3407−3415.

(32) Williams, T., Kelley, C.; et al. *GNUPLOT: An Interactive Plotting Program*, version 5.0, patch level 3, 2016.

(33) *Schrödinger Release 2015-4: Maestro*; Schrödinger, LLC: New York, 2015.

(34) Svergun, D. I. Determination of the Regularization Parameter in Indirect-transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495−503.

(35) Hurton, T.; Wright, A.; Deubler, G.; Bashir, B. SEDNTERP.

(36) Brewer, D.; Hunter, H.; Lajoie, G. NMR Studies of the Antimicrobial Salivary Peptides Histatin 3 and Histatin 5 in Aqueous and Nonaqueous Solutions. *Biochem. Cell Biol.* **1998**, *76*, 247−256.

DOI: 10.1021/acs.jcim.7b00150
J. Chem. Inf. Model. 2017, 57, 1330−1341

(37) Raj, P. A.; Marcus, E.; Sukumaran, D. K. Structure of Human Salivary Histatin 5 in Aqueous and Nonaqueous Solutions. *Biopolymers* **1998**, *45*, 51−67.

(38) Melino, S.; Rufini, S.; Sette, M.; Morero, R.; Grottesi, A.; Paci, M.; Petruzzelli, R. Zn^{2+} Ions Selectively Induce Antimicrobial Salivary Peptide Histatin-5 to Fuse Negatively Charged Vesicles. Identification and Characterization of a Zinc-binding Motif Present in the Functional Domain. *Biochemistry* **1999**, *38*, 9626−9633.

(39) Grogan, J.; McKnight, C. J.; Troxler, R. F.; Oppenheim, F. G. Zinc and Copper Bind to Unique Sites of Histatin 5. *FEBS Lett.* **2001**, *491*, 76−80.

(40) Puri, S.; Li, R.; Ruszaj, D.; Tati, S.; Edgerton, M. Iron Binding Modulates Candidacidal Properties of Salivary Histatin 5. *J. Dent. Res.* **2015**, *94*, 201−208.

(41) Campos, S. R.; Baptista, A. M. Conformational Analysis in a Multidimensional Energy Landscape: Study of an Arginylglutamate Repeat. *J. Phys. Chem. B* **2009**, *113*, 15989−16001.

Structural characterization of Histatin 5-Spermidine conjugates: A combined experimental and theoretical study

S. Jephthah,[†] J. Henriques,[†] C. Cragnell,[†] S. Puri,[‡] M. Edgerton,[‡] and M. Skepö*[,†]

†*Division of Theoretical Chemistry, Lund University, Post Office Box 124, S–221 00 Lund, Sweden*

‡*Department of Oral Biology, University at Buffalo, Buffalo, NY, 14214–3092, USA*

E-mail: marie.skepo@teokem.lu.se
Phone: +46 462 223 366

Supporting information

This document contains information regarding the charge parameterization (Table S1), the time series of the radius of gyration and end-to-end distances (Figure S1) and respective density estimates (per simulation replicate; Figure S2). The negligible effect of adding salt to the simulations is illustrated in Figure S3, and the respective average values are shown in Table S2. This supplement also contains additional experimental SAXS scattering intensities at different conjugate concentrations (Figure S4) and a comparison between the Hst5_{4-15} peptide and the Hst5_{4-15}-Spd conjugate (Figure S5).

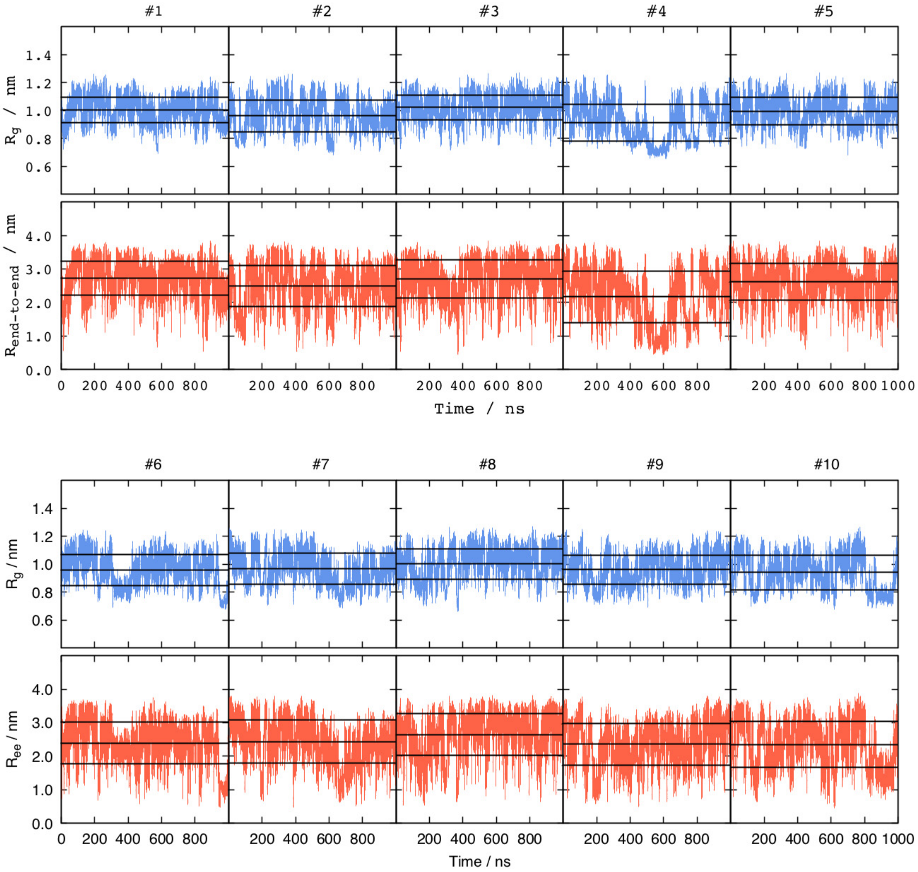
Simulation data

Charges

Table S1: List of partial charges derived for succinic acid (Sac) and spermidine (Spd). The latter is either bound by its "short" (S) or "long" (L) side. The respective AMBER force field atom types are also shown.

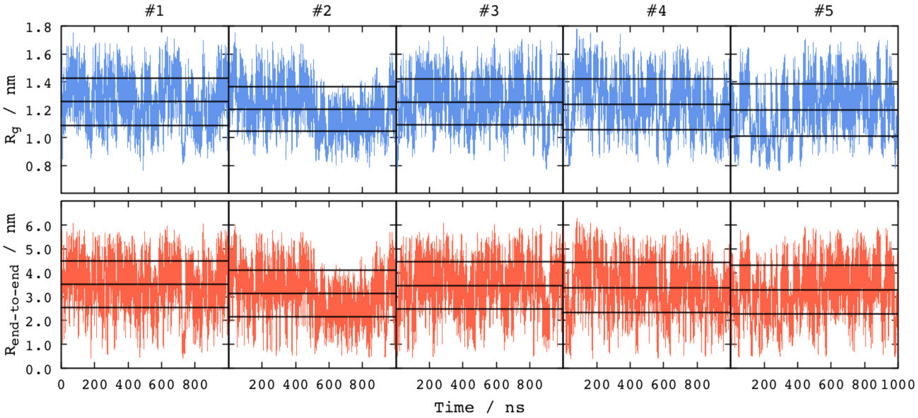
	Sac		SpdS		SpdL	
1	C	0.486928	N3	-0.329914	N3	-0.336112
2	O	-0.538039	H	0.333143	H	0.337842
3	CT	0.051854	H	0.333143	H	0.337842
4	HC	-0.000371	H	0.333143	H	0.337842
5	HC	-0.000371	CT	0.014709	CT	0.042055
6	CT	0.051854	HP	0.114991	HP	0.116958
7	HC	-0.000371	HP	0.114991	HP	0.116958
8	HC	-0.000371	CT	-0.025159	CT	-0.032479
9	C	0.486928	HC	0.057794	HC	0.056211
10	O	-0.538039	HC	0.057794	HC	0.056211
11			CT	-0.012624	CT	-0.026214
12			HC	0.035554	HP	0.094414
13			HC	0.035554	HP	0.094414
14			CT	-0.006382	N3	0.003440
15			HP	0.080819	H	0.246245
16			HP	0.080819	H	0.246245
17			N3	-0.040456	CT	-0.046316
18			H	0.256968	HP	0.096520
19			H	0.256968	HP	0.096520
20			CT	0.026892	CT	-0.009019
21			HP	0.087629	HC	0.032434
22			HP	0.087629	HC	0.032434
23			CT	-0.003188	CT	0.023243
24			HC	0.025250	HC	0.033759
25			HC	0.025250	HC	0.033759
26			CT	-0.133830	CT	-0.124888
27			H1	0.097623	H1	0.092932
28			H1	0.097623	H1	0.092932
29			N	-0.232103	N	-0.286373
30			H	0.229369	H	0.240192

Time series

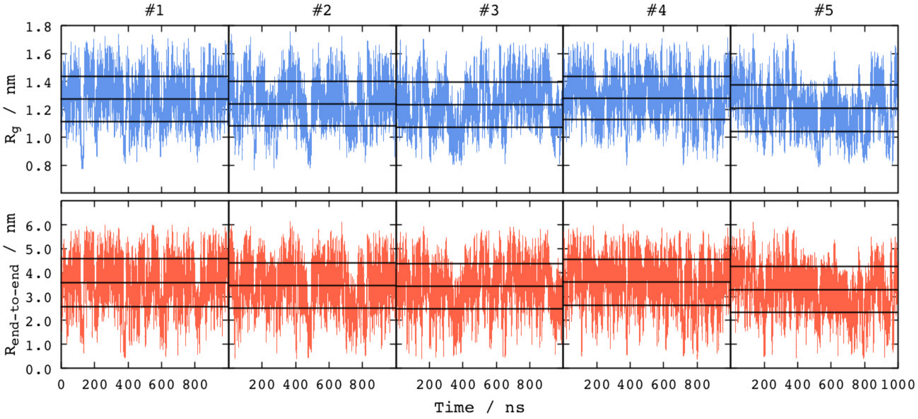


(a) Hst5 $_{4-15}$

Figure S1: Time evolution of the radius of gyration (R_g) and end-to-end distance ($R_{end-to-end}$) for all five replicates of each simulated system. Horizontal lines represent the average value ± standard deviation.

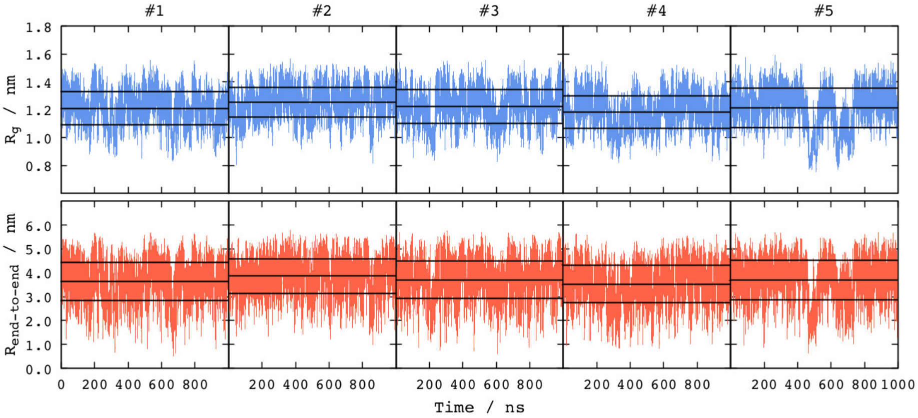


(b) NSpdS

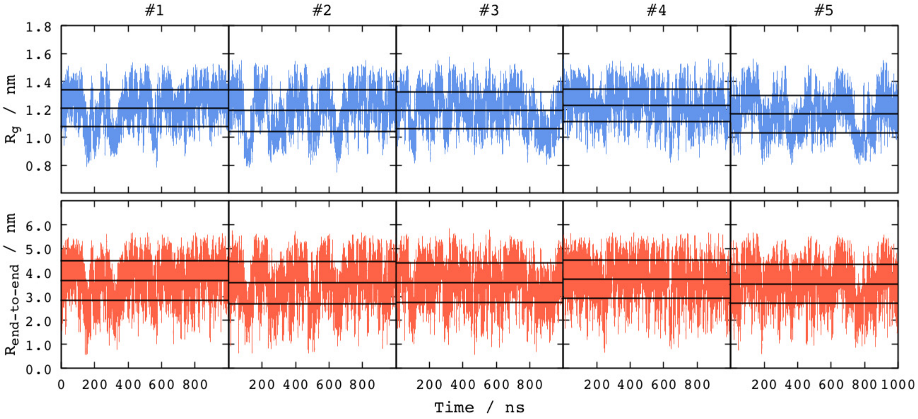


(c) NSpdL

Figure S1: Time evolution of the radius of gyration (R_g) and end-to-end distance ($R_{end-to-end}$) for all five replicates of each simulated system. Horizontal lines represent the average value ± standard deviation.



(d) CSpdS



(e) CSpdL

Figure S1: Time evolution of the radius of gyration (R_g) and end-to-end distance ($R_{end-to-end}$) for all five replicates of each simulated system. Horizontal lines represent the average value ± standard deviation.

Individual distributions

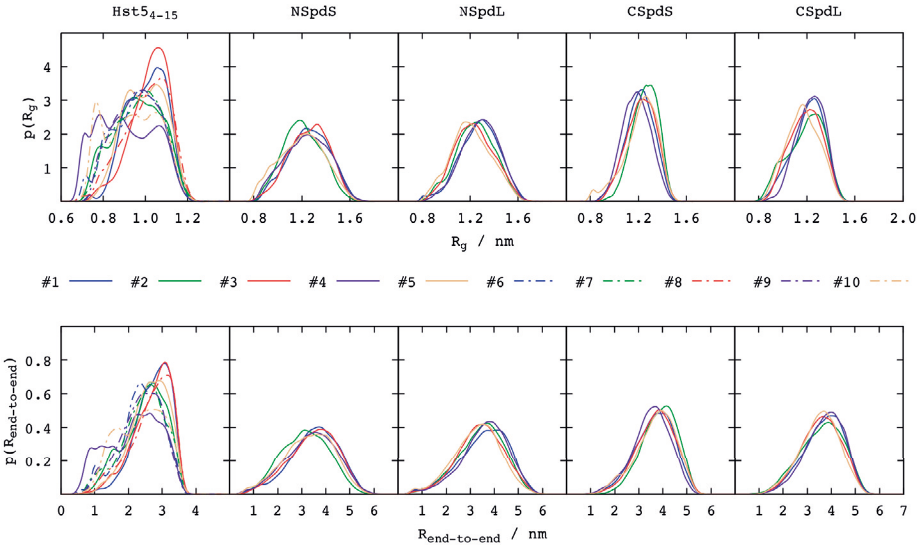
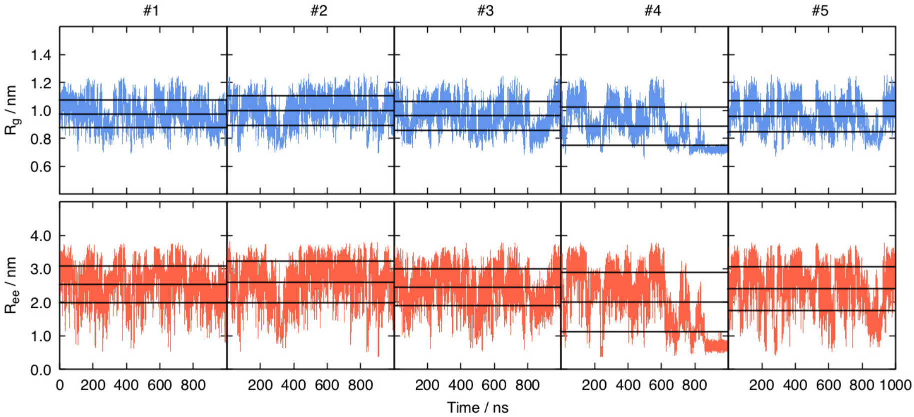
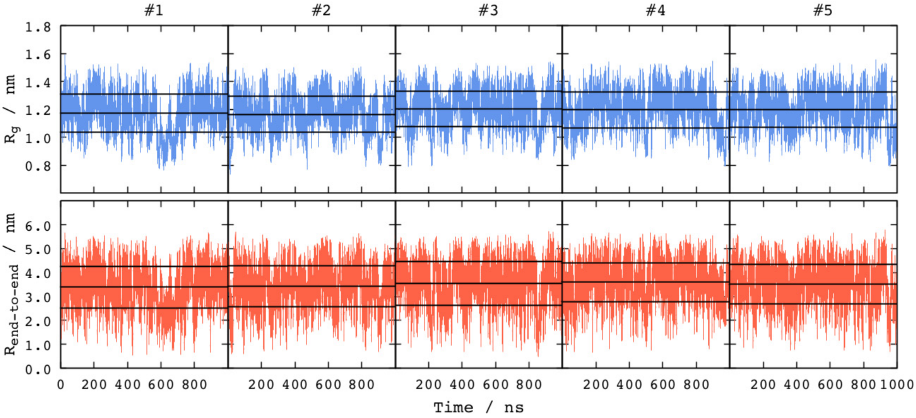


Figure S2: Density estimates of the radius of gyration (R_g) and end-to-end distance ($R_{end-to-end}$) for all replicates of each simulated system.

Time series (140 mM **NaCl**)



(a) Hst5 $_{4-15}$



(b) CSpdS

Figure S3: Time evolution of the radius of gyration (R_g) and end-to-end distance ($R_{end-to-end}$) for Hst5 $_{4-15}$ and CSpdS simulated with 140 mM NaCl. Horizontal lines represent the average value ± standard deviation.

Statistics (140 mM NaCl)

Table S2: Average and standard deviation of the radius of gyration (R_g), end-to-end distance (R_{ee}) and persistence length (L_p), for Hst5 $_{4-15}$ and CSpdS simulated with 140 mM NaCl.

	Hst5 $_{4-15}$	CSpdS
$< R_g > \pm \sigma$ (nm)	0.96 ± 0.12	1.19 ± 0.13
$< R_{ee} > \pm \sigma$ (nm)	2.40 ± 0.70	3.50 ± 0.86
$< L_p > \pm \sigma$ (bonds)	7.17 ± 3.75	6.93 ± 3.40

Experimental data

SAXS

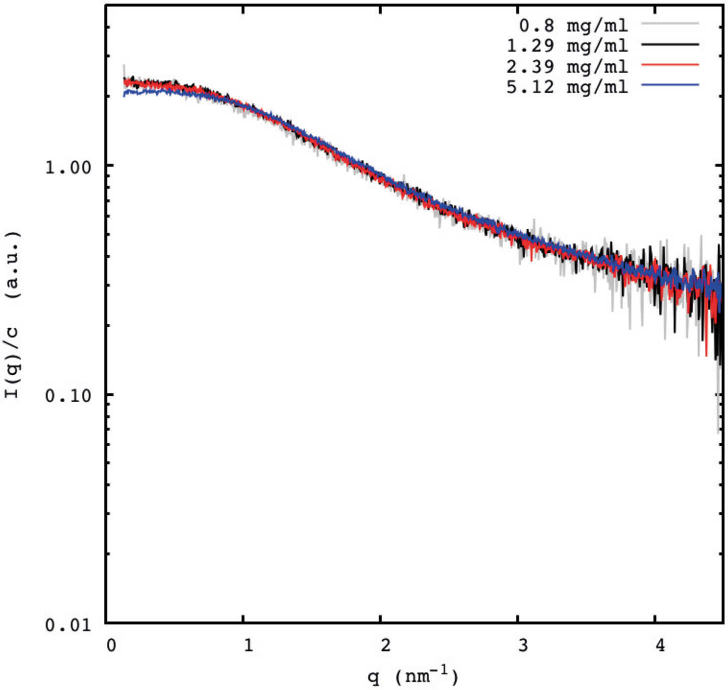


Figure S4: The scattering intensities of Hst5$_{4-15}$-Spd in 10 mM Tris buffer and 140 mM NaCl at various peptide concentrations.

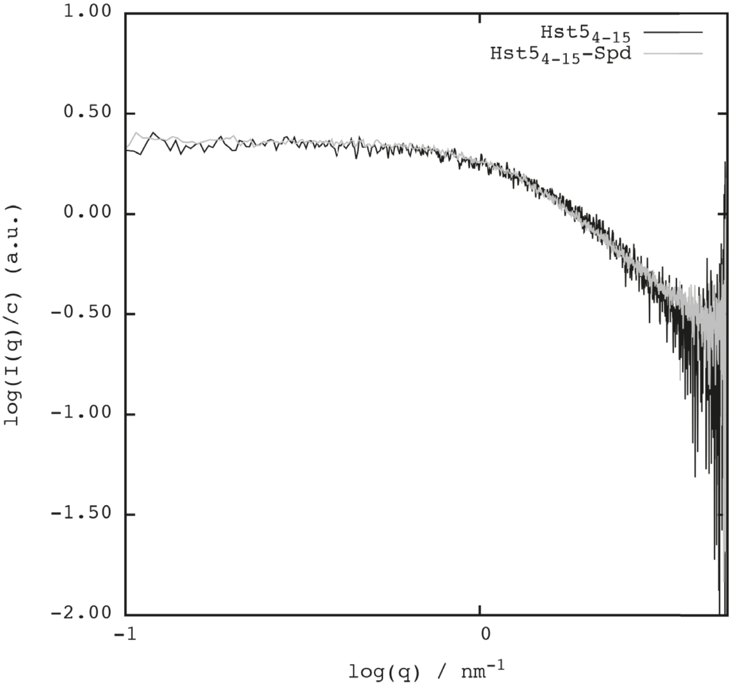


Figure S5: The scattering intensities of the Hst5$_{4-15}$ peptide and the Hst5$_{4-15}$-Spd conjugate in log-log scale. The concentrations are 1.50 mg/mL and 1.29 mg/mL for conjugate and peptide sample, respectively. The curves are flat at low q, indicating that the measured q-range is sufficient and that the samples are monodispersed without any apparent intermolecular interactions.

Paper II

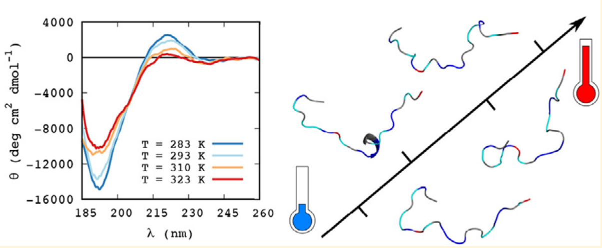
JCTC Journal of Chemical Theory and Computation
Article
Cite This: J. Chem. Theory Comput. 2019, 15, 2672−2683
pubs.acs.org/JCTC

Temperature Dependence of Intrinsically Disordered Proteins in Simulations: What are We Missing?

S. Jephthah,*[†] L. Staby,[‡] B. B. Kragelund,[‡] and M. Skepö*[†]

[†]Division of Theoretical Chemistry, Lund University, 221 00 Lund, Sweden
[‡]Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark

Supporting Information



ABSTRACT: The temperature dependence of the conformational properties in simulations of the intrinsically disordered model protein histatin 5 has been investigated using different combinations of force fields, water models, and atomistic and coarse-grained methods. The results have been compared to experimental data obtained from NMR, SAXS, and CD experiments to assess the accuracy and validity of the simulations. The results showed that neither simulations completely agreed with the experimental data, nor did they agree with each other. It was however possible to conclude that the observed conformational changes upon variations in temperature were not at all driven by electrostatic interactions. The final conclusion was that none of the simulations that were investigated in this study was able to accurately capture the temperature induced conformational changes of our model IDP.

1. INTRODUCTION

To understand the structural properties of intrinsically disordered proteins (IDPs), it is also necessary to study the possible conformational changes that could arise when the proteins are subjected to changes in temperature. While “‘regular’” globular proteins are generally unfolded upon heating, IDPs have been reported to display the opposite behavior, showing temperature-induced partial folding or secondary structure formation instead.[1,2] Several studies have suggested that disordered proteins are in fact locally ordered, possessing a well-defined polyproline II (PPII) backbone structure, and that it is the destabilization of this structure that causes the contraction of IDPs at higher temperatures.[3−5]

Using computer simulations accompanied by experiments is a good way of studying conformational changes, as the simulations can give more detailed information about the system, while the experiments can validate the simulations as well as give an overview and an overall appreciation of the results. Several studies like this have been performed on IDPs, and quite a few of them have obtained good accuracy when it comes to comparing simulated and experimental results at room temperature. The results are unfortunately not as comparable when performing similar studies at other temperatures than room temperature. In the study by Wuttke et al.,[2] Monte Carlo simulations were performed with the OPLS/ABSINTH force field and implicit solvent, in which the force field was modified to also consider the temperature-dependent dielectric constant and solvation free energies. As a result of this, the temperature-induced collapse of a selection of disordered proteins was captured, which highlighted the importance of solvation effects in simulations of IDPs using implicit solvent. Temperature effects have however not been as easily replicated in molecular dynamics simulations with explicit solvent. In 2009, Nettels et al.[6] showed that the temperature dependence in protein simulations strongly depends on the selected force field and water model. They noticed that the AMBER ff03* force field[7] in combination with the TIP3P water model[8] gave an increase in the radius of gyration, R_g, with increased temperature, whereas a decrease in the R_g was found when the same force field was used in combination with the TIP4P-Ew water model,[9] which again stressed the importance of using an appropriate water model. Using the protein force field OPLS-AA/L with TIP3P also gave a slight collapse of the R_g with temperature.[10] Similar

Received: December 20, 2018
Published: March 13, 2019

DOI: 10.1021/acs.jctc.8b01281
J. Chem. Theory Comput. 2019, 15, 2672−2683

results were observed several years later by Zerze et al.,[11] when protein simulations using the AMBER ff03ws force field[12] and the TIP4*P*/2005 water model[13] managed to qualitatively capture the temperature dependent collapse, although the Rg was always smaller than what experiments showed.

The fact that different water models give different results has been known for a long time. A small temperature dependence study (in which only two temperatures were tested) was done using the TIP4P water model[8] by Martin Neumann in 1986.[14] He showed that the dielectric constant decreases with temperature as it should, although the actual values were found to be too low. In 2004, Horn et al.[9] showed that the general tendency of decreasing dielectric constant with increased temperature was similar to the experimental values when using the TIP4P-Ew water model. The actual values were again too low, even though they were found to be considerably better than those obtained by using the TIP4P water model. Soon after, Abascal and Vega[13] determined that the dielectric constant at 298 K had better agreement with experiments when using the TIP4P-Ew water model compared to the TIP4P/2005 water model. Several studies have reported the dielectric constant of the TIP3P water model at room temperature to be higher than the expected value.[15−18] However, finding any studies in the literature concerning the temperature dependence of the dielectric constant of the TIP3P water model has been futile.

From the discussion above, it can be concluded that there have been several different studies concerning temperature dependence in simulations of proteins using different force field and water model combinations. While several simulations have shown to be sufficiently accurate at room temperature, the results from different temperatures were more varied. Thus, the question remains: *Is it possible for the currently available simulation methods to accurately mimic the experimental temperature induced structural changes in IDPs?*

In this case study, the temperature effects in simulations of the intrinsically disordered model protein histatin 5 (Hst5) have been investigated using both atomistic molecular dynamics (MD) and coarse-grained Monte Carlo (CG MC) simulations. Two of the currently available force fields appropriate for studying IDPs were chosen for the MD simulations: (i) AMBER ff99SB-ILDN[19] and (ii) CHARMM36m.[20] The AMBER ff99SB-ILDN force field was used in combination with the TIP4P-D water model,[21] but also the TIP3P water model,[8] which is known for causing too compact IDP structures when used in combination with this particular force field. The CHARMM36m force field was used in combination with the recommended TIP3P water model. For the CG MC simulations, we chose to use the Molsim simulation package (version 4.8.8).[22] To validate the simulation methods, the results were compared to data from various experimental methods such as small-angle X-ray scattering (SAXS), nuclear magnetic resonance spectroscopy (NMR), and circular dichroism spectroscopy (CD). The model protein, Hst5, is a small IDP that can normally be found in saliva, where it acts as a defense against fungal infections caused by, for example, *C. albicans*.[23,24] Hst5 is comprised of only 24 amino acid residues, of which almost 30% are histidines. In addition, seven of the other amino acid residues are positively charged, but only two are negatively charged. From this amino acid distribution, the conformational preference of Hst5 is predicted to be a polyampholytic coil or a polyelectrolytic semiflexible rod or coil, depending on the charge of the histidines.[25] When the histidines are charged, the protein obtains a more polyelectrolytic character. For the sake of simplicity, all simulated histidines were assumed to be neutral in this study. Illustrations of the atomistic and coarse-grained models of Hst5 are depicted in Figure 1.

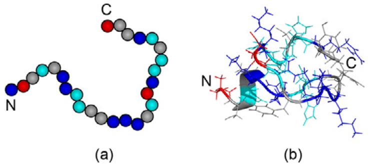


Figure 1. Illustration of the simulation models of Hst5. (a) The coarse-grained model and (b) the atomistic model overlaid with a cartoon representation to elucidate the backbone. The color scheme is as follows: red spheres represent the negative residues, blue spheres represent the positive residues, and the cyan spheres represent the histidine residues. The coarse-grained model also includes the protein termini as charged spheres.

2. METHODS

2.1. Computational Models and Methods. The model IDP, Hst5, has been simulated at different temperatures using several different combinations of simulation models and methods, see Tables 1 and 2. MC simulations were also performed for two additional CG peptides of equal length as Hst5 but with either (i) all residues being neutral or (ii) all residues having a charge of +1, see Table 2. Although all of the systems were included in the study, the main focus was on the MD simulations and the MC(IW) simulations (IW = implicit water).

Table 1. Abbreviations of the Different Combinations of Force Fields and Water Models Used in the All-Atom MD Simulations

abbreviation	force field	water model
MD(A3)	AMBER ff99SB-ILDN[19]	TIP3P[8]
MD(A4)	AMBER ff99SB-ILDN[19]	TIP4P-D[21]
MD(C3)	CHARMM36m[20]	TIP3P[8]

Table 2. Abbreviations of the Different Combinations of Chain Composition and Ionic Strength, *I*, Used in the CG MC Simulations

abbreviation	chain composition	*I* (mM)
MC_HST_10	Hst5	10
MC_NEU_10	all neutral	10
MC_POS_10	all charged	10
MC(IW)	Hst5	140
MC_NEU_140	all neutral	140
MC_POS_140	all charged	140

2.1.1. Atomistic MD Simulations. Molecular dynamics simulations were performed using the GROMACS package (version 4.6.7),[26−29] with the AMBER ff99SB-ILDN[19] and the CHARMM36m[20] force fields, together with the TIP3P[8] and the TIP4P-D[21] water models. The simulations were performed for two different kinds of simulation systems: (1) pure water and (2) Hst5. The water molecules were put in cubic boxes with edges of 2.5, 4.0, 5.0, 6.5, and 8.0 nm. A rhombic dodecahedron was used as the simulation box for Hst5, with a minimum distance of 1 nm between the solute and the box edges. The protein systems were neutralized by the addition of five chloride ions. Periodic boundary conditions were used in all directions. The equations of motion were integrated using the Verlet leapfrog algorithm[30] with a time step of 2 ps. Energies and coordinates were saved every 5 ps. A Verlet list cutoff scheme was used for the nonbonded interactions, of which the short-ranged interactions were calculated using a pair list with a cutoff of 1 nm. Long-ranged dispersion interactions were applied to the systems' energy and pressure. The long-ranged electrostatics were managed by using Particle Mesh Ewald[31] with cubic interpolation and a grid spacing of 0.16 nm. A velocity-rescaling thermostat[32] and a Parrinello–Rahman pressure coupling[33] were used to keep the temperature and the pressure constant throughout the simulations. The simulations were performed using a pressure of 1 bar. Selected temperatures in the interval of 280–323 K were used for both the pure water simulations and the Hst5 simulations. All bond lengths were constrained using the LINCS algorithm.[34] The initial Hst5 model was built as a linear structure using PyMOL.[35] Initiations were performed in two steps: (1) 500 ps NVT simulations and (2) 1000 ps NPT simulations to stabilize the temperature and the pressure of the systems. The final production runs were simulated for 50 ns each for the water simulations, and for 5 × 1 μs each for the Hst5 simulations, with the exception of the MD(A4) simulation at 293 K, which was run for a total of 7 μs. No salt was used in any of these simulations since it has already been established in our previous studies that the addition of salt does not affect the structural properties of small peptides such as Hst5.[36,37]

2.1.2. Coarse-Grained Simulation Model. A coarse-grained model was used for the Monte Carlo simulations, in which each amino acid residue and the N- and C-termini are represented by hard spheres (monomers/beads) connected to each other by harmonic bonds (see Figure 1). The bead radius was set to 2 Å and the equilibrium bond length was set to 4.1 Å, with a force constant of 0.4 N/m. Each sphere was either neutral or charged according to the amino acid sequence at pH 7. The interactions between the nonbonded beads were simulated using a short-ranged attraction of 0.6×10^4 kJ Å^6/mol (corresponding to 0.6 kT), and an extended Debye–Hückel potential using the inverse Debye screening length, κ, and the dielectric constant, ε_r, of the solvent as input variables. Further details of the simulation model are described in the recent paper by Cragnell et al.[38]

2.1.3. Monte Carlo Simulations. The coarse-grained Metropolis Monte Carlo simulations were performed using an in-house extended version of the Molsim simulation package (version 4.8.8)[22] in the canonical ensemble (NVT). The coarse-grained Hst5 peptide was simulated in a cubic box with edges of 250 Å. Periodic boundary conditions were applied in all directions. The solvent in the box was implicit water with an ionic strength of 10 mM or 140 mM. The investigated temperatures in the simulation box were in the interval of 283–323 K. Each simulation was equilibrated for 10 × 2 ×10^5 steps, followed by 10 × 10^6 steps in the production run. Four different simulation moves were used throughout the simulations: (i) single bead translation, (ii) pivot rotation, (iii) chain translation, and (iv) chain slithering. Results from the MC(IW) simulations were used for the comparison with the MD simulations and the experiments. This particular system was chosen to facilitate the comparison with the experimental results in particular, in which physiological ionic strength was used.

2.1.4. Simulation Analyses. The average radius of gyration, $\langle R_g \rangle$ in the MD simulations was obtained by using the GROMACS tools g_polystat, and the dielectric constant was computed using g_dipoles. Frames for representative structures were obtained by using g_cluster and illustrated by using PyMOL.[35] The minimum distance between the periodic images of each MD simulation was calculated using g_mindist. By monitoring that the minimum distance never approached the cutoff distance of the nonbonded interactions in the system, it was ensured that the simulated protein chain did not interact with its periodic images. Theoretical SAXS intensities from the MD simulations were obtained using CRYSOL (version 2.8.2). The secondary structure was analyzed using the DSSP program (version 2.2.1),[39] with modifications according to Chebrek et al.[40] to include analysis of the PPII structure. Probability density functions of the R_g, $p(R_g)$, were used to analyze and confirm the accuracy of the sampling in the simulations (see the Supporting Information). Estimations of the full width half-maximum, FWHM, of the $p(R_g)$ were obtained according to Cragnell et al.,[38] that is, by fitting the $p(R_g)$ curves with the Gaussian function

$$f(x) = a \cdot \exp\left[-\frac{(x-b)^2}{c^2}\right] \tag{1}$$

and computing the FWHM according to the following equation:

$$\text{FWHM} = 2c\sqrt{\ln(2)} \tag{2}$$

The variables a, b, and c are fitting parameters, and the FWHM values were reported with a 95% confidence interval.

2.2. Experiments and Techniques. Three experimental techniques have been used in this study: (i) SAXS, (ii) NMR, and (iii) CD. Two buffers were used in the experiments: (i) a 20 mM Na_2HPO_4/NaH_2PO_4 phosphate buffer at pH 7.0 and (ii) a 20 mM Tris buffer at pH 7.5. The phosphate buffer would be preferred to use in all cases due to its ideal buffer range and negligible temperature dependence. Unfortunately, the phosphate buffer was not suitable for the SAXS experiments, in which it has been shown to cause severe radiation damage. Thus, the Tris buffer was used for these experiments instead but with a slightly increased pH to maintain a decent buffering action. The ionic strength of the solutions was set to 140–150 mM in the SAXS and CD experiments. This was done to mimic physiological conditions and to exclude any intermolecular interactions. The effect of the ionic strength is, however, assumed to be negligible because of the small size of the Hst5 peptide, which has also been seen in previous SAXS measurements and simulations.[36,37] Thus, the NMR measurements remained salt-free in this study.

2.2.1. SAXS and CD Sample Preparations. The Hst5 peptide was obtained from Genemed Synthesis Inc. (San Antonio, TX, USA) as white powder with 95.73% purity and trifluoroacetate (TFA) as counterion. Tris buffers (Saveen Werner AB, >99.9% purity, CAS Registry no. 77-86-1) for the SAXS measurements were prepared at a concentration of 20 mM in Milli-Q water, and acidified with HCl so that the buffers would maintain pH 7.5 at the appropriate measuring temperatures. Calculations of the pH at different temperatures were performed by using a $d(pK_a)/dT$ value of −0.028. Sodium phosphate monobasic monohydrate (NaH_2PO_4, Sigma-Aldrich, ≥99.0% purity, CAS Registry no. 10049-21-5) and sodium phosphate dibasic dihydrate (Na_2HPO_4, Sigma-Aldrich, ≥99.5% purity, CAS Registry no. 10028-24-7) were used to prepare a 20 mM phosphate buffer in Milli-Q water for the CD measurements. The pH of the phosphate buffer was set to 7.0 with NaOH. The ionic strength of the buffer solutions was set to 140−150 mM. NaCl was used to adjust the ionic strength for the SAXS measurements, and NaF (VWR International, ≥99.5% purity, CAS Registry no. 7681-49-4) was used for the CD measurements. Before any peptide was added, the buffers were filtered through a 0.2 μm hydrophilic polypropylene membrane (Pall Corporation). After the peptide was dissolved, a concentration cell (Vivaspin 20, 2.0 kDa MWCO, product no. VS02H92, Sartorius Stedim Biotech GmbH, Göttingen, Germany) was used to remove low molecular weight impurities. The samples were rinsed with buffer by centrifugation to a maximum of 3500 RCF at 15−18 °C, and the amount of buffer used for rinsing was equal to at least 10 times the sample volume. Dialysis of the SAXS samples was done using dialysis cups (Slide-A-Lyzer MINI, 2.0 kDa MWCO, product no. 69580, Thermo Scientific, United States), and the CD samples were dialyzed using dialysis cassettes (Slide-A-Lyzer Dialysis Cassettes, 2.0 kDa MWCO, product no. 66205, Thermo Scientific, United States). All samples were dialyzed for at least 12 h in a buffer volume of at least 100× the sample volume to ensure exact background.

2.2.2. SAXS Measurements.

SAXS measurements were performed at beamline BM29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The incident-beam wavelength was 0.99 Å, and the distance between the sample and the PILATUS 1 M detector was 2867 nm, giving a scattering vector range of 0.037−4.928 nm^{-1}. For each sample and pure solvent, at least ten successive 1 s frames were recorded and analyzed. Special attention was paid to radiation damage by comparing the successive frames prior to further processing of the data. The measurements were performed at selected temperatures of 10, 20, 37 and 50 °C. Before the SAXS measurements, the samples were further centrifuged at 18400 RCF and 6 °C for at least 30 min to remove aggregates. Protein concentrations were measured after preparation and again immediately before the SAXS measurements using a Nanodrop spectrophotometer ($\varepsilon = 2560\ M^{-1}\ cm^{-1}$, $\lambda = 280$ nm). To obtain the proper SAXS spectra, the background (i.e., the pure solvent) was subtracted from the corresponding sample spectrum. Normalization was done by converting $I(0)$ to an absolute scale by measuring the scattering of pure water.

2.2.3. CD Measurements.

Far-UV (185−260 nm) CD spectra were obtained using a Jasco J-715 CD spectrometer with a model PTC-348WI Peltier type temperature control system (Hachioji, Tokyo, Japan). The samples were filtered though a Millex-GV Filter with a pore size of 0.22 μm (Merck Millipore Ltd., Ireland), before being added to a quartz cuvette with a path length of 0.1 cm. Selected temperatures of 10, 20, 37, and 50 °C were used for the measurements. The scanning rate was 20 nm/min, with a bandwidth of 2.0 nm, and a response time of 2 s. At least five scans were performed and averaged for each measurement. All spectra were smoothed using Savitzky−Golay filtering (SGF) with a bandwidth of 75. The background spectra were subtracted from the respective sample spectra and checked for oversmoothing before presenting the final data (see example in the Supporting Information). The smoothing process was done following the protocol of Norma J Greenfield.[41]

2.2.4. NMR Measurements.

All NMR experiments were acquired on a Bruker Avance III 600 MHz (^1H) spectrometer equipped with a TCI cryogenic probe. Samples contained 1 mM Hst5 in a 20 mM mixture of Na_2HPO_4/NaH_2PO_4 (pH 7.0), 10% D_2O, 0.25 mM DSS, and 0.25% (v/v) 1,4-dioxane. Pulsed-field gradient (PFG) NMR diffusion experiments were recorded using a standard Bruker PFG-LED pulse sequence with bipolar gradients and solvent presaturation during the relaxation delay of 3 s.[42] A total of 32 spectra with gradient strengths ranging from 2% to 98% of its maximum value were recorded. The pseudo-2D data was Fourier transformed and baseline corrected in Topspin (Bruker) and subsequently processed in Dynamics Center (Bruker) to obtain peak intensities. Diffusion constants were determined by fitting of peak intensity decays against the gradient strength using the Stejskal−Tanner equation:

$$I = I_0 e^{-g^2\gamma^2\delta^2(\Delta-\delta/3)D} \tag{3}$$

where I is the intensity, g the gradient strength, γ the gyromagnetic ratio of ^1H, δ the gradient length, Δ is the diffusion time, and D is the diffusion constant. The diffusion constants of a selection of peaks from the aromatic and aliphatic side-chain regions were plotted as histograms and fitted to a Gaussian

$$f(x) = A\cdot e^{-(x-\mu)^2/2\sigma^2} \tag{4}$$

where μ is the mean, σ the standard deviation, and A the amplitude. Different bin sizes where tested to check the robustness of the fit. A small amount of 1,4-dioxane with a known hydrodynamic radius, $R_{H,ref}$, of 2.12 Å was used as an internal reference and used for the calculation of $R_{H,prot}$ according to Wilkins et al.[43]

$$R_{H,prot} = \frac{D_{ref}}{D_{prot}}\cdot R_{H,ref} \tag{5}$$

3. RESULTS AND DISCUSSION

In this study, both atomistic MD simulations and coarse-grained MC simulations of Hst5 have been performed at different temperatures. Due to the extensive computer resources needed for the MD simulations, only four temperatures were tested with this method. By using a coarse-grained model for the MC simulations, a substantial amount of computational time was saved, which enabled the study of a wider range of temperatures. Analysis of the simulated data provided R_g probability density functions, as well as the average radius of gyration, $\langle R_g\rangle$, and SAXS profiles of each simulated system. Secondary structure analysis could also be obtained from the MD simulations. Discussion about convergence and sampling is referred to the Supporting Information. The results from the experimental methods, that is, SAXS, NMR, and CD, were utilized to assess the validity of the simulations and to discern any temperature-dependent conformational changes. In this section, we will start by presenting and discussing changes in $\langle R_g\rangle$ for all the different methods that have been used in this study (if applicable). Subsequently, the shape, flexibility, and secondary structure of

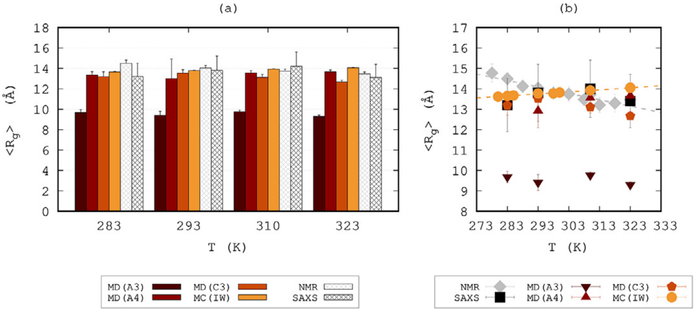


Figure 2. Radius of gyration, $\langle R_g \rangle$, as a function of temperature from all six methods. (a) Histograms to emphasize the difference in $\langle R_g \rangle$ between the methods at the different temperatures. (b) The change in temperature of the different methods. The dashed lines in panel b are linear regression trend lines. All data points include error bars, but the errors are too small to be visible in some cases.

Table 3. $\langle R_g \rangle$ **of All Methods at the Four Selected Temperatures**

	$\langle R_g \rangle$ (Å)					
T (K)	SAXS	NMR	MD(A3)	MD(A4)	MD(C3)	MC(IW)
283	13.2 ± 1.3	14.49 ± 0.31	9.68 ± 0.28	13.34 ± 0.33	13.19 ± 0.47	13.65 ± 0.03
293	13.8 ± 1.4	14.03 ± 0.26	9.41 ± 0.39	12.92 ± 0.82	13.52 ± 0.34	13.78 ± 0.03
310	14.0 ± 1.4	13.73 ± 0.20	9.77 ± 0.15	13.55 ± 0.22	13.11 ± 0.29	13.92 ± 0.04
323	13.4 ± 1.3	13.46 ± 0.19	9.30 ± 1.13	13.66 ± 0.18	12.67 ± 0.15	14.05 ± 0.03

our model peptide will be analyzed before the finalizing concluding remarks.

3.1. Temperature Correlations in the Conformational Ensemble. The $\langle R_g \rangle$ at different temperatures was obtained from both the MD and the MC simulations, as well as from SAXS and NMR. A conversion factor of 1.1 was used to estimate the $\langle R_g \rangle$ from the hydrodynamic radius, R_H, which was obtained from the NMR experiments. The conversion factor was computed using the method presented by Nygaard et al.[44] The $\langle R_g \rangle$ from SAXS was obtained from PRIMUS using the Guinier approximation with $qR_g(\max) < 0.8$. A comparison between the $\langle R_g \rangle$ from all different methods is presented in Figure 2. Figure 2a emphasizes the difference in $\langle R_g \rangle$ between all methods at the selected temperatures, and Figure 2b illustrates the $\langle R_g \rangle$ changes as a function of temperature. Linear regression analysis was done on all data sets, assuming statistical significance for $p > 0.05$ (see Table S1 for more detailed information). The $\langle R_g \rangle$ values at four selected temperatures are also displayed in Table 3. Visual inspection of Figure 2a revealed that the $\langle R_g \rangle$ from all methods except MD(A3) were similar at the selected temperatures. The reason for the lower $\langle R_g \rangle$ values with this force field and water type combination has already been frequently and thoroughly discussed[19,45−48] and will, thus, not be explained here. However, the observed trends in Figure 2b did give a different view of the results: while the NMR data displayed a distinct decreasing behavior with increased temperature, the $\langle R_g \rangle$ from the SAXS measurements showed a small increase until the peak value of 310 K, after which it decreased again. Taking the linear regression analysis into consideration, the $\langle R_g \rangle$ from the NMR data was confirmed to have a strong and significant correlation with the temperature. The same could not be concluded about the SAXS results, for which no linear correlation could be identified. The $\langle R_g \rangle$ from the MD(A3) and the MD(C3) simulations showed a correlation that was similar to the NMR results, although weaker and not significant in the selected confidence level. An opposite correlation between the $\langle R_g \rangle$ and the temperature was found in the MD(A4) simulations compared to the other MD simulations, but this correlation was also proven to be insignificant. The results from the MC(IW) showed a positive correlation with increasing temperature, which was also confirmed to be the only simulated correlation of statistical significance. Because of the known problem of sampling,[46] the lack of significant correlation in the MD simulations was not surprising. However, the change in $\langle R_g \rangle$ that would arise because of improved sampling would most likely not be of enough magnitude to change the observed correlations, based on the results found by Hicks and Zhou,[49] where they only observed small differences (≤0.5 Å) for the $\langle R_g \rangle$ with and without enhanced sampling methods. This value corresponds to approximately one-fourth of the bead radius in our coarse-grained model, and is only slightly larger than the error estimates of the MD simulations. In fact, the $\langle R_g \rangle$ only changed from 13.00 to 12.92 Å upon extending two of the replicates in the MD(A4) simulation from 1 to 2 μs.

Since the SAXS and NMR experiments gave different results, it is difficult to determine which one of the temperature-dependent conformational changes is the most reliable. There were several basic factors differing between the two experiments: (i) the buffer, (ii) the pH, and (iii) the ionic strength. However, as already mentioned, neither of these differences is expected to give rise to such huge deviations as was seen in the results. Thus, we are looking for answers elsewhere in future studies. One observation that was done in the SAXS experiments at 310 K and higher, was that the computed

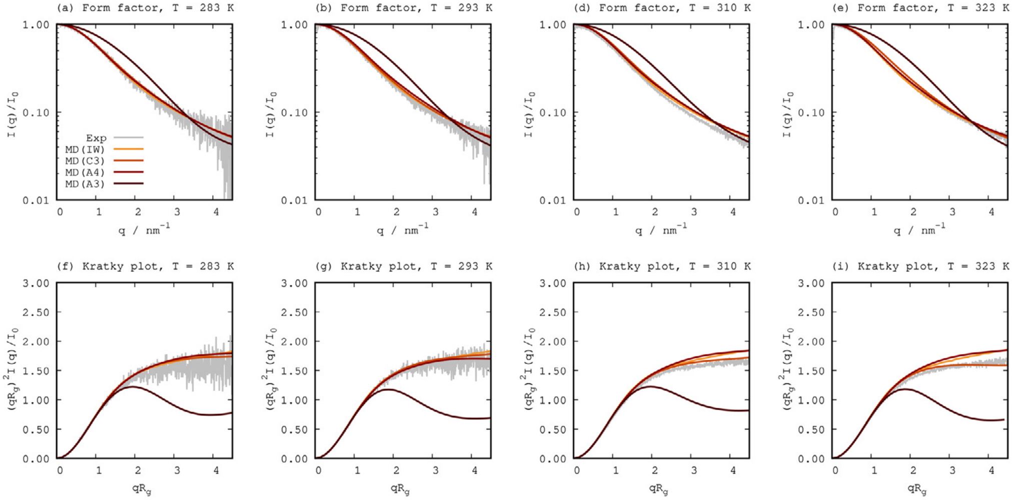


Figure 3. Form factor (upper row, a−e) and the normalized Kratky plot (lower row, f−i) from SAXS measurements (gray) compared to computed SAXS curves from simulations at four different temperatures (left to right).

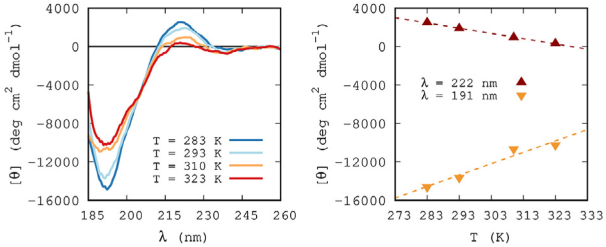
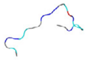
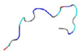
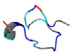
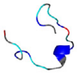
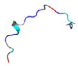
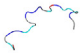


Figure 4. (a) Normalized CD spectra of Hst5 at different temperatures in samples with an ionic strength of 150 mM and (b) the change in ellipticity at wavelengths of 191 and 222 nm with temperature.

molecular mass became significantly lower (i.e., a deviation of more than 10%) for samples of low concentration, which could potentially mean that there is some sort of destabilization of the peptide under these conditions. This is however only a speculation, and the SAXS data presented in this paper had their molecular masses within the acceptable deviation limits. Another possibility is that the change in $\langle R_g \rangle$ is too small for being properly distinguished by the SAXS.

3.2. Shape and Flexibility. By studying the SAXS spectra, information on the shape and flexibility of the peptide can be obtained. Comparing the experimental data to that from simulations provide an important tool in validation of the simulation models. Comparison of the results from the simulations and the experiment in Figure 3 makes it clear that the MD(A4), the MD(C3), and the MC(IW) simulations managed to imitate the experimental form factor very well at 293 K (Figure 3b). The normalized Kratky plot at this temperature (Figure 3g) also agree on the shape of the peptide, which from the plateau indicated a high degree of flexibility. The MD(A3) simulations did not agree with the experiments, which can be seen by the indication of a more globular shape in both the form factor and the Kratky plot. All SAXS results at 293 K were in agreement with previous studies.[36,47] Visual inspection of the form factor suggested that the agreement between simulations and experiment seems to hold at 283 and 323 K but not at 310 K. Comparison of the Kratky plot does, however, show larger differences between the simulations and the experiments at all temperatures other than 293 K. The deviations are more pronounced at higher qR_g values, where the simulations generally have slightly higher values. It is difficult to discern if this difference is significant or not since the Krakty plot is very sensitive in this region. Even small alterations due to, for example, the background subtraction can cause the Kratky plot to change at high qR_g values. Steeper positive slopes usually indicates increasingly extended or rigid structures, but the difference is expected to be larger for such a significant structural change. Furthermore, it is important to keep in mind that experimental SAXS has quite low resolution compared to the SAXS curves from the simulations. Another reason for the differences between the simulations and the experiments could be that the histidines are assumed to be neutral throughout all of the simulations, even though histidines are known to charge regulate in solution. This

Table 4. Cartoon Depiction of the Representative Structures from the MD Simulations and Their Corresponding Secondary Structure Per Amino Acid According to DSSPPII Analysis

T (K)	MD(A3)	MD(A4)	MD(C3)	MD(A4)/MD(C3)
283	 -TTT-EETTEE--TTT-S--	 -STT----SS--TTS--	 --S-TT-SS-TT-S-S--	
293	 -TTTTS-SS--S-SS-TT--	 -TTS-TTS-S-SS-SS--	 ----STTTTT----S--	
310	 -TT-S-SGGGT-S--TTT-	 --SS-STT--S-S--TT--	 -S-SS--SS-SSTTTSGGG-	
323	 -HHHHTT-STT-SS--S---	 -SSS-TTSSS--SS-S--	 --SSSS---SSS-SS--	

does however seem unlikely since a previous study by Kurut et al.[54] has shown that the charge regulation of the histidines is very low at neutral pH in simulations. Additionally, protonation of the histidines results in a more extended structure due to the additional electrostatic repulsion that accompanies the protonation (see Figure S4). Inspection of the experimental Kratky curves proved no change in the shape and flexibility of the peptide with the temperature. The differences between the simulated SAXS spectra at different temperatures were also found to be visually negligible for each model and are thus only shown in Figure S5.

3.3. Secondary Structure. Temperature dependence in the secondary structure was studied by CD spectroscopy, see Figure 4a. The spectra showed a negative band around 191 and 240 nm, as well as a positive band around 222 nm. The absolute intensity of the bands was shown to decrease as the temperature was increased. The change in ellipticity at 191 and 222 nm as a function of temperature is illustrated in Figure 4b. Linear regression analysis revealed a significant correlation between the ellipticity and the temperature at the selected bands (see Table S1), which indicates destabilization of PPII structure with increasing temperature.[55] Since the PPII structure is quite extended, one could speculate that the peptide becomes less extended when the temperature is increased, which agrees very well with the NMR results. This is, however, not supported by the SAXS results, which indicated no significant change in shape and flexibility with the temperature. Shifts in the bands toward longer wavelengths were observed, which is also associated with a destabilization of the PPII structure. Measurements at high temperatures and shorter wavelengths are however less reliable due to increased tension (HT) voltage. This makes it difficult to deduce whether or not this observation at 191 nm was the expected natural response with some discrepancies because of the limitations of the method or if it was a random variation because of insufficient resolution.

Cluster analysis was performed to acquire a central, or representative, structure of each simulation, on which secondary structure analysis was performed in addition to the complete MD trajectories (see Figure S6). Contrary to the experiments, the representative structures did not show any traces of PPII structure for any force field or temperature, see Table 4. Instead the representative structures were found to be dominated by bends (S), turns (T), and random coils (−). These structures were also dominating in the complete simulation trajectories. As expected, the MD(A3) simulations gave rise to more compact turn and helical structures (H, α-helix; G, 3_{10}-helix) at a temperature of 293 K and above. At 283 K, the central part of the peptide was instead interpreted as an extended strand (E). The turn content was more distinguished in the MD(C3) simulations compared to the MD(A4) simulations. One deviation from the unstructured behavior was found for MD(C3) at 310 K, which representative structure suggested a 3_{10}-helix toward the C-terminus of the protein. Visual inspection of the representative structures from the MD(A4) and MD(C3) simulations showed similar conformations and are, thus, aligned in the last column of the table to visually enhance the comparison. In the simulations using the AMBER force fields, a few residue sequences were found to consistently adopt turn conformations. In the MD(A3) simulations, the preserved structure was found around His-3, Ala-4, and Lys-5. These residues were found in a turn conformation at all temperatures except 323 K, where they were found in an α-helix conformation instead. The turn in the MD(A4) was not present at 283 K but could otherwise be found at His-8 and Gly-9. The MD(C3) was not seen to possess such specific structure propensity. It is not unreasonable to expect turns around these type of amino acids since they are known to be disorder-promoting.[56] However, further studies are required to discern if this structural preservation is a true effect or simply due to insufficient sampling. While $\langle R_g \rangle$ might not change significantly, the secondary structure analysis is more likely to benefit from enhanced sampling. Although improbable, inadequate sampling could also be a reason for the lack of PPII structure in the simulations. Only extending the simulations is not likely to be sufficient to change the conformation of the representative structure of the simulation. This was observed when comparing the 5 μs version of the MD(A4) simulation to the extended 7 μs version, for which the structures were identical (see Figure S7). However, since we are only interested in the overall secondary structure content and not specific details about every possible conformation, and although it would be interesting to expand the study using enhanced sampling methods, this is out of scope in our study.

3.4. Temperature-Dependent Interactions in the Simulations. To exclude the possibility that the water models could cause an inaccurate temperature dependence in the MD simulations, pure water simulations were also performed using the AMBER ff99SB-ILDN force field in combination with the two water models. The dielectric constant was computed from the water simulations at different temperatures, and compared to the reference values, see Figure 5. The figure showed that the dielectric constant of both the TIP3P and the TIP4P-D water model decreased with increased temperature, as expected. The actual magnitudes of the obtained dielectric constants were, however, shown to deviate, resulting in larger values than the reference values while using the TIP3P water model, and values smaller than the reference values while using the TIP4P-D water model. Since decreasing correlations were captured in the water models, that is, a correct temperature dependence was observed, any possible erroneous contribution to the behavior of the Hst5 simulations can most likely be disregarded.

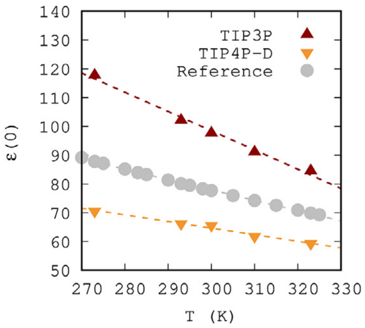


Figure 5. Dielectric constant of the TIP3P and the TIP4P-D water models at different temperatures compared to the reference values from Malmberg and Maryotte (1956),[50] Owen et al. (1961),[51] Uematsu and Frank (1980),[52] and Fernandez et al. (1997).[53]

Further investigation of the dielectric constant in the simulations was done by performing additional MC simulations using the dielectric constants from the MD simulations instead of the standard reference values. The resulting $\langle R_g \rangle$ values were found to be basically identical to those from the MC(IW) simulations (and are thus not shown), which indicates that the electrostatic contribution is not dominating the temperature dependence in the MC simulations. This observation was further explored by performing additional coarse-grained MC simulations of two other chains of equal length as Hst5, but consisting completely of (i) neutral beads and (ii) beads with a charge of +1 each (see Table 2). The results are shown in Figure 6, where the temperature dependence of the $\langle R_g \rangle$ was found to be the same regardless of what fraction of charge the chain possesses. The temperature dependence was also found to be independent of the ionic strength. These results combined strengthens the notion of that the temperature dependence is not electrostatically driven in the coarse-grained MC simulations.

The strength of the electrostatic interactions is dependent on the temperature through the dielectric constant, and the interactions become stronger at higher temperatures. Consequently, if the electrostatic interactions were dominating, the simulated peptide would most likely either (i) become more extended at higher temperatures because of increasing repulsion between the positive residues in the chain, (ii) show an increased interaction between residue Glu-16 and any of the positive amino acid residues, or (iii) display a combination of these two effects. Since neither of the proposed effects was observed in the MD results, this could indicate that the effect of temperature on the conformational properties of

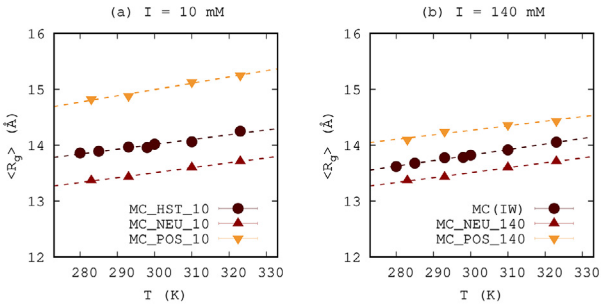


Figure 6. Comparison of the change in $\langle R_g \rangle$ with temperature for three chains of equal length but with different charge compositions. The simulations were performed using two different ionic strengths: (a) 10 and (b) 140 mM. See Table 2 for a more detailed description of the simulated systems.

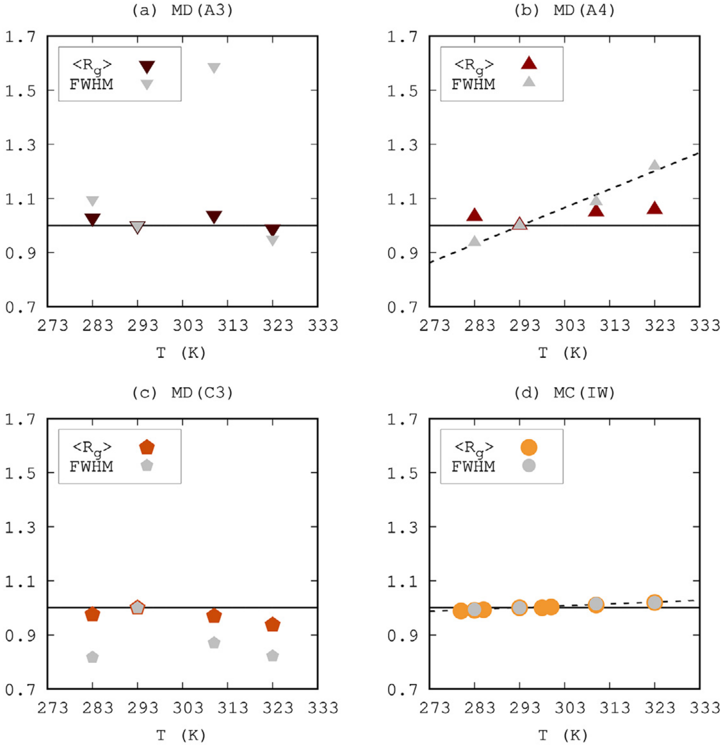


Figure 7. $\langle R_g \rangle$ compared to the FWHM as a function of temperature for the four different simulation methods. All points are normalized with their respective value at room temperature.

Hst5 is not primarily driven by the electrostatic interactions in the MD simulations.

Assuming that the electrostatic interactions are not the main driving force for the temperature dependence of conformational changes in Hst5 in the simulations leaves only one option: the entropic effect. To evaluate the conformational entropy in the simulated systems, the FWHM of $p(R_g)$ (Figure S2) was determined. The results were normalized and compared to the normalized $\langle R_g \rangle$, see Figure 7. Linear correlation between the FWHM and the temperature was found for the MD(A4) and the MC(IW) simulations only. Since the correlations were found to be positive, this means that the chain entropy was increasing with the temperature. Unfortunately, several of the data points of the remaining two methods cannot be considered reliable because of bad Gaussian fitting caused by seemingly two or more radius distributions under the same curve (e.g., 310 K in Figure S2a). Thus, another fitting model or improved sampling would be needed to properly evaluate the conformational entropy of the MD(A3) and MD(C3) simulations. Another observable correlation was identified between the $\langle R_g \rangle$ and the FWHM in the MC(IW) simulation, which suggests that the conformational changes heavily depend on the entropic forces. To investigate the correlation further, linear regression analysis was performed to prove its significance (see Figure S3 and Table S1).

To summarize, the results indicated that the temperature dependence of the conformational properties of Hst5 most likely is driven more by entropic effects than by the electrostatic interactions in the simulations, especially in the case of the MC simulations. This observation is in agreement with the loss off PPII structure that was observed in the experiments. Nevertheless, neither of the simulations was able to predict any accurate temperature dependence of the conformational ensemble, which gives rise to the question in the title of this study: What are we missing? Considering the MD simulations, both the TIP4P-D water model and the CHARMM36m force field have been modified to sustain more favorable protein−water dispersion interactions. As mentioned in the introduction, MC simulations with added solvation free energies have also been shown to give more accurate results. Thus, although several studies have shown that proper balancing of protein−solvent interactions is a key element to obtain good simulation results, it is obviously not enough to only scale these interactions to capture conformational properties of IDPs at other temperatures than room temperature. In addition, since the electrostatic interactions and the entropic effects are the only temperature dependent interactions in the simulations, these are either not completely accurate, or it might also be needed to adjust the strength of the short-ranged interactions with respect to the temperature.

4. CONCLUSION

The case study presented here has focused on investigating the accuracy and validity of several simulation models and methods considering temperature induced structural changes. The simulated results have been compared to each other, but also to experimental results. As noted in several previous studies, we again concluded that the choice of force fields and water models heavily influences the outcome of the MD simulations, and also when it comes to temperature effects. It was observed that all three studied MD systems displayed different responses upon changes in temperature. Weak correlations between the temperature and the $\langle R_g \rangle$ were seen, but these were statistically insignificant, which is most likely attributed to the known problem of insufficient sampling. While enhanced sampling would decrease the errors and stabilize the values, and thereby also give more prominent correlations, it is unlikely that it would change the general trends observed in this study. The only simulation method that gave an actual significant temperature correlation was the CG MC method. This correlation was not supported by the NMR experiments, which instead suggested the opposite temperature dependence. Secondary structure analysis of the MD simulations was also unable to match the experimental results. The CD strongly indicated a loss of PPII structure, but this structure was not found in any of the simulated systems, and neither did these systems show any consistent correlation in secondary structure changes. However, since the secondary structure analysis was based on the representative structure from each simulation, enhanced sampling could possibly lead to a different distribution of secondary structure without significantly change averages of the other studied properties. Furthermore, we also deduced that it is unlikely that the water models are responsible for the different temperature behaviors and that the temperature induced correlations are not electrostatically driven in neither of the simulation methods. In conclusion, this study has shown that several of the simulation models and methods that are currently available for simulations of IDPs do *not* manage to capture the temperature induced structural changes, since none of the methods presented here was able to accurately explain the experimentally observed correlations.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b01281.

> Notes on CD subtraction and smoothing, as well as simulation convergence and sampling; linear regression statistics, additional FWHM plots, and supplementary Kratky plots; and time evolution of the secondary structure and supplementary representative structures from the MD simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: stephanie.jephthah@teokem.lu.se. Phone: +46-46-222 17 54.
*E-mail: marie.skepo@teokem.lu.se. Phone: +46-46-222 33 66.

ORCID

S. Jephthah: 0000-0003-4287-506X
M. Skepö: 0000-0002-8639-9993

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge financial support from NordForsk's Neutron Science Programme, the Crafoord Foundation, and the Danish Research Council (#4181-00344 to BBK). Computer resources for the simulations were provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC). We also thank the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for providing beamtime, and Dr. Bart van Laer for assistance at beamline BM29. Finally, we would also like express our gratitude to Dr. João Henriques and Carolina Cragnell for providing valuable guidance regarding computer simulations and SAXS experiments, respectively.

■ REFERENCES

(1) Uversky, V. N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* **2009**, *28*, 305−325.

(2) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5213−5218.

(3) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chem. Rev.* **2006**, *106*, 1877−1897.

(4) Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α−helices or loss of polyproline II? *Protein Sci.* **2010**, *19*, 1555−1564.

(5) Sreerama, N.; Woody, R. W. *Methods in Enzymology*; Elsevier, 2004; Vol. *383*, pp 318−351.

(6) Nettels, D.; Müller-Späth, S.; Küster, F.; Hofmann, H.; Haenni, D.; Rüegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; et al. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20740−20745.

(7) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

(9) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(10) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(11) Zerze, G. H.; Best, R. B.; Mittal, J. Sequence-and temperature-dependent properties of unfolded and disordered proteins from atomistic simulations. *J. Phys. Chem. B* **2015**, *119*, 14622−14630.

(12) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein−Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.

(13) Abascal, J. L.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(14) Neumann, M. Dielectric relaxation in water. Computer simulations with the TIP4P potential. *J. Chem. Phys.* **1986**, *85*, 1567−1580.

(15) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* **1998**, *108*, 10220−10230.

(16) Höchtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. Rationalization of the dielectric properties of common three-site water models in terms of their force field parameters. *J. Chem. Phys.* **1998**, *109*, 4927−4937.

(17) Wu, Z.; Cui, Q.; Yethiraj, A. A new coarse-grained model for water: the importance of electrostatic interactions. *J. Phys. Chem. B* **2010**, *114*, 10524−10529.

(18) Braun, D.; Boresch, S.; Steinhauser, O. Transport and dielectric properties of water and the influence of coarse-graining: Comparing BMW, SPC/E, and TIP3P models. *J. Chem. Phys.* **2014**, *140*, 064107.

(19) 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., Genet.* **2010**, *78*, 1950−1958.

(20) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71.

(21) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.

(22) Jurij, R.; Per, L. MOLSIM: A modular molecular simulation software. *J. Comput. Chem.* **2015**, *36*, 1259−1274.

(23) Raj, P. A.; Edgerton, M.; Levine, M. Salivary histatin 5: dependence of sequence, chain length, and helical conformation for candidacidal activity. *J. Biol. Chem.* **1990**, *265*, 3898−3905.

(24) Tsai, H.; Raj, P. A.; Bobek, L. A. Candidacidal activity of recombinant human salivary histatin-5 and variants. *Infect. Immun.* **1996**, *64*, 5000−5007.

(25) Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102−112.

(26) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(27) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(28) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(29) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(30) Berendsen, H.; Van Gunsteren, W. Practical algorithms for dynamic simulations. *Molecular-Dynamics Simulation of Statistical-Mechanical Systems* **1986**, 43−65.

(31) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(33) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. (Melville, NY, U. S.)* **1981**, *52*, 7182.

(34) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(35) Schrödinger, LLC. *PyMOL Molecular Graphics System*, version 1.2r1; 2009.

(36) Cragnell, C.; Durand, D.; Cabane, B.; Skepö, M. Coarse-grained modelling of the intrinsically disordered protein Histatin 5 in solution: Monte Carlo simulations in combination with SAXS. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 777−791.

(37) Jephthah, S.; Henriques, J.; Cragnell, C.; Puri, S.; Edgerton, M.; Skepö, M. Structural Characterization of Histatin 5−Spermidine Conjugates: A Combined Experimental and Theoretical Study. *J. Chem. Inf. Model.* **2017**, *57*, 1330−1341.

(38) Cragnell, C.; Rieloff, E.; Skepö, M. Utilizing Coarse-Grained Modeling and Monte Carlo Simulations to Evaluate the Conformational Ensemble of Intrinsically Disordered Proteins and Regions. *J. Mol. Biol.* **2018**, *430*, 2478.

(39) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

(40) Chebrek, R.; Leonard, S.; de Brevern, A. G.; Gelly, J.-C. PolyprOnline: polyproline helix II and secondary structure assignment database. *Database* **2014**, *2014*, No. bau102.

(41) Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **2007**, *1*, 2876.

(42) Wu, D.; Chen, A.; Johnson, C. An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *J. Magn. Reson., Ser. A* **1995**, *115*, 260−264.

(43) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, *38*, 16424−16431.

(44) Nygaard, M.; Kragelund, B. B.; Papaleo, E.; Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys. J.* **2017**, *113*, 550−557.

(45) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95*, L07−L09.

(46) Henriques, J.; Cragnell, C.; Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theory Comput.* **2015**, *11*, 3420−3431.

(47) Henriques, J.; Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: on the accuracy of the TIP4P-D water model and the representativeness of protein disorder models. *J. Chem. Theory Comput.* **2016**, *12*, 3407−3415.

(48) Huang, J.; MacKerell, A. D. Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2018**, *48*, 40−48.

(49) Hicks, A.; Zhou, H.-X. Temperature-induced collapse of a disordered peptide observed by three sampling methods in molecular dynamics simulations. *J. Chem. Phys.* **2018**, *149*, 072313.

(50) Malmberg, C.; Maryott, A. Dielectric Constant of Water from 00 to 1000 C. *J. Res. Natl. Bur. Stand* **1956**, *56*, 2641.

(51) Owen, B. B.; Miller, R. C.; Milner, C. E.; Cogan, H. L. The Dielectric Constant of Water As a Function of Temperature and Pressure1, 2. *J. Phys. Chem.* **1961**, *65*, 2065−2070.

(52) Uematsu, M.; Frank, E. Static dielectric constant of water and steam. *J. Phys. Chem. Ref. Data* **1980**, *9*, 1291−1306.

(53) Fernández, D. P.; Goodwin, A.; Lemmon, E. W.; Levelt Sengers, J.; Williams, R. A formulation for the static permittivity of water and steam at temperatures from 238 to 873 K at pressures up to 1200 MPa, including derivatives and Debye−Hückel coefficients. *J. Phys. Chem. Ref. Data* **1997**, *26*, 1125−1166.

(54) Kurut, A.; Henriques, J.; Forsman, J.; Skepö, M.; Lund, M. Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 657−667.

(55) Chemes, L. B.; Alonso, L. G.; Noval, M. G.; de Prat-Gay, G. In *Intrinsically Disordered Protein Analysis*; Uversky, V. N., Dunker, A. K., Eds.; Humana Press: Totowa, NJ, 2012; Vol. *895*; pp 387−404.

(56) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **2014**, *114*, 6561−6588.

Supporting information for: Temperature Dependence of Intrinsically Disordered Proteins in Simulations: What are We Missing?

S. Jephthah,*,† L. Staby,‡ B. B. Kragelund,‡ and M. Skepö*,†

†*Division of Theoretical Chemistry, Lund University, Lund, Sweden*
‡*Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Copenhagen, Denmark*

E-mail: stephanie.jephthah@teokem.lu.se; marie.skepo@teokem.lu.se
Phone: +46-46-222 17 54; +46-46-222 33 66

1 CD subtraction and smoothing

All spectra were smoothed using Savitzky–Golay filtering (SGF) with a bandwidth of 75. The spectra were checked for over-smoothing by (a) visually comparing the smoothed spectra to the raw data, and by (b) subtracting the smoothed curve from the raw data, see Figure S1. Since the data points in Figure S1(b) are evenly distributed around zero, the data was assumed to not be over-smoothed. The smoothed background spectra were then subtracted from the respective smoothed sample spectra. The smoothing process was done following the protocol of Norma J Greenfield.[S1]

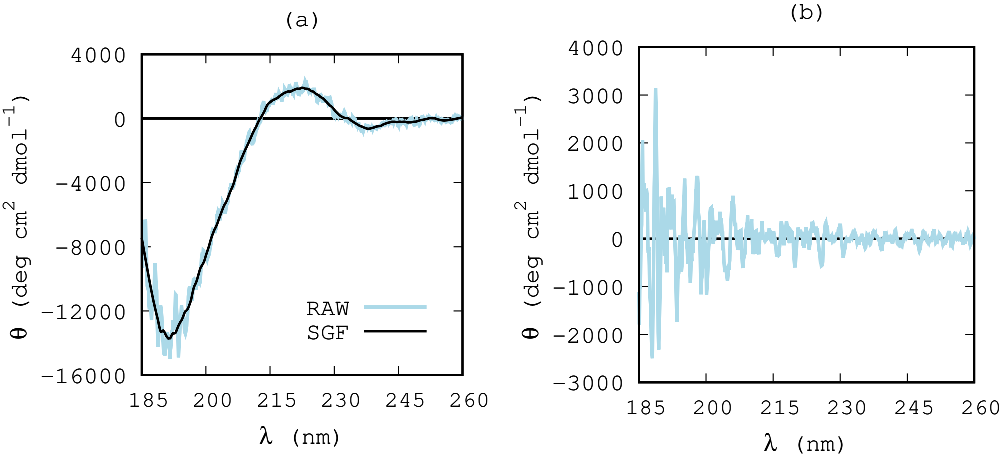


Figure S1: Example of smoothening of a CD spectra using Savitzky-Golay filtering (SGF) with a bandwidth of 75. (a) The smoothened data compared to the raw data, and (b) the difference between the raw data and the smoothened data. The data used for this example is Hst5 in a solution with an ionic strength of 150 mM at 20 °C. All curves are normalized.

2 Simulation convergence and sampling

Probability density function estimates of the radius of gyration, $p(R_g)$, at different temperatures are shown in Figure S2 for the four different combinations of simulation models and methods. As expected, the `MD(A3)` simulations sample a more narrow range of conformations at smaller values compared to the rest of the simulation methods. The figure also shows that the `MD(A4)`, the `MD(C3)` and the `MC(IW)` simulations sample similar ranges of R_g. The peaks of the `MC(IW)` probability density functions shifts slightly towards lower radii as the temperature is decreased, but no immediate change in the shape of the curves is observed, suggesting that the chain flexibility does not change significantly upon the temperature changes using this simulation method. Possible peak shifts and changes in flexibility are more difficult to analyze in the MD simulations, due to the known issues with conformational sampling of IDPs[S2] in combination with only a small variation with temperature. Even though the simulations have been run for at least 5 µs, more sampling (e.g. by simulating for an even longer time or by applying enhanced sampling) is needed to discern these type of changes from the probability density function estimates in this case.

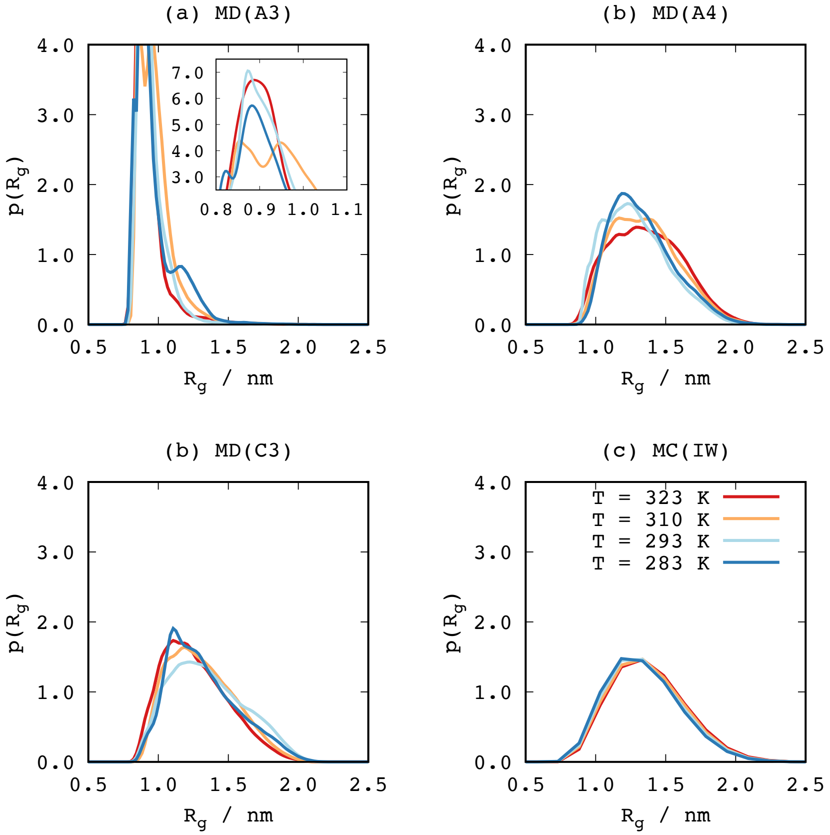


Figure S2: The distance distribution probability functions of the radius of gyration, R_g, as obtained from the four different simulation methods.

3 Statistics

Table S1: Linear regression statistics. Slope (m), intercept (c), correlation coefficient (r), coefficient of determination (r^2), and probability value (p).

	m	c	r	r^2	p
From R_g vs. T					
NMR	-0.031	23.224	-0.935	0.873	0.000
SAXS	0.005	12.068	0.247	0.061	0.753
`MD(A3)`	-0.005	11.037	-0.398	0.158	0.602
`MD(A4)`	0.013	9.488	0.698	0.488	0.302
`MD(C3)`	-0.015	17.740	-0.775	0.601	0.225
`MC(IW)`	0.010	10.872	0.996	0.991	0.000
`MC_HST_10`	0.009	13.772	0.964	0.930	0.036
`MC_NEU_10`	0.009	10.862	0.998	0.995	0.002
`MC_POS_10`	0.011	11.638	0.991	0.983	0.009
`MC_NEU_140`	0.009	10.862	0.998	0.995	0.002
`MC_POS_140`	0.008	11.829	0.980	0.959	0.020
From FWHM plots					
`MD(A3)`	0.002	0.825	0.106	0.011	0.894
`MD(A4)`	0.037	-5.349	0.989	0.977	0.011
`MD(C3)`	-0.008	8.382	-0.260	0.067	0.740
`MC(IW)`	0.004	4.950	0.989	0.979	0.011
From FWHM vs. R_g					
`MD(A3)`	1.435	-12.166	0.827	0.684	0.173
`MD(A4)`	1.378	-12.686	0.683	0.466	0.317
`MD(C3)`	1.269	-10.785	0.781	0.610	0.219
`MC(IW)`	0.425	0.328	0.982	0.965	0.018
From θ vs. T					
$\lambda(191)$	0.069	-28.068	0.973	0.946	0.027
$\lambda(222)$	-0.040	13.071	-0.998	0.996	0.002
From $\varepsilon(0)$ vs. T					
TIP3P	-0.609	299.106	-0.995	0.991	0.000
TIP4P-D	-0.229	133.326	-0.991	0.982	0.001
Reference	-0.359	185.773	-0.999	0.999	0.000

4 FWHM

The *FWHM* was plotted as a function of the $\langle R_g \rangle$ to illustrate possible correlations, see Figure S3. The linear regression statistics is found in Table S1. A somewhat positive correlation between the *FWHM* and the $\langle R_g \rangle$ can be seen from the r-value, but the linear fit was not impressive. The p-value did however show that the correlation is not statistically significant, except for in the `MC(IW)` simulations.

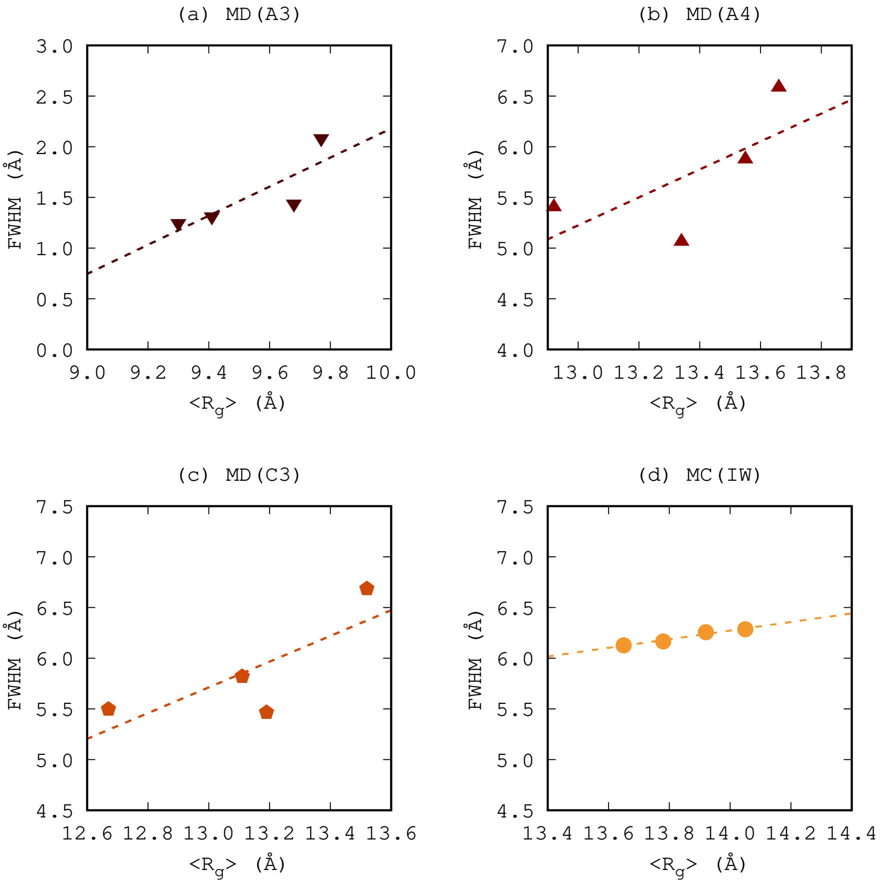


Figure S3: The FWHM as a function of the $\langle R_g \rangle$. The linear regression stats are available in Table S1.

5 SAXS

The difference in structure of Hst5 when its histidines are either all neutral or fully protonated is depicted in Figure S4. The simulations emphasizing the differences have been performed at 300 K, but the experimental data and the `MD(A4)` simulation at 293 K are also included for comparison.

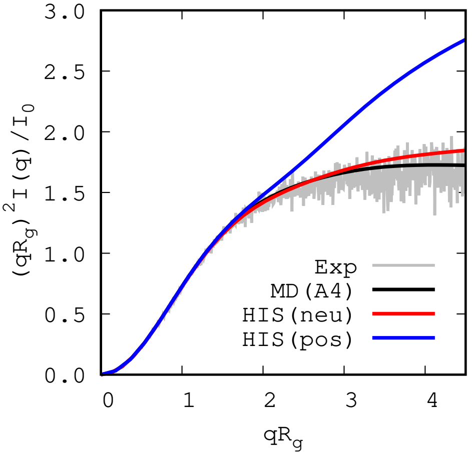
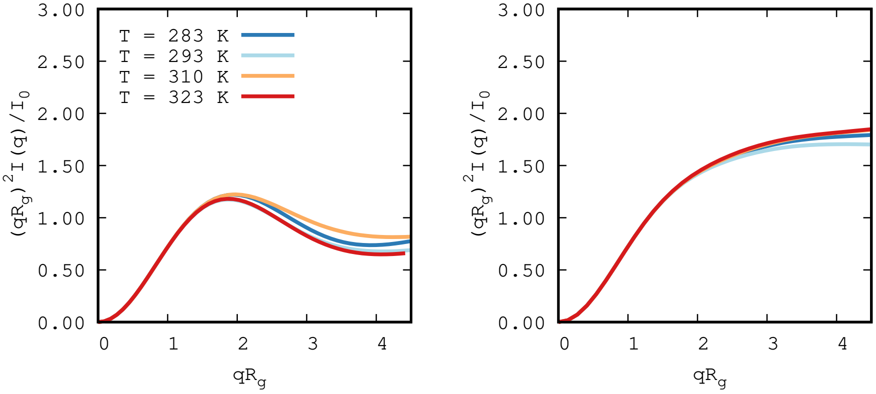
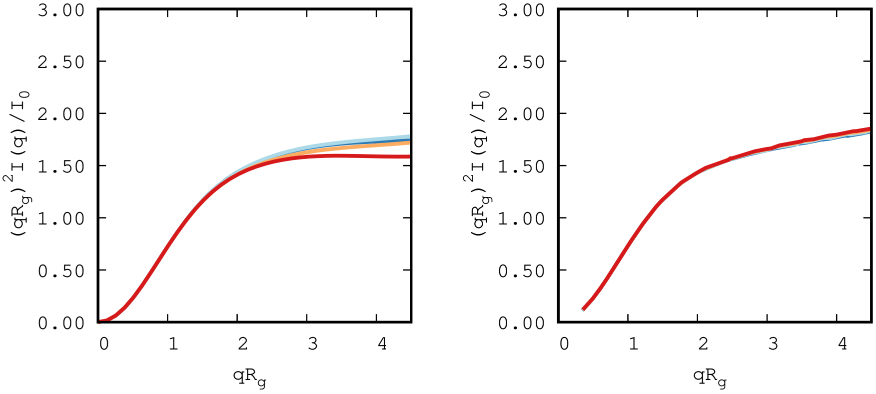


Figure S4: Comparison of the normalized Kratky plots of the Hst5 model with neutral histidines only (black and red) and with fully protonated histidines (blue). The experimental SAXS data (gray) is also included for comparison.


(a) MD(A3)

(b) MD(A4)


(c) MD(C3)

(d) MC(IW)

Figure S5: Comparison of the normalized Kratky plots of the different simulation methods at four different temperatures.

6 Secondary structure

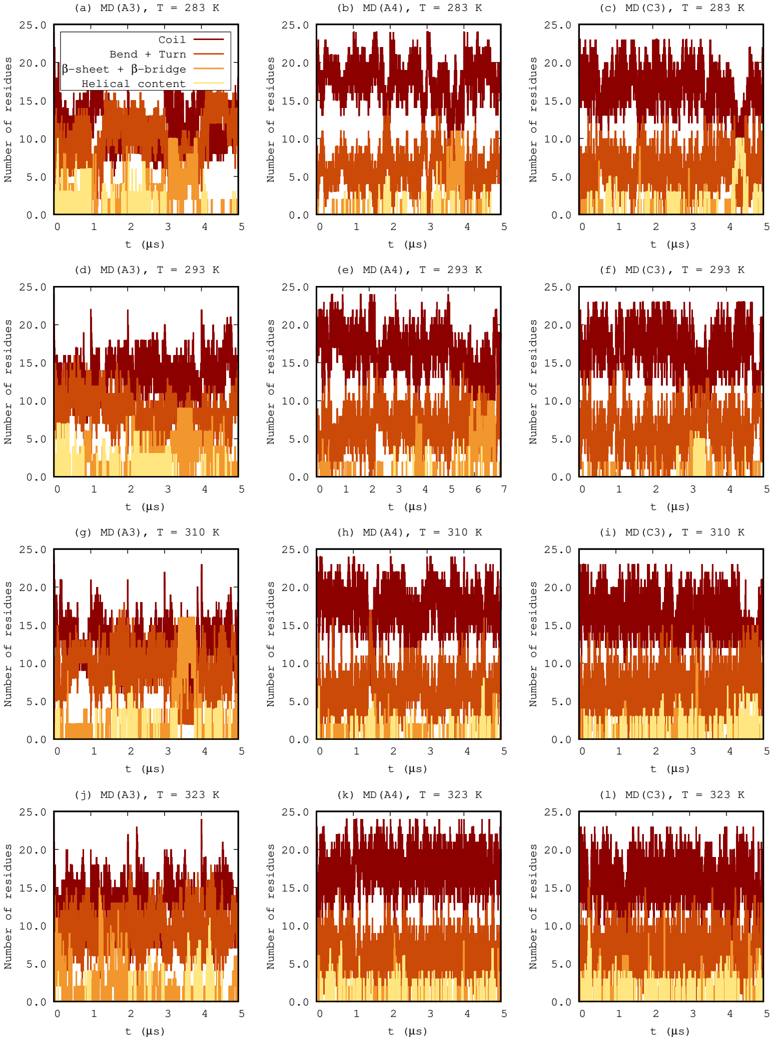


Figure S6: The change in secondary structure over time in the MD simulations.

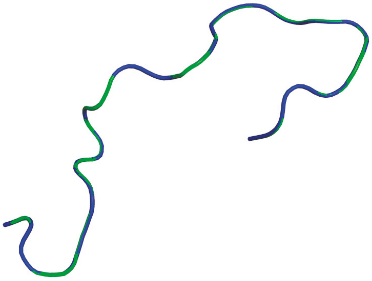


Figure S7: Aligned representative structures of the `MD(A4)` simulation at 293 K after 5 µs (green) and 7 µs (blue).

References

(S1) Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **2006**, *1*, 2876.

(S2) Henriques, J.; Cragnell, C.; Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theory Comput.* **2015**, *11*, 3420–3431.

Paper III

biomolecules

MDPI

Article

Physicochemical Characterisation of KEIF—The Intrinsically Disordered N-Terminal Region of Magnesium Transporter A

Stéphanie Jephthah [1], Linda K. Månsson [1], Domagoj Belić [2], Jens Preben Morth [3] and Marie Skepö [1,4,*]

[1] Division of Theoretical Chemistry, Department of Chemistry, Lund University, Naturvetarvägen 14, 221 00 Lund, Sweden; stephanie.jephthah@teokem.lu.se (S.J.); linda.mansson@live.com (L.K.M.)
[2] Division of Physical Chemistry, Department of Chemistry, Lund University, Naturvetarvägen 14, 221 00 Lund, Sweden; domagoj.belic@fkem1.lu.se
[3] Enzyme and Protein Chemistry, Section for Protein Chemistry and Enzyme Technology, Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, 2800 Kgs. Lyngby, Denmark; premo@dtu.dk
[4] Lund Institute of Advanced Neutron and X-ray Science (LINXS), Scheelevägen 19, 233 70 Lund, Sweden
* Correspondence: marie.skepo@teokem.lu.se; Tel.: +46-46-222-33-66

Received: 24 March 2020; Accepted: 13 April 2020; Published: 17 April 2020

Abstract: Magnesium transporter A (MgtA) is an active transporter responsible for importing magnesium ions into the cytoplasm of prokaryotic cells. This study focuses on the peptide corresponding to the intrinsically disordered N-terminal region of MgtA, referred to as KEIF. Primary-structure and bioinformatic analyses were performed, followed by studies of the undisturbed single chain using a combination of techniques including small-angle X-ray scattering, circular dichroism spectroscopy, and atomistic molecular-dynamics simulations. Moreover, interactions with large unilamellar vesicles were investigated by using dynamic light scattering, laser Doppler velocimetry, cryogenic transmission electron microscopy, and circular dichroism spectroscopy. KEIF was confirmed to be intrinsically disordered in aqueous solution, although extended and containing little β-structure and possibly PPII structure. An increase of helical content was observed in organic solvent, and a similar effect was also seen in aqueous solution containing anionic vesicles. Interactions of cationic KEIF with anionic vesicles led to the hypothesis that KEIF adsorbs to the vesicle surface through electrostatic and entropic driving forces. Considering this, there is a possibility that the biological role of KEIF is to anchor MgtA in the cell membrane, although further investigation is needed to confirm this hypothesis.

Keywords: membrane proteins; intrinsically disordered proteins; circular dichroism spectroscopy; small-angle X-ray scattering; cryogenic transmission electron microscopy; molecular-dynamics simulations; protein–vesicle interactions; magnesium transporter; secondary structure

1. Introduction

In a society where antimicrobial resistance is constantly manifesting in new ways, the demand for effective antibiotics is naturally increasing. In order to rationalise the design of new antibiotics, and to find new potential cellular targets, bacterial biochemical functions must be mapped and fully understood.

The magnesium ion Mg^{2+} is the most abundant divalent cation in any biological system and, it being an essential mineral nutrient and thus an absolute requirement for life, is present in every cell type in every living organism [1]. In cells, Mg^{2+} is an essential cofactor for more than

600 enzymes, including important DNA and RNA polymerases; it is also required for stabilisation of the ribosome–protein complex during protein synthesis. In adenosine triphosphate (ATP)-dependent enzymes, Mg^{2+} binds an ATP, the main unit of cellular energy, in the catalytic pocket, thus activating the phosphate ester towards hydrolysis [2]. In non-ATP-dependent enzymes, the role of Mg^{2+} is instead to hold a water molecule in a specific position, and this water molecule in turn helps to hold a particular structure in place or participates directly in the enzymatic reaction mechanism [2].

In bacteria and archaea, three major classes of Mg^{2+} transporters have been identified as responsible for the translocation of Mg^{2+} across the cell membrane: channel protein CorA, gated channel protein magnesium transporter E (MgtE), and pump magnesium transporters A and B (MgtA and MgtB) [2]. Only MgtA and MgtB serve as primary active transporters, but unlike other P-type ATPases, MgtA and MgtB mediate Mg^{2+} influx down, rather than against, the electrochemical gradient. Mg^{2+} acts as a product inhibitor for MgtA, which is activated by free Mg^{2+} concentrations below 10 µM and strongly inhibited by concentrations above 1 mM [3].

MgtA consists of 898 amino acids and has a molecular weight of 99.5 kDa. In a recent study by Subramani et al., (2016) [3], using the DISOPRED3 server for intrinsically disordered region (IDR) prediction [4], the first 33 amino acids (1–33) of MgtA from *Escherichia coli (E. coli)* were classified as intrinsically disordered. How the disordered nature of this IDR affects the biological function of MgtA however remains unknown. Thus, the focus of this study is to investigate this N-terminal, intrinsically disordered part of MgtA, hereafter referred to as KEIF. KEIF has the following amino acid sequence:

M F K E I F T R L I R H L P S R L V H R D P L P G A Q Q T V N T V.

At physiological pH, KEIF carries five positively charged (blue) and two negatively charged (red) amino acids, giving it a net charge of +3. The majority of charged amino acids are evenly distributed in the first half of the peptide.

Here, we present the first-ever physicochemical characterisation of KEIF, which was performed using a variety of computational and experimental techniques. First, primary-structure and bioinformatic analyses were performed in order to make predictions about the overall structure and behaviour of the peptide. Second, experimental techniques such as circular-dichroism (CD) spectroscopy and small-angle X-ray scattering (SAXS) were used in combination with atomistic molecular-dynamics (MD) simulations to characterise the undisturbed single chain in aqueous solution. Interactions with neutral and anionic large unilamellar vesicles (LUVs) were investigated by using dynamic light scattering (DLS), laser Doppler velocimetry (LDV), cryogenic transmission electron microscopy (cryo-TEM), and CD spectroscopy. Corroborating Subramani's DISOPRED3-based prediction [3], we found KEIF to display typical characteristics of a disordered peptide in aqueous bulk solution. Interestingly, dissolution in an organic solvent or the presence of anionic vesicles serve both to induce an increase of helical structure within the peptide. The obtained results might help shine some light on KEIF's role for the function of MgtA, as the question still remains: does KEIF play an important role in MgtA function, or should KEIF be regarded merely as a passive appendix?

2. Materials and Methods

2.1. Sample Preparation

2.1.1. Samples for CD Spectroscopy

KEIF powder (95.67%; Genemed Synthesis Inc., San Antonio, TX, USA) was dissolved in, and was purified by dialysis (100–500 Da MWCO Biotech Cellulose Ester (CE) Dialysis Membrane Tubing; SpectrumLabs, Piraeus, Greece) against, a 20 mM tris(hydroxymethyl)aminomethane (TRIS) buffer at pH 7.4 and at 6 °C. Following purification, the concentration of the peptide stock solution was determined using a Thermo Scientific NanoDrop 2000 Spectrophotometer (UV-Vis at 214 nm, $\varepsilon/1000 = 60.805\,\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$, $Mw = 3.871$ kDa). The stock solution was diluted with additional buffer to prepare 0.2 mg mL^{-1} (52 µM) samples for CD measurements, supplemented with either 10 or

150 mM NaF, and sometimes with 1 mM (corresponding to ∼20 eq.) $CaCl_2$, $MgCl_2$ or $ZnCl_2$. Samples were filtered (0.22 µm MILLEX-GV Filter Unit) prior to measurements.

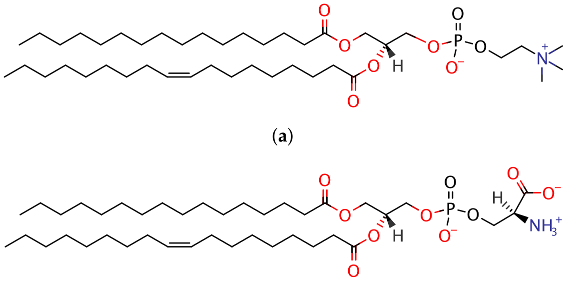
To prepare a sample of KEIF in 2,2,2-trifluoroethanol (TFE), KEIF powder was first dissolved in, and was purified by dialysis (100–500 Da MWCO Biotech Cellulose Ester (CE) Dialysis Membrane Tubing; SpectrumLabs, Piraeus, Greece) against, milliQ water at 6 °C. The sample was then lyophilised and the resulting purified powder was dissolved in TFE (>99%; Sigma-Aldrich, Stockholm, Sweden) to yield 0.2 mg mL^{-1} (52 µM). This sample was not filtered prior to measurements as filter units do not withstand organic solvents.

2.1.2. SAXS Samples

A 20 mM TRIS buffer was prepared and acidified with HCl to maintain a pH of 7.5. The ionic strength of the buffer was set to 140 mM using NaCl. KEIF powder was dissolved in the buffer and dialysed against the same buffer at 4 °C, using a 500–1000 Da MWCO Regenerated Cellulose (RC) Dialysis Membrane Tubing (SpectrumLabs, Piraeus, Greece). Before SAXS measurements, samples were centrifuged at 18,400 RCF at 6 °C for at least 30 min to remove impurities and aggregates. Protein concentrations were measured immediately before SAXS measurements using a NanoDrop One Microvolume UV-Vis Spectrophotometer (UV-Vis at 214 nm).

2.1.3. LUV Samples

Neutral LUVs were prepared using 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC; Avanti Polar Lipids, Alabaster, AL, USA), whereas anionic ones were prepared using a 3:1 (mol:mol) mixture of POPC and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS; Avanti Polar Lipids, Alabaster, AL, USA) (Figure 1). Lipids were (co)dissolved in 3:7 (v:v) methanol:chloroform in a glass vial. Solvents were evaporated under a stream of nitrogen, after which the resulting lipid film was further dried under reduced pressure (0.8 bar) overnight. The lipids were then hydrated with 20 mM TRIS buffer at pH 7.4 to a total lipid concentration of 30 mM, and the sample was then subjected to five freeze–thaw cycles before being extruded 31 times through a 0.1 µm polycarbonate membrane filter (Avanti Polar Lipids, Alabaster, AL, USA).



(**a**)

(**b**)

Figure 1. Lipids used for preparation of large unilamellar vesicles (LUVs): (**a**) 1-palmitoyl-2 -oleoyl-glycero-3-phosphocholine (POPC) and (**b**) 1-palmitoyl-2-oleoyl-sn-glycero-3- phospho-L-serine (POPS).

2.1.4. LUV–KEIF Samples

KEIF (purified by dialysis as described in Section 2.1.1) was mixed with POPC or 3:1 POPC:POPS LUVs to give final peptide and lipid concentrations of 750 µM and 12.0 mM, respectively. The used buffer was 20 mM TRIS at pH 7.4, and ionic strength was set to 10 mM with NaF. Samples were incubated overnight before DLS (Section 2.2.3) and LDV measurements (Section 2.2.4), and CD spectroscopy (Section 2.2.1). Cryo-TEM imaging (Section 2.2.5) was performed the following day.

Samples were used as for cryo-TEM imaging, but were diluted six times (to 125 µM peptide and 2.0 mM lipid) prior to CD measurements, and five more times (to 25 µM peptide and 0.4 mM lipid) prior to DLS and LDV measurements.

2.2. Experiment Methods

2.2.1. CD Spectroscopy

Far-UV CD spectra were recorded in 0.1 nm intervals (typically) between 190 and 260 nm, with four accumulations, on a JASCO J-715 spectropolarimeter equipped with a photomultiplier tube detector. Samples were measured in a 1 mm quartz cuvette (Hellma Analytics 110-QS). Temperature control was ensured using a PTC-348WI peltier-type temperature-control system. The measurement temperature was 20 °C, and samples were equilibrated for 5 min at this temperature prior to measurements. All spectra were corrected by subtracting a reference spectrum obtained from a sample lacking KEIF, but which was otherwise identical. Ellipticity is reported as mean residue molar ellipticity θ (degcm^{2}dmol^{-1}) according to Equation (1), where θ_{obs} is ellipticity (deg), mrw is mean residue molecular weight, c is protein concentration (g mL^{-1}), and l is the optical path length of the cell (cm).

$$\theta = \theta_{obs}(mrw)/10lc \tag{1}$$

Some of the obtained CD spectra were subject to BeStSel [5,6] fitting through a web-server (http://bestsel.elte.hu/index.php) to access the corresponding secondary structure elements. Fitted residuals are presented in Figures S1 and S2.

2.2.2. SAXS Experiments

SAXS experiments were performed at beamline BM29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The incident-beam wavelength was 0.99 Å, and the distance between sample and PILATUS 1M detector was set to 2869 mm. The temperature of the storage and exposure cells was 20 °C. By measuring the scattering of pure water, the forward scattering $I(0)$ was converted to an absolute scale. At least ten successive frames with an exposure time of 1 s were recorded for each sample. Scattering from the pure solvent was also measured both before and after each individual protein-sample measurement, and subtracted from the corresponding protein-sample spectrum. Special attention was paid to radiation damage by comparing successive frames prior to background subtraction in order to avoid inclusion of faulty data. Data were processed and analysed using the ATSAS package [7]. The ensemble optimisation method (EOM) [8,9] was used to fit theoretical scattering intensities to the experiment data.

2.2.3. DLS Measurements

A Malvern Zetasizer Nano-ZS (Malvern Instruments Ltd., Malvern, UK) equipped with a 633 nm 4 mW HeNe laser with automatic laser attenuator was used for DLS measurements. Disposable PMMA cuvettes (BRAND GMBH, Wertheim, Germany) were used as sample cells. The measurement temperature was 20 °C and samples were equilibrated for 5 min prior to measurements. Measurements were performed at a fixed scattering angle of 173° using the noninvasive backscatter (NIBS) technique. Data were analysed by the cumulant method provided by the instrument software. Hydrodynamic radius R_H (Z-average radius, or "cumulant mean", given by the software) is given as the average of five consecutive measurements of 60 s where standard deviation represents the error.

2.2.4. LDV Measurements

A Malvern Zetasizer Nano-ZS (Malvern Instruments Ltd., Malvern, UK) equipped with a 633 nm 4 mW HeNe laser with an automatic laser attenuator was used for LDV measurements to obtain estimates of electrophoretic mobility. Disposable folded capillary cells (Malvern DTS1070) were used as sample cells. The measurement temperature was 20 °C, and samples were equilibrated for 5 min prior to measurements. Measurements were performed at a fixed scattering angle of 17° using the M3-PALS laser interferometric technique. Electrophoretic mobility is given as the average of three consecutive measurements where standard deviation represents the error.

2.2.5. Cryo-TEM Imaging

Cryo-TEM sample preparation and imaging were performed at the National Center for High-Resolution Electron Microscopy within Lund University. Lacey formvar-carbon film on 200 mesh copper TEM grids (Ted Pella, Redding, CA, USA) were glow-discharged in a Quorum GloCube system (Quorum Technologies, Laughton, UK). Then, 4 µl of vesicle suspension was pipetted onto the TEM grid in a Leica EM GP automatic plunge freezer (Leica Microsystems, Stockholm, Sweden) operating at 21 °C and relative humidity of $>90\%$, backside-blotted for 2.5 s, and plunged into liquid ethane. Samples were transferred onto a Fischione 2550 cryogenic sample holder and imaged on a JEOL JEM-2200FS (JEOL, Tokyo, Japan) transmission electron microscope equipped with an omega energy filter, operating at an accelerating voltage of 200 kV. Sample temperature was kept below −174 °C during imaging. The zero-loss images were acquired on an F416.0 camera (TVIPS, Gauting, Germany) using Serial EM software [10] running in low-dose mode (total electron dose per image $<15\,e^{-}\,\text{Å}^{-2}$). The acquired cryo-TEM images were processed using ImageJ software [11].

2.3. Calculations

2.3.1. Isoelectric-Point Calculation

Theoretical isoelectric-point (pI) calculation was performed using the Swiss Institute of Bioinformatics (SIB) Bioinformatics Resource Portal ExPASy (Expert Protein Analysis System) [12] `Compute pI/Mw` tool through a web-server (https://web.expasy.org/compute_pi/).

2.3.2. Partitioning-Free-Energy Calculation

Water-to-bilayer partitioning-free-energy calculation was performed using the `Membrane Protein Explorer` (`MPEx`) tool [13] and the Wimley–White octanol scale [14,15]. His was considered neutral, Lys and Arg positive, Asp and Glu negative, and both termini charged.

2.4. Simulations

2.4.1. Atomistic MD Simulations

Atomistic MD simulations were performed using the GROMACS package (version 4.6.7) [16–18], with an AMBER ff99SB-ILDN force field [19] and TIP4P-D water model [20]. A rhombic dodecahedron was used as a simulation box, with periodic boundary conditions in all directions. A minimal distance of 1 nm was set between solute and box edges. An initial, linear structure of KEIF was built using PyMOL [21]. The two His residues in the amino acid sequence were set to be neutral throughout the simulations, giving the peptide a net charge of +3. Three chloride ions were added to neutralise the system. Simulations were performed without the addition of any salt.

The equations of motion were integrated using the Verlet leap-frog algorithm [22] with a time step of 2 fs. A Verlet list cut-off scheme was used for the nonbonded interactions. Short-ranged interactions were calculated using a pair list with a cut-off of 1 nm. Long-ranged dispersion interactions were applied to the systems' energy and pressure, and long-ranged electrostatics was managed by using Particle Mesh Ewald [23] with cubic interpolation and a grid spacing of 0.16 nm. All bond lengths were constrained using the LINCS algorithm [24]. A velocity-rescaling thermostat [25] with a relaxation time of 0.1 ps was used to keep a temperature of 300 K. A Parrinello–Rahman pressure coupling [26] was used to keep pressure constant at 1 bar throughout the simulations. Relaxation time was 2 ps, and isothermal compressibility was set to that of water, i.e., 4.5×10^{-5} bar^{-1}.

Energy minimisation was done using the steepest-descent algorithm. Initiation was performed in two steps with position restraints on the peptide to equilibrate the temperature and pressure of the systems: (1) 500 ps NVT simulations, and (2) 1000 ps NPT simulations. Five replicates with different starting seeds were used for each simulation. The production runs were also performed in the NPT ensemble, and were run for a total of 10 µs (5×2 µs).

2.4.2. Simulation Analyses

The average radius of gyration (R_g) and end-to-end distance (R_{ee}) were obtained using the GROMACS tool `g_polystat`. To assess the convergence of the simulations, autocorrelation functions and block-error estimates of R_g and R_{ee} were computed using the GROMACS tool `g_analyze`. Principal component analysis (PCA) of the peptide backbone, by Campos and Baptista [27], and by using only the first two PCs, was also used for the evaluation of convergence and sampling. The minimal distance between periodic images in the simulations was monitored by the GROMACS tool `g_mindist` to ensure that the simulated peptide did not interact with its periodic images. Cluster analysis was done with the GROMACS tool `g_cluster`, using the GROMOS method [28], and was also used to obtain frames for representative structures. All peptide structures were visualised and rendered using PyMOL [21]. Distance matrices were obtained by using the GROMACS tool `g_mdmat`, and were used to create distance maps to show the distance between amino acid residues within the peptide. To create a contact map for the entire trajectory, MDTraj research software was used [29]. The secondary structure was analysed using the GROMACS tool `g_rama` and the DSSP program (version 2.2.1) [30], as well as DSSPPII analysis, that is, the DSSP program with modifications by Chebrek et al. [31] to include detection of the polyproline II (PPII) helix. Theoretical scattering intensities were obtained by using CRYSOL (version 2.8.2) [32].

3. Results and Discussion

The physicochemical characterisation of KEIF was performed in three parts by using a variety of methods. First, the primary structure was analysed to predict the overall structure and behaviour of the peptide. This was done by evaluating the charge, disorder propensity, and hydrophobicity/hydrophilicity per amino acid residue in the KEIF sequence. In addition, the sequence was compared with other protein sequences and sequence motifs from various protein databases. Second, the single chain was studied using far-UV CD spectroscopy, SAXS, and atomistic MD simulations. These results yielded basic structural properties, such as the average radius of gyration, maximal dimension ($D_{\max}$), and end-to-end distance, as well as distance-distribution functions ($P(r)$) and secondary-structure information. Third, interactions with neutral and anionic LUVs were investigated using DLS, LDV, cryo-TEM, and CD spectroscopy.

3.1. Primary-Structure Analysis

3.1.1. Charge-Distribution, Isoelectric-Point, and Das–Pappu Analysis

The estimated charge per amino acid at pH 7.0 is shown in Figure 2a, where the majority of the charged residues were located in the N-terminal half of the sequence. At this pH, the contribution from the histidines (pK_a 6.0) to the total charge was assumed to be negligible, giving a peptide net charge of (about) +3. The isoelectric point of KEIF was calculated to be 11.54 using the ExPASy [12] tool `Compute pI/Mw`. KEIF was predicted to belong to the R1 region of the Das–Pappu plot [33,34] in Figure 2b, which predicted that KEIF assumes a globular structure in aqueous solution.

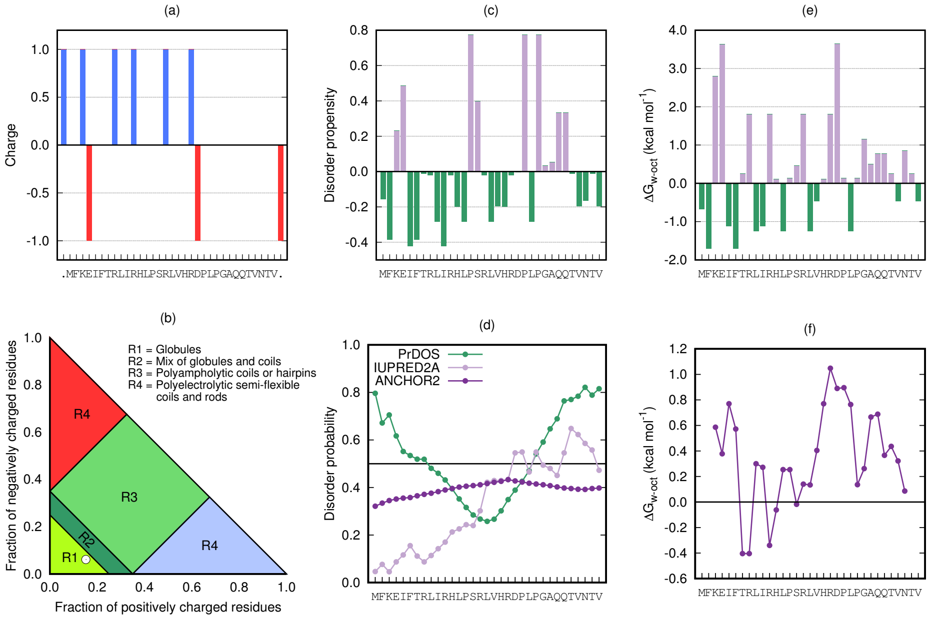


Figure 2. Primary structure analyses of KEIF: (**a**) Estimated charge per amino acid residue. Charge of peptide termini included as separate residues and marked as dots in the x-axis label. (**b**) Das–Pappu plot [33,34]. KEIF location is indicated by white circle in Region R1. (**c**) Disorder propensity per amino acid $(C_{\mathrm{DisProt}} - C_{\mathrm{PDB}})/C_{\mathrm{PDB}}$, as described by Uversky (2013) [35]. (**d**) Probability prediction of disordered regions and disordered binding regions using PrDOS (green) [36], IUPred2A (light purple), and ANCHOR2 (dark purple) algorithms [37]. (**e**) Whole-residue Wimley–White hydrophobicity indices [14,15] per amino acid residue. (**f**) Kyte–Doolittle [38] smoothed (five amino acid sliding-window) hydropathy plot based on whole-residue Wimley–White indices.

3.1.2. Disorder Propensity and Probability

Figure 2c shows the disorder propensity per amino acid based on fractional difference $(C_{\mathrm{DisProt}} - C_{\mathrm{PDB}})/C_{\mathrm{PDB}}$ as described by Uversky (2013) [35]. Overall, the sequence did not seem to contain a substantial amount of disorder-promoting residues, although a cluster of disorder-promoting residues was found closer to the C-terminal end of the sequence (residues Pro-24–Gln-28), suggesting that this part of the peptide has higher propensity for disordered conformations. The obtained prediction by the IUPred algorithm [37] in Figure 2d also suggested a low probability of disorder that increased sightly towards the C-terminus. This observation is, however, not supported by PrDOS analysis [36], which instead predicted disorder at both termini, and only a low probability of disorder in the central part of the sequence. The probability of disordered binding regions (by ANCHOR2 [37]) was found to be low across the entire amino acid sequence.

3.1.3. Distribution of Hydrophobic and Hydrophilic Amino Acids

The distribution of hydrophobic and hydrophilic amino acid residues in KEIF is depicted in Figure 2e, where the whole-residue Wimley–White hydrophobicity indices [14,15]—corresponding to the free energy ΔG of transfer from water to *n*-octanol—were taken as a measure of amino acid hydrophobicity/hydrophilicity. As revealed by the hydropathy plot shown in Figure 2f, obtained from Kyte–Doolittle sliding-window analysis [38], the peptide was overall (slightly) hydrophilic in character, and no transmembrane regions could be identified; with a typical bilayer thickness of 30 Å, an α-helical transmembrane segment would have to involve approximately twenty amino acids, and a

β-strand nine. The hydrophilic character returned from analysis suggested that KEIF does not reside in the transmembrane part of MgtA, but likely protrudes into the surrounding aqueous environment either intracellularly or extracellularly.

3.1.4. Sequence and Motif Alignment

The amino acid sequence of KEIF was compared with sequences from other proteins in the UniProtKB/Swiss-Prot database [39] with a protein-similarity search [40–43] to see if there were any similar sequences with known function. Pairwise sequence alignment of the top six results is displayed in Table 1. Excluding MgtA sequences, neither one of the resulting sequences had a high score, and the expected values indicated no biological significance. In addition, all matching sequences were found well within their corresponding proteins, which made any similarities of function unlikely. Smaller fragments of the charged part of the sequence (residues 3–21) were investigated in the same way, but yielded no different results. The full KEIF sequence was tested for containing any sequence motifs using ScanProsite [44] and MOTIF [45], but none was found.

Table 1. Pairwise sequence alignment, score, and expect value for top six results obtained from amino acid sequence similarity search of KEIF.

	Start Res. No.	Sequence	Stop Res. No.	Score (bits)	E-Value
KEIF	1	`MFKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV`	33	-	-
Query	1	`MFKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV`	33		
		`MFKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV`		70.5	1×10^{-15}
Sbjct 1	1	`MFKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV`	33		
Query	3	`KEIFTRLIRHLPSRLVHRDPLPGAQQTVN`	31		
		`+++F RL RHLP RLVHRDPLPGAQ  VN`		48.5	6×10^{-8}
Sbjct 2	7	`RQLFARLNRHLPYRLVHRDPLPGAQTAVN`	35		
Query	2	`FKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV`	33		
		`FKE+  +L+  L   ++HR+P P  Q   N V`		28.5	0.8
Sbjct 3	785	`FKEVEVQLLPELEEMILHRNPFPALQTLRNRV`	816		
Query	2	`FKEIFTRLIRHLPSRLVHRD`	21		
		`F+E+ T + RHLP  L H+D`		26.9	3.0
Sbjct 4	178	`FEEVDTNVTRHLPHELQHKD`	197		
Query	2	`FKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV`	33		
		`FKE+  +L+  L   ++HR+P P  Q   N V`		26.2	5.6
Sbjct 5	785	`FKEVEVQLLPELEEMILHRNPFPALQTLRNRV`	816		
Query	7	`TRLIRHLPSRLVHRDPLPG`	25		
		`TR++RH  +R + R+P PG`		25.8	7.7
Sbjct 6	129	`TRILRHAMTRHIFREPAPG`	147		

Sbjct 1 = `P0ABB8` Magnesium-transporting ATPase, P-type 1 *(Escherichia coli)*; Sbjct 2 = `P36640` Magnesium-transporting ATPase, P-type 1 *(Salmonella typhimurium)*; Sbjct 3 = `Q14667` Protein KIAA0100 *(Homo sapiens)*; Sbjct 4 = `Q758B8` GPI ethanolamine phosphate transferase 2 *(Ashbya gossypii)*; Sbjct 5 = `Q5SYL3` Protein KIAA0100 *(Mus musculus)*; Sbjct 6 = `C9K7C0` O-methyltransferase AMT9 *(t)*.

3.2. *Single Chain*

3.2.1. CD Spectroscopy

CD spectroscopy, a technique widely used to study the conformation of proteins in solution [46–48], was used with KEIF in order to obtain information about the peptide's secondary structure. CD spectra were recorded at 10 and 150 mM 1:1 salt (NaF), on the addition of Mg^{2+}, Ca^{2+} and Zn^{2+} cations in the form of chloride salts, as well as in organic solvent TFE (Figure 3a–c, respectively). In aqueous solution (TRIS buffer) and irrespective of salt concentration, the obtained CD spectra were characteristic of a disordered structure [46,47], and appeared to be completely insensitive to a 15-fold change in salt concentration (Figure 3a). The disordered structure is likely promoted by intrachain electrostatic repulsion caused by the relatively high density of positively charged amino acid residues. As expected, on the basis of the high similarity of the two spectra, the BeStSel [5,6] fitting of the two datasets returned highly similar secondary-structure elements where irregular (other) structures constituted the largest portion (see Table 2). The fits also pointed to a considerable fraction of β-strands, whereas helical structure elements were absent.

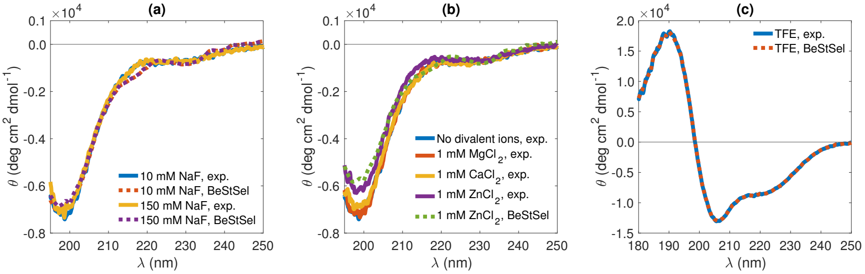


Figure 3. KEIF circular-dichroism (CD) spectra (solid lines) with BeStSel [5,6] fits (dashed lines), showing the effect of (**a**) varying salt (NaF) concentration (in TRIS buffer), (**b**) introducing various divalent cations (10 mM NaF in TRIS buffer), and (**c**) switching to organic solvent (TFE).

Table 2. Estimated secondary structure content in KEIF, returned from BeStSel [5,6] fitting of CD spectra in Figure 3.

	10 mM NaF (aq.)	**150 mM NaF (aq.)**	**1 mM ZnCl$_2$ (aq.)**	**TFE (org.)**
Fitted Range (nm)	**190–250**	**190–250**	**190–250**	**180–250**
Helix (%)	0.0	0.0	0.0	30.3
β-strand (%)	38.5	38.8	41.5	10.4
Turn (%)	14.9	14.9	14.8	15.9
Others * (%)	46.6	46.3	43.7	43.4

* 3_{10}-helix, π-helix, bends, β-bridge, and irregular/loop.

Whereas KEIF secondary structure appeared to be essentially insensitive to the presence of divalent Ca^{2+} and Mg^{2+} cations, again deduced from recorded CD spectra, the addition of Zn^{2+} ions served to make the minimum at around 200 nm somewhat less pronounced (Figure 3b); however, the effect on the corresponding structural elements returned from BeStSel [5,6] fitting is almost negligible. KEIF's apparent insensitivity to the presence of divalent cations was not surprising, considering that amino acids typically involved in metal ion co-ordination via their polar side-chain atoms—thiolate-carrying Cys (C), imidazole-carrying His (H), and carboxylate-carrying Glu (E) and Asp (D), collectively known as CHED [49]—are scarce. Moreover, the rather high density of cationic amino acid residues, as opposed to anionic ones, likely makes KEIF–cation interactions electrostatically unfavourable.

The situation was very different when KEIF is suspended in TFE (Figure 3c). In this organic solvent, as indicated by the development of a double minimum at 208 and 220 nm and a maximum at 192 nm [46,47], helical content considerably increases, mainly at the expense of the portion of β-strands (Table 2). Similar observations were made for the human-saliva protein histatin 5, which has disordered conformation in aqueous solution [50], but adopts a more helical conformation in TFE [51,52].

3.2.2. SAXS Measurements

Conformational information about the single chain of KEIF was obtained by performing SAXS experiments. The resulting form factor, Kratky plot, and distance-distribution function are depicted in Figure 4, in comparison to the EOM fit and obtained results from MD simulations. Figure 4a shows the obtained form factor, whose shape indicated natively unfolded behaviour. Further investigation of the data, in the form of the Kratky plot (Figure 4b), revealed the typical curve shape of a fully flexible and extended protein/peptide. The EOM fit conformed well with the experiment data ($\chi^2 = 1.143$).

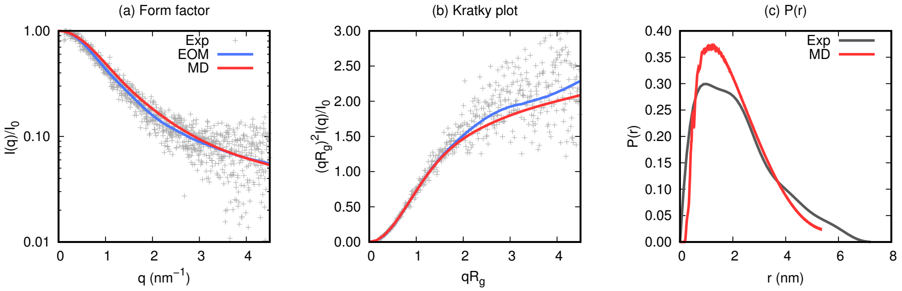


Figure 4. Experiment small-angle X-ray scattering (SAXS) results (grey) compared to ensemble-optimisation-method (EOM; blue) and molecular-dynamics (MD) simulations (red). (**a**) Form factors, (**b**) Kratky plot, and (**c**) distance-distribution functions.

Estimations of the radius of gyration were obtained using Guinier approximation (up to $qR_g \leq 0.8$), the $P(r)$ and the EOM. As shown in Table 3, Guinier approximation provides the smallest estimation, and $P(r)$ the largest, although the difference between the two was only 0.1 nm (5.5%). The estimation from the EOM was close to an average of the two values, and corresponded to deviations of only 2.2–3.3%. Estimations of the maximal dimension were also obtained from the $P(r)$ and the EOM, which are also shown in Table 3. A larger discrepancy of approximately 2 nm (33.3%) was found between the two estimated values.

Table 3. Ensemble averages of R_g and $D_{\max}$ (if applicable) as obtained from various methods.

	R_g (nm)	$D_{\max}$ (nm)
Guinier	1.76 ± 0.11	-
P(r) *	1.86	7.22
EOM *	1.80	5.16
MD	1.64 ± 0.05	-

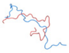
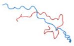
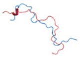
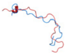
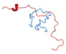
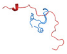
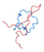
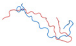
* No explicit errors were given using these analysis methods.

3.2.3. Atomistic Simulations

Atomistic MD simulations were performed to complement the experiment studies, and to obtain additional insight about the conformational properties of KEIF in bulk solution. Simulation convergence was assessed considering probability-distribution functions, autocorrelation functions, and block-average-error estimates of the radius of gyration and end-to-end distance (see Figures S3–S6). PCA was also utilised for this assessment (Figure S7). Discussion of the convergence is referred to the Supplementary Materials. To assess the validity of the simulations, simulation results were compared to the experiment results. Scattering curves were procured from the concatenated simulation trajectory by the use of CRYSOL (version 2.8.2) [32] and compared to the experiment SAXS curves and the curves from the EOM (see Figure 4). The curves were found to be very similar. The radius of gyration from the simulation was, however, found to be smaller than what was obtained from analysis of the experiment data (see Table 3), although the percentage difference was only 7.1–12.6%. Because of the good correspondence with the experiment SAXS results, the simulated data were considered to be sufficiently valid to be used as accurate single-chain representation.

Cluster analysis was performed on the concatenated MD simulation trajectory to obtain representative structures. Eight clusters were found with an RMSD cutoff of 0.99 Å, and the top six (99.75%) were compared to the six structures that were obtained from EOM analysis in Table 4. A large majority of the MD structures were found in the first two clusters at this cutoff. However, if using an RMSD cutoff of 0.70 Å or 0.50 Å, cluster sizes became of more equal size, and the top eight clusters summed up to 58.72% and 16.74%, respectively. For more thorough analysis of the structures, distance maps showing the distance between amino acid residues in the representative structures were created (see Figure 5). By studying these maps, details otherwise unnoticed were found. For example, evidence of cation–π interactions was observed between Phe-6 and (i) Arg-20 in the MD 3 structure (see Figure 6), (ii) Gln-27 in the MD 4 structure, and (iii) Lys-3 in the MD 5 structure. The remaining close distances seemed to arise due to hydrogen bonds and dispersion interactions, although a few electrostatic interactions were also observed. A contact map, instead showing the probability of contacts within a cutoff of 4.0 Å throughout the entire concatenated simulation, is presented in Figure 7. Here, the most probable contact was found between Leu-23 and Gln-27. Other notable contacts were found between residues Leu-13 and Arg-16, Arg-16 and Val-30, as well as between Leu-17 and Arg-20.

Table 4. Comparisons between representative structures from cluster analysis of MD simulations (red) and structures obtained from EOM analysis of experiment SAXS data (blue). The percentage of all structures that belonged to each cluster is given in the parentheses; MD clusters summed up to 99.75%, and EOM structures summed up to ∼100%. RMSD values (Å) of aligned atoms given below each comparison.

	EOM 1 (∼30%)	**EOM 2 (∼30%)**	**EOM 3 (∼10%)**	**EOM 4 (∼10%)**	**EOM 5 (∼10%)**	**EOM 6 (∼10%)**
MD 1 (61.72%)	 12.84	 10.88	 11.32	 12.92	 10.58	 10.45
MD 2 (24.44%)	 7.15	 7.80	 5.81	 7.95	 16.36	 16.33
MD 3 (8.35%)	 11.56	 12.76	 13.45	 15.00	 8.74	 9.25
MD 4 (3.05%)	 15.98	 16.62	 16.69	 19.10	 10.09	 10.28
MD 5 (1.40%)	 9.19	 7.61	 6.44	 7.26	 15.96	 16.26
MD 6 (0.80%)	 16.57	 15.78	 17.16	 18.21	 5.92	 7.28

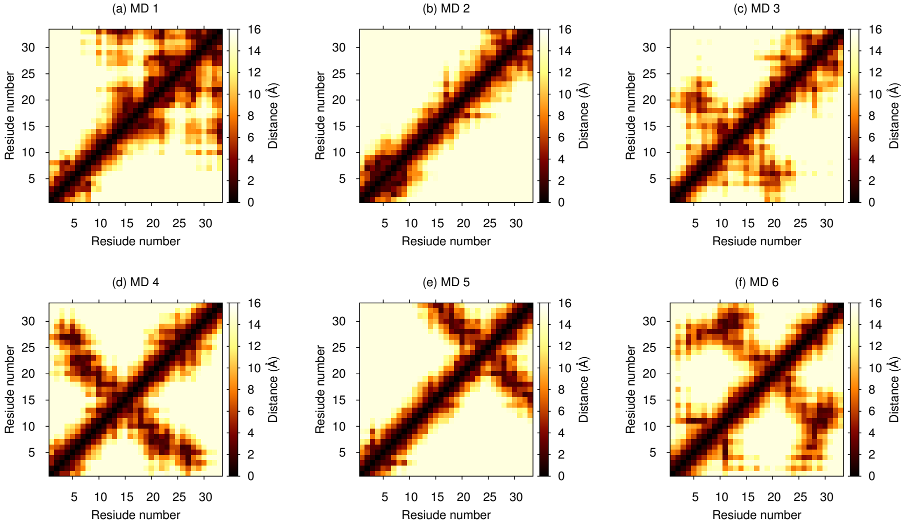


Figure 5. Distance maps depicting distance between amino acid residues for each representative structure of top six clusters from MD simulation.

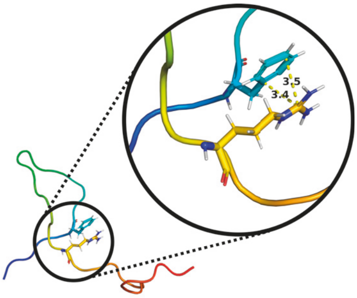


Figure 6. Illustration of cation–π interaction between Phe-6 and Arg-20 in MD 3 structure. marked distances given in Ångström (Å).

The secondary structure per amino acid of KEIF from the MD simulation was analysed using the DSSP algorithm, and is visualised in Figure 8a. Although most of the structure was dominated by coils and bends, a few residues also showed propensity for turns and β-structures. The helical content was found to be negligible. Unfortunately, this analysis did not include the PPII helix. To account for PPII helices, DSSPPII analysis was utilised on the representative structures from the top six clusters of the MD simulations (see Table 5). While the N-terminal half of the first structure was dominated by random coil conformation, a more local order was found towards the C-terminus as distinguished turns around a small PPII helix at Asp-21–Pro-22–Leu-23, followed by an isolated β-bridge between Thr-29 and Thr-32. The small PPII helix around residues 21–24 was conserved in the top three cluster structures, although PPII helices were present in all structures. Particularly, the fourth structure seemed to have strong PPII propensity. A small 3_{10}-helix was found at residues 4–6 in the second structure, whereas the sixth structure contained evidence of β-sheet formation. These results did not contradict what was observed by CD spectroscopy (Figure 3). The presence of PPII in the conformational ensemble of KEIF was also in line with what was seen in the Kratky plot (Figure 4), that is, mainly flexible but extended conformations. A Ramachandran plot (Figure 8b) was also produced from the simulated results that showed a high count in the region of $(\phi, \psi) = (-75°, +145°)$, which also supported a significant PPII content. The plot also shows a fairly high count of β-structures, but only little α-helical content, which corroborated the CD spectroscopy results (Table 2).

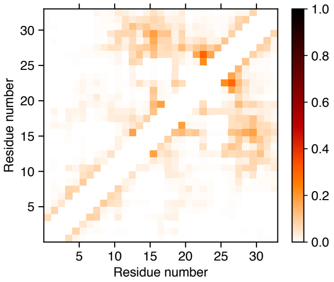


Figure 7. Contact map showing the probability of amino acid residues being closer to each other than cutoff of 4.0 Å. Darker colour indicates higher probability, and white corresponds to zero probability. Residue interactions with themselves, as well as two neighbouring residues on each side, were excluded from analysis.

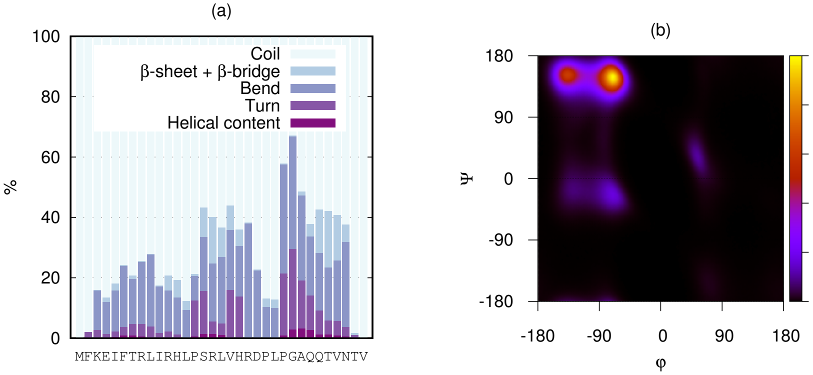


Figure 8. (**a**) Stacked secondary-structure histograms per amino acid residue of KEIF as obtained from MD simulations. Three different types of helices included in helical content: (i) α-helix, (ii) 3_{10}-helix, and (iii) π-helix. This analysis did not include the PPII helix. (**b**) Ramachandran plot of KEIF as obtained from MD simulations.

Table 5. Secondary structure per amino acid of representative structures from top six clusters of MD simulation of KEIF, as obtained by using DSSPPII analysis [31]. Secondary structure per amino acid represented according to standard DSSP classification.

#	MFKEIFTRLIRHLPSRLVHRDPLPGAQQTVNTV
1	----SS----S----TTTS-PPPTTTT-BTTB-
2	---GGGTSPP--------S--SPP-S--SS---
3	-----------SS---SPP-PPPTT--SS----
4	---PPPBPP----TTS-----B-TTTTPPPP--
5	------PP------SS--B-PP-TT--SB----
6	----PP-SS-EE-PPPP--SS-----SSEE---

3.3. Interactions with Vesicles

The interaction of KEIF with lipid bilayers, provided by neutral (POPC) and anionic (POPC:POPS) LUVs, was characterised by partitioning-free-energy calculations, DLS, LDV, cryo-TEM, and CD spectroscopy, as outlined in the following subsections. The anionic bilayer may be regarded as a model of the bacterial-cell membrane, while the neutral one was used for comparison to elucidate the effect of membrane charge and the importance of electrostatic interactions.

3.3.1. Partitioning Free Energies

The hydrophilic/hydrophobic character of a peptide or a protein naturally influences its interactions with a bilayer. In Section 3.1.3 it was concluded that KEIF has an overall (slightly) hydrophilic character, and should consequently prefer bulk water rather than the hydrophobic interior of a bilayer. To support this hypothesis, the `MPEx` tool [13] was used to calculate the free energy ΔG of partitioning of KEIF from bulk water into a bilayer. By applying the `MPEx` tool to KEIF, water-to-bilayer partitioning free energy of $+18.21\,\mathrm{kcal\,mol^{-1}}$ was obtained, signifying the unfavourableness of water-to-bilayer partitioning. Even if partitioning would present an opportunity to reduce the free energy by a partitioning–folding coupling mechanism—corresponding to approximately $0.4\,\mathrm{kcal\,mol^{-1}}$ per amino acid residue [53]—free energy would still be positive ($+5.01\,\mathrm{kcal\,mol^{-1}}$), and water-to-bilayer partitioning thus unflavoured. KEIF–vesicle interactions discussed in the following subsections are instead attributed to electrostatic interactions between net positive KEIF and net negative POPC:POPS vesicles.

3.3.2. DLS and LDV

Zwitterionic lipid POPC (carrying one positively charged and one negatively charged functional group; net charge ±0) and anionic lipid POPS (carrying one positively charged and two negatively charged functional groups; net charge -1) were used for the preparation of LUVs. The vesicles were prepared by extrusion, resulting in monomodal size distributions (polydispersity index, $PdI < 0.1$) with diameters D centred at 108.0 nm and 102.9 nm for the neutral POPC and anionic 3:1 POPC:POPS vesicles, respectively, as measured by DLS (correlation functions are shown in Figure S8). LDV, in turn, was used to determine vesicle electrophoretic mobility, μ. Whereas a considerable net negative mobility was measured for the POPC:POPS vesicles due to the anionic nature of POPS, the mobility of the POPC vesicles could not be accurately determined due to their extremely weak net charge; the value given for the POPC vesicle mobility has to thus be taken with a grain of salt. Vesicle-size and electrophoretic-mobility data are summarised in Table 6.

Initially, KEIF-vesicle interactions were probed by by DLS. DLS is highly sensitive to changes in particle size, as scattered intensity I scales with the sixth power of particle radius r ($I \propto r^6$). The day after KEIF addition ($\sim$18 h), vesicle-size distributions were still monomodal (correlation functions are shown in Figure S8), reflecting the absence of vesicle aggregation. Whereas the size of neutral POPC vesicles slightly increased upon addition of KEIF (Table 6)—possibly an indication of fusion of a small number of vesicles—the size of anionic 3:1 POPC:POPS ones was, instead, somewhat reduced. We hypothesise that cationic KEIF electrostatically adsorbs to the surface of the anionic vesicles to neutralise some of the negative charges, thereby reducing the lateral head-group repulsion and allowing lipids to pack closer.

Second, electrophoretic-mobility measurements were performed to study the adsorption of KEIF to vesicle surface. Whereas the mobility of the neutral POPC vesicles became only somewhat more positive upon addition of positively charged KEIF (remember, however, that these low mobility values are not too accurate), the mobility of the anionic 3:1 POPC:POPS vesicles became significantly more positive (Table 6). On the basis of these measurements, we hypothesise that KEIF, only to a (very) small extent, adsorbs to the POPC vesicles; KEIF can possibly access the negative charge on a small number of the POPC head groups. The apparent high affinity of KEIF for the anionic POPC:POPS vesicles signifies that a net-negative vesicle charge is important for adsorption, with charge neutralisation and concomitant entropy gain resulting from counter ion release being the main driving force for adsorption. The importance of charges and electrostatic interactions was also reported elsewhere [54]. Moreover, the fact that mobility—a measure of surface charge—becomes significantly more positive upon KEIF addition to the anionic 3:1 POPC:POPS vesicles supports the hypothesis that KEIF adsorbs to the vesicle surface and does not partition into the hydrophobic interior of the lipid bilayer. That into-bilayer partitioning is unfavourable stems from the hydrophilic nature of KEIF (Figure 2f), which is reflected in the positive partitioning free energies presented in Section 3.3.1.

Table 6. Z-average diameter, polydispersity index, and electrophoretic mobility of POPC and 3:1 POPC:POPS vesicles in absence and presence of KEIF, at 10 mM NaF in TRIS buffer. Lipid:KEIF molar ratio was 16:1.

	D **(nM)**	**PdI**	μ **(10^{-8} m^2 V^{-1} s^{-1})**
POPC vesicles	108.0 ± 0.7	0.09 ± 0.02	−0.12 ± 0.03 *
POPC vesicles + KEIF	111.1 ± 0.6	0.06 ± 0.01	+0.16 ± 0.03 *
3:1 POPC:POPS vesicles	102.9 ± 0.7	0.06 ± 0.02	−3.99 ± 0.61
3:1 POPC:POPS vesicles + KEIF	95.7 ± 0.3	0.07 ± 0.02	−2.02 ± 0.40

* These values should be taken with a grain of salt, as measurement accuracy was low due to the very weak charge.

3.3.3. Cryo-TEM

In order to obtain complementary information to that provided by DLS, POPC and 3:1 POPC:POPS vesicles were imaged by cryo-TEM [55], in the absence and presence of KEIF (Figure 9; additional images are shown in Figures S9 and S10). Whereas the majority of the anionic 3:1 POPC:POPC vesicles were unilamellar, multilamellar vesicles were frequently observed in the case of the neutral POPC vesicles. Vesicles with one or two smaller-sized internalised vesicles were observed in both cases, but were more common with POPC. In both cases, the degree of polydispersity appeared to be larger than that indicated by DLS (Table 6), which was supported by the larger standard deviation returned from cryo-TEM image analysis of vesicle size (Figure S11).

The addition of KEIF to the neutral POPC vesicles had a dramatic effect on vesicle stability, as evidenced by a severe degree of polydispersity—likely caused by fusion of the original, small-sized vesicles—and the presence of a large number of ruptured vesicles and free lipid bilayers. This state did not alter the macroscopic appearance of the sample, nor was it picked up by the DLS measurements, which still indicated a low polydispersity index (Table 6). We hypothesise that extremely weakly charged POPC vesicles, on addition of KEIF polyelectrolytes, become electrostatically destabilised, and thus prone to rupture and fusion. In contrast to POPC vesicles, the anionic 3:1 POPC:POPS vesicles seemed to be unaffected by the addition of KEIF, as these vesicles remained intact. Corroborating the observations made by DLS, a smaller size was measured for the POPC:POPS vesicles in the presence of KEIF (Figure S11). As previously discussed, adsorption of KEIF to the vesicles likely reduces lipid head-group repulsion, thus enabling lipids to pack closer.

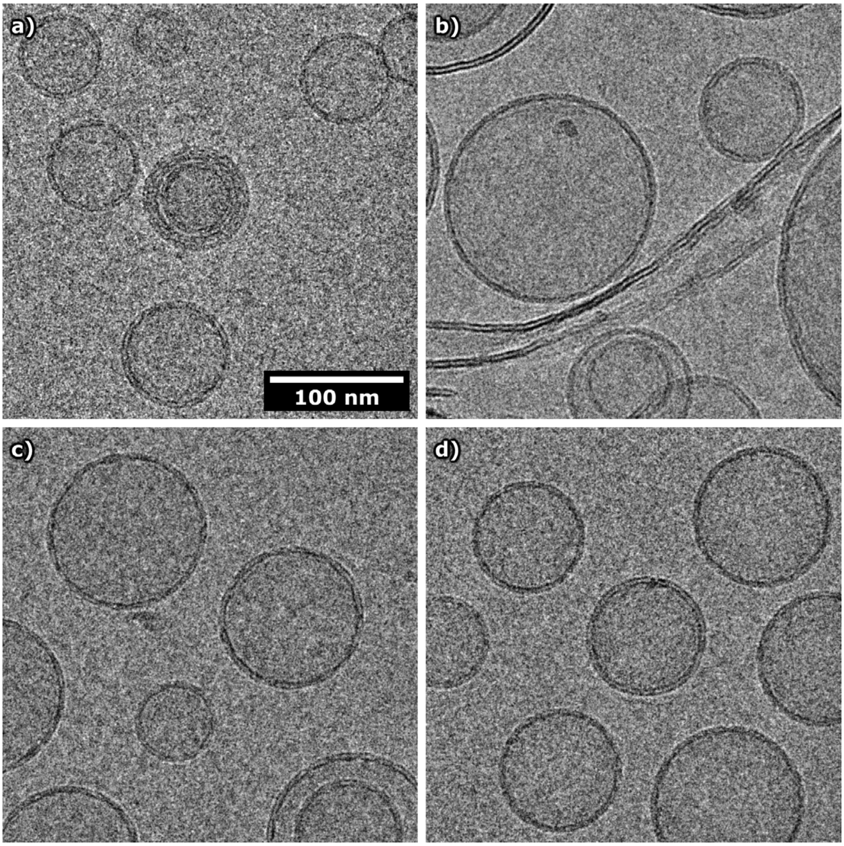


Figure 9. Representative cryo-TEM images of (**a**,**b**) POPC and (**c**,**d**) 3:1 POPC:POPS vesicles in the (**a**,**c**) absence and (**b**,**d**) presence of KEIF, at 10 mM NaF in TRIS buffer. The lipid:KEIF molar ratio was 16:1. The scale bar applies to all images.

3.3.4. CD Spectroscopy

CD spectroscopy is a useful tool for monitoring conformational changes in membrane-active proteins and peptides. In such studies, as model membranes, small unilamellar vesicles (SUVs) made by sonication have been used almost exclusively, as the common belief has been that light scattering associated with LUVs is unacceptably high. It was shown, however, that undistorted CD spectra can be obtained at wavelengths above 200 nm in the presence of up to 3 mM LUVs [56]. By choosing LUVs over SUVs, vesicle-curvature effects that may cause anomalous peptide partitioning are reduced, and, as LUVs are thermodynamically stable structures (as different from SUVs which are only metastable), equilibrium thermodynamic measurements can be performed.

Using CD spectroscopy, the possible induction of secondary-structure elements in KEIF upon adsorption onto the aforementioned neutral POPC and anionic 3:1 POPC:POPS LUVs was studied. Obtained CD spectra, recorded the day after KEIF addition to the vesicles, showed clear differences between POPC and POPC:POPS vesicles (Figure 10). First, the spectrum recorded in the presence of neutral POPC vesicles was highly similar to that recorded in the absence of vesicles, and BeStSel [5,6] fitting returned similar amounts of various secondary structural elements (Table 7). This means that either adsorption to the POPC vesicles does not induce any conformational changes or, more likely and supported by obtained results form DLS and LDV (Section 3.3.2), KEIF simply does not posses any electrostatic affinity for the POPC vesicles and consequently does not adsorb.

The spectrum recorded in the presence of anionic POPC:POPS vesicles (Figure 10) is clearly different from that recorded in the presence of neutral POPC vesicles (or in the absence of vesicles), which again highlights the importance of electrostatic interactions to drive the adsorption of KEIF to the vesicles. Changes to the CD spectrum induced by the POPC:POPS vesicles suggest an increase in the secondary-structure content of KEIF, from a largely disordered conformation in solution to more ordered conformation when absorbed to the vesicles. The spectrum showed the same characteristic features as the spectrum recorded in TFE, with a double minimum at 208 and 220 nm that is indicative of helical structure [46,47]. Indeed, BeStSel [5,6] fitting of the spectrum showed an increase in helical content (see Table 7).

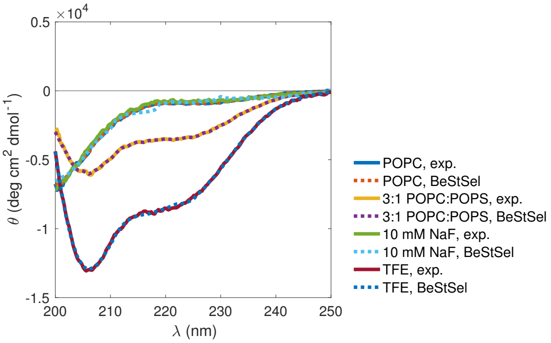


Figure 10. KEIF CD spectra recorded in presence of POPC and 3:1 POPC:POPS vesicles, in TRIS buffer supplemented with 10 mM NaF. Lipid:KEIF molar ratio was 16:1. For comparison, spectra recorded in TRIS buffer supplemented with 10 mM NaF and in TFE are also shown. Dashed lines represent BeStSel [5,6] fits.

Table 7. Estimated secondary-structure content in KEIF, returned from BeStSel [5,6] fitting of CD spectra in Figure 10.

	POPC (aq.)	**3:1 POPC:POPS (aq.)**	**10 mM NaF (aq.)**	**TFE (org.)**
Fitted Range (nm)	**200–250**	**200–250**	**200–250**	**200–250**
Helix (%)	1.8	9.9	0	20.5
β-strand (%)	31.3	27.6	31.0	16.6
Turn (%)	17.7	16.0	17.7	15.2
Others * (%)	49.4	46.5	51.3	47.7

* 3_{10}-helix, π-helix, bends, β-bridge, and irregular/loop.

3.4. Summary of, and Correlations between, Main Results

Primary-structure analysis predicted a rather globular conformation of KEIF, with larger disorder propensity towards the C-terminus of the peptide. No disordered binding regions were predicted, nor were any sequence motifs or transmembrane regions found. Both CD and SAXS experiments showed that KEIF is indeed a disordered peptide in aqueous solution, in agreement with Subramani's DISOPRED3-based prediction [3]. These results were also supported by MD simulations, for which the scattering curve was in good agreement with experiment SAXS data (Figure 4). Contrary to predictions, SAXS results suggested that KEIF is fully flexible and extended instead of globular. This was also supported by the simulations, where the radius of gyration was found to be similar (within 7.1–12.6%; Table 3), and the largest conformation clusters were represented by fairly extended structures (Table 4). Thus, solely relying on net charge and the fraction of charged residues for predicting the conformation and shape of the peptide seemed to be insufficient in this particular case, and implied that the location of the charges in the sequence is of significance. With all of the positive charges evenly distributed in the N-terminal half of the sequence, it is not unexpected that electrostatic repulsion could cause the extension of this part of the peptide chain.

Analysis of the CD spectroscopy results suggested a little amount of β-structures in the single chain in aqueous solution (Table 2), which was also observed in the DSSP analysis and in the Ramachandran plot of the MD simulations (Figure 8). By performing DSSPPII analysis on the simulated peptide, PPII structure was also revealed to be an important secondary-structure component, mainly found towards the C-terminal end of the peptide (Table 5). This discovery cannot be disregarded by the experiment results, since the PPII helix was not included in theBeStSel analysis. Furthermore, CD spectroscopy of KEIF in organic solvent TFE induced a considerable increase of helical content.

Partitioning-free-energy analysis deemed water-to-bilayer partitioning of KEIF to be unfavourable. DLS and LDV results seemed to agree with this prediction, and provided data suggesting that KEIF instead has the ability to adsorb to the surface of anionic POPC:POPS vesicles (Table 6). In addition, CD spectroscopy revealed a conformational change for KEIF with this type of membrane, inducing an increase in helical content (Figure 10 and Table 7), similarly to what was observed with KEIF suspended in TFE, but to a smaller extent. We hypothesise that these observations are attributed to the electrostatic adsorption of cationic KEIF to the anionic vesicle surface, driven by charge neutralisation and a concomitant release of counterions. This would also explain the apparent low propensity of KEIF to adsorb to neutral POPC vesicles.

4. Conclusions

This paper featured the extensive physicochemical characterisation of KEIF, the N-terminal disordered region of MgtA, using an approach combining various experimental techniques and MD simulations. Both the experimental techniques and the complementary simulations confirmed that KEIF is an extended intrinsically disordered peptide with little propensity towards β-structures, and possibly PPII structure. In addition, experiments showed that KEIF adopts a more helical structure in organic solvent TFE. Further experimental investigation of interactions between KEIF and vesicles proved that it is unlikely for KEIF to traverse the bilayer, and that it instead seems to adsorb to the surface of anionic vesicles. Because of opposite charges of the vesicles and the peptide, it is hypothesised that the interaction is electrostatically driven. KEIF adsorption to the vesicle surface also implies a release of counterions, which is entropically favourable and constitutes an additional driving force for adsorption.

By performing this study, we provided comprehensive insight to the structure–function relationship of KEIF that, in turn, might aid in providing a more holistic understanding of the function of MgtA. Because of the observed interactions between KEIF and anionic lipid bilayers, it is reasonable to believe that this intrinsically disordered region of MgtA actually has an important function in the biological context. Considering KEIF's hydrophilic character, together with its electrostatic affinity for the surface of anionic lipid bilayers (the bacterial cell membrane being a prime example), its role might be to anchor the large MgtA protein in the bilayer, as schematically depicted in Figure 11. Whereas additional research, beyond the scope of this paper, is required in order to elucidate the exact role of KEIF, it does seem likely that KEIF is more than just a mere "appendix" of MgtA.

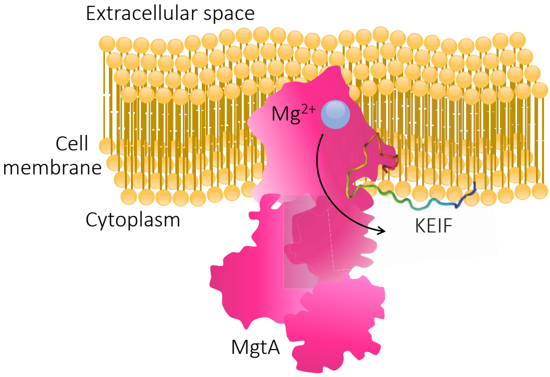


Figure 11. Schematic cross-section of an MgtA-carrying cell membrane. It is possible that KEIF plays the role of an anchor, helping to stabilise the large protein complex in the membrane by locking it in place via electrostatic interactions with anionic lipid head groups.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/10/4/623/s1. Figure S1: Residuals from BeStSel fitting of CD spectra 1; Figure S2: Residuals from BeStSel fitting of CD spectra 2; Figure S3: MD probability-distribution functions; Figure S4: MD autocorrelation functions; Figure S5: MD block-average-error estimates; Figure S6: MD convergence; Figure S7: PCA free-energy landscapes; Figure S8: DLS autocorrelation functions; Figure S9: Cryo-TEM images 1; Figure S10: Cryo-TEM images 2; Figure S11: Cryo-TEM size distributions.

Author Contributions: conceptualisation, S.J., L.K.M., and M.S.; methodology, S.J., L.K.M., and D.B.; validation, S.J., L.K.M., and M.S.; formal analysis, S.J., L.K.M., and D.B.; investigation, S.J., L.K.M., and D.B.; resources, M.S.; data curation, S.J., L.K.M., D.B., and M.S.; writing—original-draft preparation, S.J., L.K.M., and M.S.; writing—review and editing, S.J., L.K.M., D.B., J.P.M., and M.S.; visualisation, S.J., L.K.M., and D.B.; supervision, M.S.; project administration, M.S.; funding acquisition, J.P.M. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NordForsk Neutron Science Program (grant number 81912) and by the Science Faculty project grant program for research on neutrons and synchrotron light (Lund University Strategic funds for MAX-IV and European Spallation Source).

Acknowledgments: The authors thank the European Synchrotron Radiation Facility (ESRF) for providing beamtime and Bart van Laer at ESRF for providing assistance in using beamline BM29. Simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC). Cryo-TEM imaging was performed at the National Center for High Resolution Electron Microscopy (nCHREM) at Lund University. D.B. acknowledges the financial support from the Crafoord Foundation, and thanks Anna Carnerup for the generous technical assistance.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Reinhart, R.A. Magnesium metabolism: A review with special reference to the relationship between intracellular content and serum levels. *Arch. Intern. Med.* **1988**, *148*, 2415–2420. [CrossRef]
2. Maguire, M.E. Magnesium transporters: Properties, regulation and structure. *Front. Biosci.* **2006**, *11*, 3149–3163. [CrossRef] [PubMed]
3. Subramani, S.; Perdreau-Dahl, H.; Morth, J.P. The magnesium transporter A is activated by cardiolipin and is highly sensitive to free magnesium in vitro. *eLife* **2016**, *5*, e11407. [CrossRef] [PubMed]
4. Jones, D.T.; Cozzetto, D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863. [CrossRef] [PubMed]
5. Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.H.; Goto, Y.; Réfrégiers, M.; Kardos, J. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3095–E3103. [CrossRef] [PubMed]

6. Micsonai, A.; Wien, F.; Bulyáki, É.; Kun, J.; Moussong, É.; Lee, Y.H.; Goto, Y.; Réfrégiers, M.; Kardos, J. BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **2018**, *46*, W315–W322. [CrossRef]
7. Franke, D.; Petoukhov, M.; Konarev, P.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.; Kikhney, A.; Hajizadeh, N.; Franklin, J.; Jeffries, C.; et al. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Cryst.* **2017**, *50*, 1212–1225. [CrossRef]
8. 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*, 5656–5664. [CrossRef]
9. Tria, G.; Mertens, H.D.; Kachala, M.; Svergun, D.I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2015**, *2*, 207–217. [CrossRef]
10. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **2005**, *152*, 36–51. [CrossRef]
11. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef] [PubMed]
12. 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*, 3784–3788. [CrossRef] [PubMed]
13. Snider, C.; Jayasinghe, S.; Hristova, K.; White, S.H. MPEx: A tool for exploring membrane proteins. *Protein Sci.* **2009**, *18*, 2624–2628. [CrossRef]
14. Wimley, W.C.; Creamer, T.P.; White, S.H. Solvation energies of amino acid side chains and backbone in a family of host- guest pentapeptides. *Biochemistry* **1996**, *35*, 5109–5124. [CrossRef]
15. White, S.H.; Wimley, W.C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365. [CrossRef]
16. Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. [CrossRef]
17. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef]
18. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [CrossRef]
19. 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. Bioinf.* **2010**, *78*, 1950–1958. [CrossRef]
20. Piana, S.; Donchev, A.G.; Robustelli, P.; Shaw, D.E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113–5123. [CrossRef]
21. Schrödinger LLC. *The PyMOL Molecular Graphics System*; Version 1.2r1; 2009. Available online: https://pymol.org/2/support.html?#citing (accessed on 16 April 2020).
22. Berendsen, H.; Van Gunsteren, W. Practical algorithms for dynamic simulations. In *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*; North-Holland: Amsterdam, The Netherlands, 1986; pp. 43–65.
23. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089. [CrossRef]
24. Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [CrossRef]
25. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [CrossRef] [PubMed]
26. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182. [CrossRef]
27. Campos, S.R.; Baptista, A.M. Conformational analysis in a multidimensional energy landscape: Study of an arginylglutamate repeat. *J. Phys. Chem. B* **2009**, *113*, 15989–16001. [CrossRef] [PubMed]
28. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W.F.; Mark, A.E. Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240. [CrossRef]
29. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.P.; Lane, T.J.; Pande, V.S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. doi:10.1016/j.bpj.2015.08.015. [CrossRef]

30. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. [CrossRef]
31. Chebrek, R.; Leonard, S.; de Brevern, A.G.; Gelly, J.C. PolyprOnline: Polyproline helix II and secondary structure assignment database. *Database* **2014**, *2014*. [CrossRef]
32. Svergun, D.; Barberato, C.; Koch, M.H. CRYSOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **1995**, *28*, 768–773. [CrossRef]
33. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13392–13397. [CrossRef] [PubMed]
34. Das, R.K.; Ruff, K.M.; Pappu, R.V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112. [CrossRef] [PubMed]
35. Uversky, V.N. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord. Proteins* **2013**, *1*, e24684. [CrossRef] [PubMed]
36. Ishida, T.; Kinoshita, K. PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460–W464. [CrossRef] [PubMed]
37. Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [CrossRef]
38. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [CrossRef]
39. Consortium, U. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]
40. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]
41. Schäffer, A.A.; Aravind, L.; Madden, T.L.; Shavirin, S.; Spouge, J.L.; Wolf, Y.I.; Koonin, E.V.; Altschul, S.F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **2001**, *29*, 2994–3005. [CrossRef]
42. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinf.* **2009**, *10*, 421. [CrossRef]
43. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [CrossRef] [PubMed]
44. De Castro, E.; Sigrist, C.J.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **2006**, *34*, W362–W365. [CrossRef]
45. Kanehisa, M. Linking databases and organisms: GenomeNet resources in Japan. *Trends Biochem. Sci.* **1997**, *22*, 442–444. [CrossRef]
46. Kelly, S.M.; Price, N.C. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **2000**, *1*, 349–384. [CrossRef] [PubMed]
47. Kelly, S.M.; Jess, T.J.; Price, N.C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta Proteins Proteom.* **2005**, *1751*, 119–139. [CrossRef]
48. Greenfield, N.J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **2006**, *1*, 2876. [CrossRef]
49. Babor, M.; Gerzon, S.; Raveh, B.; Sobolev, V.; Edelman, M. Prediction of transition metal-binding sites from apo protein structures. *Proteins Struct. Funct. Bioinf.* **2008**, *70*, 208–217. [CrossRef]
50. Brewer, D.; Hunter, H.; Lajoie, G. NMR studies of the antimicrobial salivary peptides histatin 3 and histatin 5 in aqueous and nonaqueous solutions. *Biochem. Cell Biol.* **1998**, *76*, 247–256. [CrossRef]
51. Melino, S.; Rufini, S.; Sette, M.; Morero, R.; Grottesi, A.; Paci, M.; Petruzzelli, R. Zn^{2+} ions selectively induce antimicrobial salivary peptide histatin-5 to fuse negatively charged vesicles. Identification and characterization of a zinc-binding motif present in the functional domain. *Biochemistry* **1999**, *38*, 9626–9633. [CrossRef]

52. Jephthah, S.; Henriques, J.; Cragnell, C.; Puri, S.; Edgerton, M.; Skepö, M. Structural Characterization of Histatin 5–Spermidine Conjugates: A Combined Experimental and Theoretical Study. *J. Chem. Inf. Model.* **2017**, *57*, 1330–1341. [CrossRef]
53. Wimley, W.C.; White, S.H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **1996**, *3*, 842–848. [CrossRef] [PubMed]
54. Nygren, P.; Lundqvist, M.; Liedberg, B.; Jonsson, B.H.; Ederth, T. Secondary structure in de novo designed peptides induced by electrostatic interaction with a lipid bilayer membrane. *Langmuir* **2010**, *26*, 6437–6448. [CrossRef] [PubMed]
55. Patterson, J.P.; Xu, Y.; Moradi, M.A.; Sommerdijk, N.A.; Friedrich, H. CryoTEM as an advanced analytical tool for materials chemists. *Acc. Chem. Res.* **2017**, *50*, 1495–1501. [CrossRef] [PubMed]
56. Ladokhin, A.S.; Fernández-Vidal, M.; White, S.H. CD spectroscopy of peptides and proteins bound to large unilamellar vesicles. *J. Membr. Biol.* **2010**, *236*, 247–253. [CrossRef]

Supplementary Materials: Physicochemical Characterisation of KEIF – The Intrinsically Disordered N-terminal Region of Magnesium Transporter A

Stéphanie Jephthah [1], Linda K. Månsson [1], Domagoj Belić [2], Jens Preben Morth [3] and Marie Skepö [1,4]*

[1] Division of Theoretical Chemistry, Department of Chemistry, Lund University, Naturvetarvägen 14, 221 00 Lund, Sweden
[2] Division of Physical Chemistry, Department of Chemistry, Lund University, Naturvetarvägen 14, 221 00 Lund, Sweden
[3] Enzyme and Protein Chemistry, Section for Protein Chemistry and Enzyme Technology, Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, 2800, Kgs. Lyngby, Denmark
[4] Lund Institute of Advanced Neutron and X-ray Science (LINXS), Scheelevägen 19, 233 70, Lund, Sweden
* Correspondence: marie.skepo@teokem.lu.se; Tel.: +46-46-222 33 66

1. Residuals from BeStSel fitting of CD spectra

Some of the obtained CD spectra were subject to BeStSel [1,2] fitting to access KEIF's corresponding secondary structure elements. Fitted residuals are given in Figure S1 and Figure S2, and represent the goodness of the fittings that are graphically displayed in Figure 3 and Figure 10, respectively.

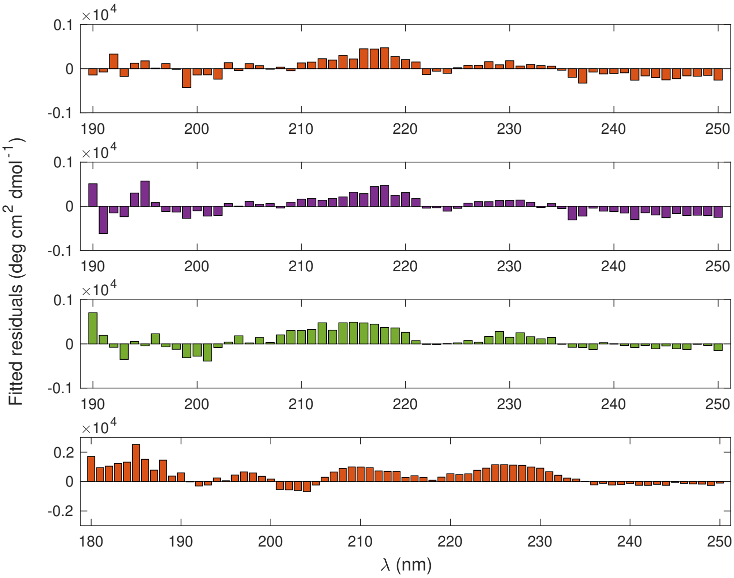


Figure S1. Fitted residuals for spectra recorded in the presence of (from top to bottom) 10 mM NaF (aq.), 150 mM NaF (aq.), 1 mM $ZnCl_2$ (aq.), and TFE (org.). Note that the x- and y-axis ranges vary between the subplots. The fits are graphically presented in Figure 3 in the paper.

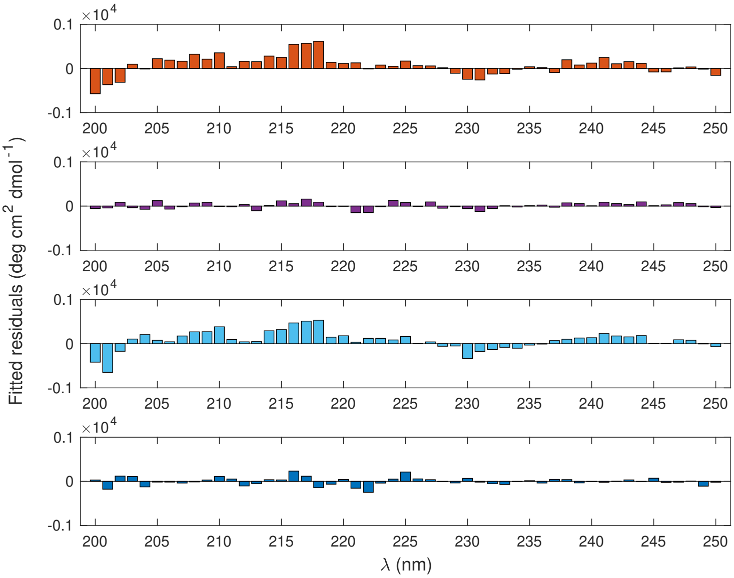


Figure S2. Fitted residuals for spectra recorded in the presence of (from top to bottom) POPC vesicles (aq.), 3:1 POPC:POPS vesicles (aq.), 10 mM NaF (aq.), and TFE (org.). The fits are graphically presented in Figure 10 in the paper.

2. Molecular dynamics simulations

Simulation convergence was assessed by investigating probability distribution functions, auto-correlation functions, and block average error estimates of the radius of gyration and the end-to-end distance, see Figure S3-S6. Although the probability distribution functions in Figure S3 are not identical, they still cover the same region, which is a good indication of a converged system. The auto-correlation functions of the radius of gyration and the end-to-end distance are expected to decrease rapidly and vary around zero for a converged system. In Figure S4 the correlation of all replicates is seen to reach zero within 0.3 µs for the radius of gyration, and within 0.2 µs for the end-to-end distance. Although both properties remain varying around zero, the end-to-end distance seem to be more converged than the radius of gyration. Especially the second replicate (red) seem to have more difficulty of converging. However, when consulting the block average error estimates in Figure S5, the errors are seen to converge to plateau values even for the second replicate. Instead, here the fourth replicate (blue) is shown to be less converged compared to the rest of the replicates. Analysing the concatenated trajectory in Figure S6, gives smooth and well-shaped curves for the probability distribution functions, and both the auto-correlation functions and the block average error estimates look sufficiently converged.

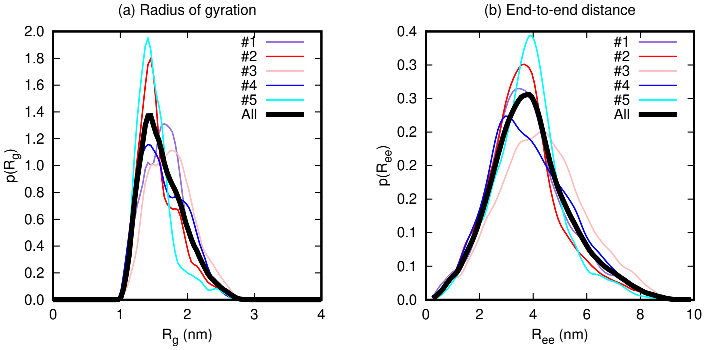


Figure S3. Probability distribution functions of (a) the radius of gyration and (b) the end-to-end distance for each replicate (1-5) and the concatenated trajectory (all).

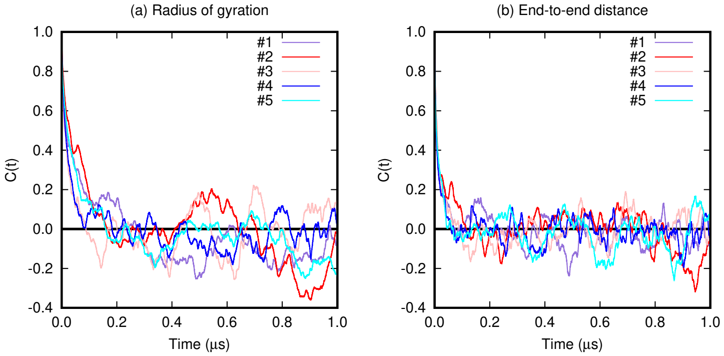


Figure S4. Auto-correlation functions of (a) the radius of gyration and (b) the end-to-end distance for each replicate (1-5).

To further investigate the convergence of the simulation, PCA was performed, and is illustrated for each replicate (1-5) and the concatenated trajectory (all) in Figure S7. Unfortunately, there were fairly large differences between the replicates, so even though they sample the same conformational space in general, different replicates

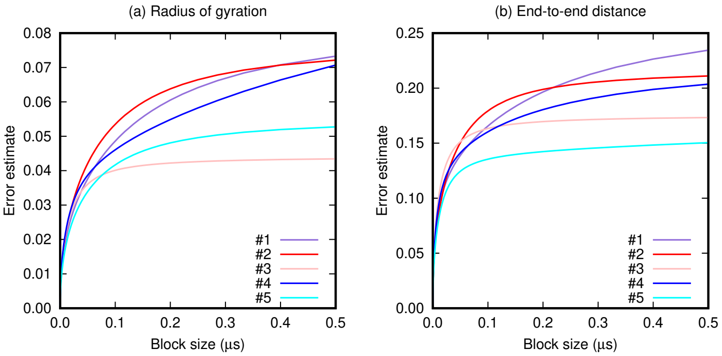


Figure S5. Block average error estimates of (a) the radius of gyration and (b) the end-to-end distance for each replicate (1-5).

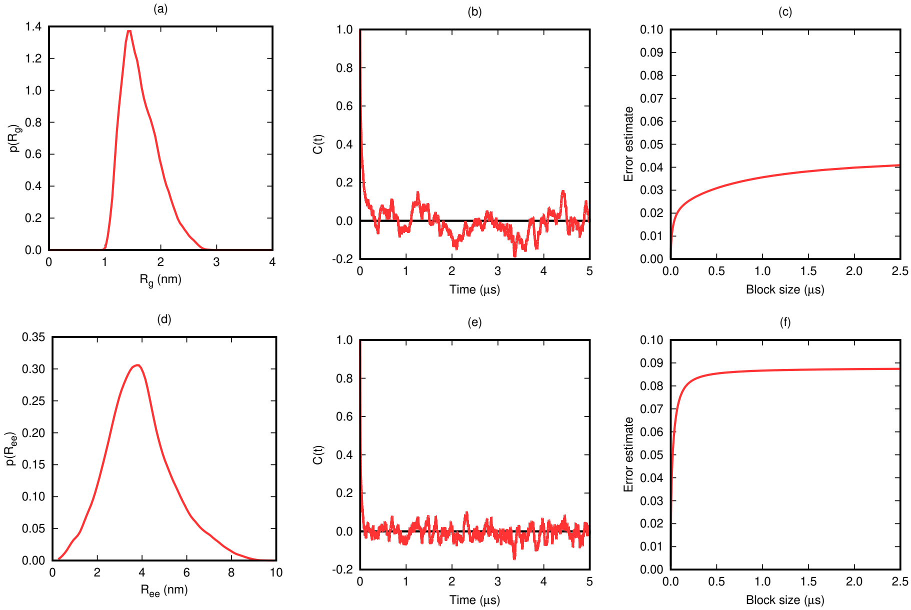


Figure S6. From left to right: probability distribution function, auto-correlation functions, and block average error estimate of (a-c; top row) the radius of gyration and (d-f; bottom row) the end-to-end distance for the concatenated trajectory.

seem to sample different local minima more than others. This is particularly true for replicate 1, 2 and 5. However, PCA of the concatenated trajectory seem to have sampled the entire conformational space sufficiently well.

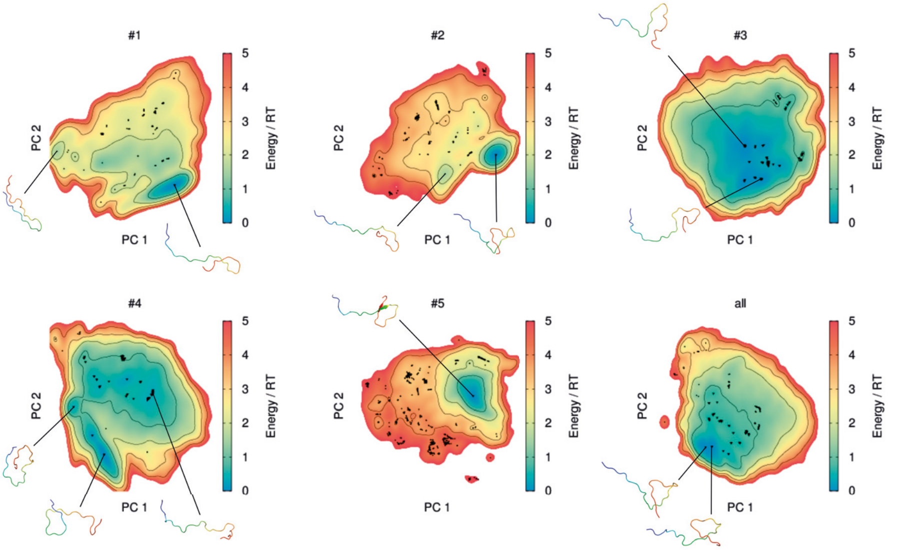


Figure S7. Free energy landscapes of KEIF for each individual replicate (1-5), as well as for the concatenated trajectory (all), using the first two principal components obtained from PCA. Triangles mark energy minima with $RT \leq 1$, and dots mark other local minima with $1 < RT \leq 5$. Snapshots of the representative structure of selected minima are also included.

32 3. LUVs investigated by dynamic light scattering

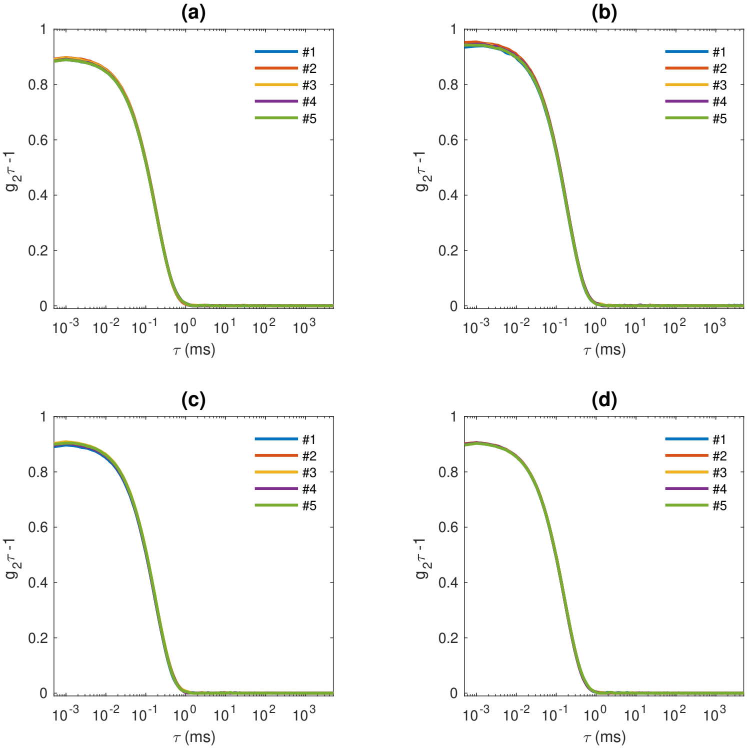


Figure S8. DLS auto-correlation functions obtained with (a) POPS vesicles, (b) POPC vesicles with added KEIF, (c) 3:1 POPC:POPS vesicles, and (d) 3:1 POPC:POPS vesicles with added KEIF. Five consecutive 60 s measurements were performed at scattering angle 173°.

4. Additional cryo-TEM images of LUVs

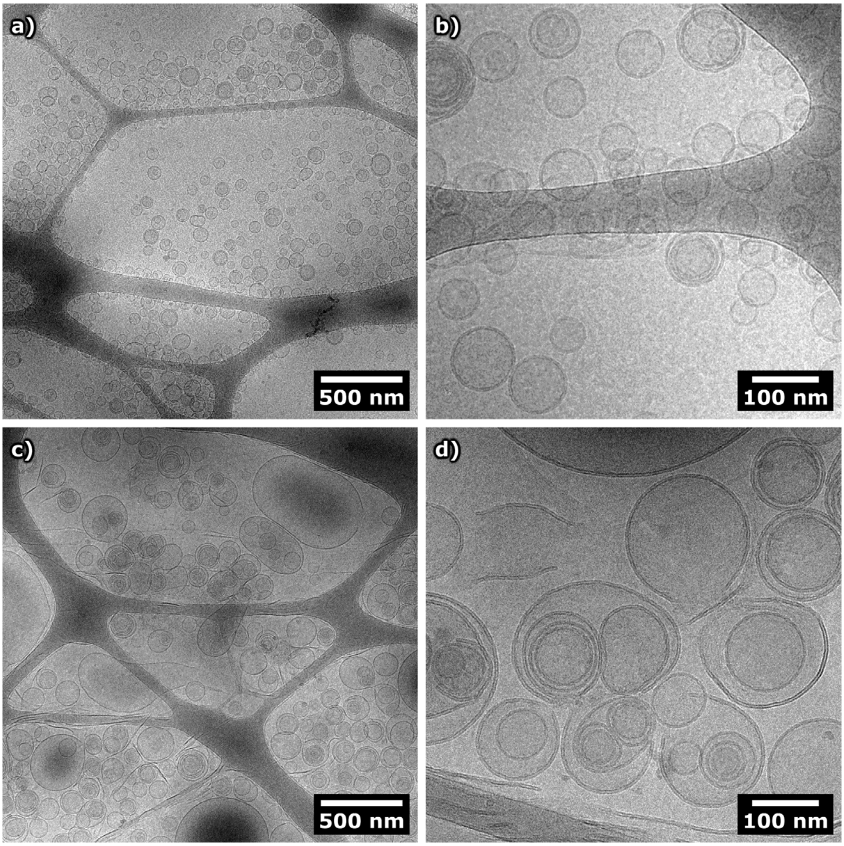


Figure S9. Representative cryo-TEM images of POPC vesicles in the (a, b) absence and (c, d) presence of KEIF, at 10 mM NaF in 20 mM TRIS buffer pH 7.4. The lipid:KEIF molar ratio was 16:1.

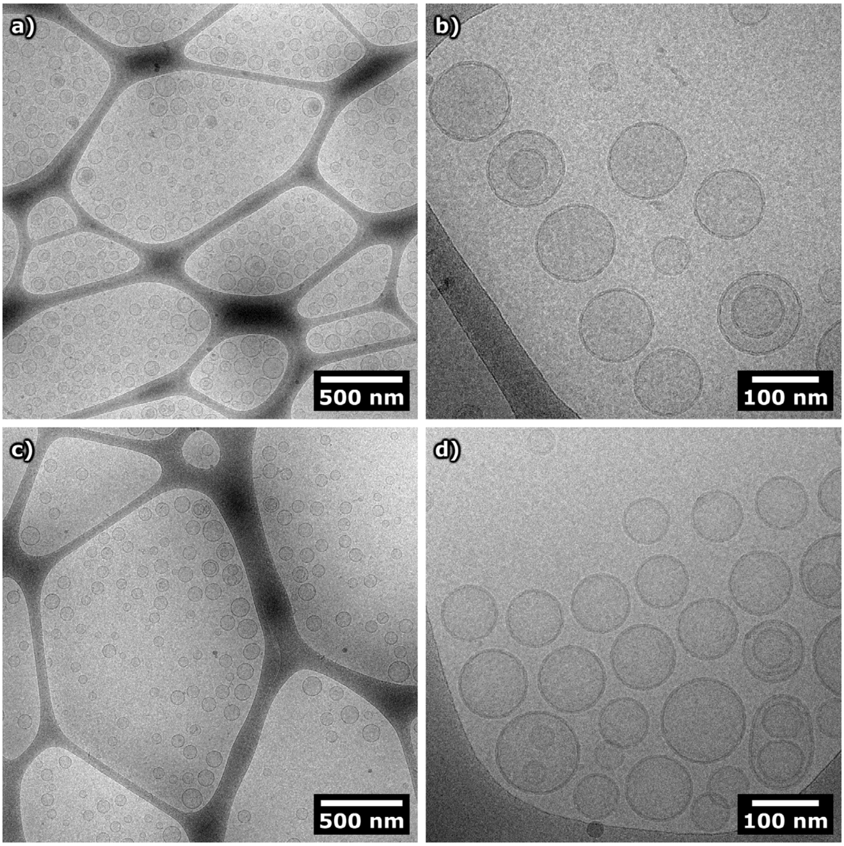


Figure S10. Representative cryo-TEM images of 3:1 POPC:POPS vesicles in the (a, b) absence and (c, d) presence of KEIF, at 10 mM NaF in 20 mM TRIS buffer pH 7.4. The lipid:KEIF molar ratio was 16:1.

5. LUVs' size distributions

Based on the acquired cryo-TEM images, the LUVs' size distributions were obtained by determining their surface area in ImageJ [3], from which their diameter was calculated assuming completely round circles (i.e. spheres in 3D). Considering the small size of the vesicles, flattening effects were assumed negligible; the vast majority of the vesicles were below 150 nm in diameter, and larger (>150 nm) vesicles represented less than 0.5 % of all vesicles analysed. In total, 1822 vesicles were analysed, and the resulting size distributions are displayed in Figure S11. Please, note that the histograms are based on analysing only intact vesicles, and remember that in (b) (i.e. POPC vesicles in the presence of KEIF) the majority of the structures were of irregular shape (incomplete or burst vesicles, discontinuous bilayers, etc.), for which a diameter cannot be defined, and these structures are hence not represented in the corresponding histogram.

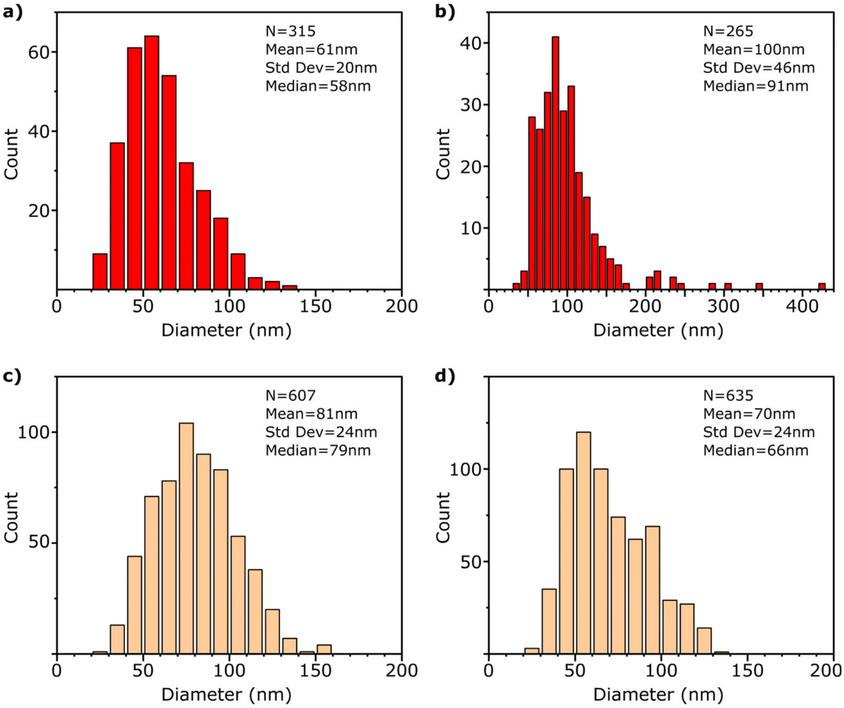


Figure S11. Size distributions, obtained from analysis of cryo-TEM images, of (a, b) POPC and (c, d) 3:1 POPC:POPS vesicles in the (a, c) absence and (b, d) presence of KEIF, at 10 mM NaF in TRIS buffer. The lipid:KEIF molar ratio was 16:1.

References

1. Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.H.; Goto, Y.; Réfrégiers, M.; Kardos, J. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E3095–E3103.
2. Micsonai, A.; Wien, F.; Bulyáki, É.; Kun, J.; Moussong, É.; Lee, Y.H.; Goto, Y.; Réfrégiers, M.; Kardos, J. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **2018**, *46*, W315–W322.
3. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675.

Paper IV

Force field effects in simulations of proteins with varying polyproline II propensity

Stéphanie Jephthah,[*,†] Francesco Pesce,[*,‡] Kresten Lindorff-Larsen,[*,‡] and Marie Skepö[*,†]

†*Division of Theoretical Chemistry, Lund University, Lund, Sweden*

‡*Structural Biology and NMR Laboratory & the Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark*

E-mail: stephanie.jephthah@teokem.lu.se; francesco.pesce@bio.ku.dk; lindorff@bio.ku.dk; marie.skepo@teokem.lu.se

Phone: +46 46 222 17 54; +45 35 32 66 25; +45 35 32 20 27; +46 46 222 33 66

Abstract

Five peptides known to possess polyproline II (PPII) structure have been investigated using atomistic molecular dynamics simulations to compare how well four different force fields known for simulating intrinsically disordered proteins sufficiently well (Amber ff99SB-disp, Amber ff99SB-ILDN, CHARM36IDPSFF, and CHARMM36m) can capture this secondary structure element. The results revealed that all force fields sample PPII structure but to different extents and with different propensities toward other secondary structure elements, in particular, the β-sheet and "random coils". Conformation cluster analysis of simulations of histatin 5 also revealed that the conformational ensembles of the force fields are quite different. We were unable to determine which method was the most accurate; nevertheless, it is most likely that all methods need to be improved to accurately predict the secondary structure of flexible peptides.

1 Introduction

Intrinsically disordered proteins and regions (IDPs and IDRs), are characterized by the lack of a well-defined tertiary structure in aqueous solution. The conformations of IDPs and IDRs are known to vary significantly, which makes it difficult to study them by standard methods. For example, because of their disordered behavior, IDPs and IDRs cannot be crystallized, and until relatively recently many molecular simulations of IDPs tended to sample too compact and too stable structures.

Despite being classified as "unordered", studies originating from the seventies discovered that a few natively unfolded peptides possessed some degree of local order in their backbone, identified as the left-handed polyproline II helix (PPII).[1] As a secondary structure element, the PPII helix is decidedly different from the α-helix and the β-sheet, and perhaps less well-known, although being frequently occurring in many proteins. The PPII helix has backbone dihedral angles of approximately $(\phi, \psi) = (-75°, +145°)$ with a helical pitch of 9.3 Å/turn and 3.0 residues/turn, which causes it to become quite extended.[2,3]

One popular experimental technique for determining the secondary structure of proteins is circular dichroism (CD) spectroscopy. In the CD spectrum, the PPII helix is often associated with a strong band with negative ellipticity around 198 nm and a weak positive band around 218 nm.[1,4,5] There are several software packages available for analyzing CD data and providing estimates of the relative secondary structure content, at least in terms of α-helices and β-sheets. Unfortunately, these algorithms may fall short when it comes to analyzing spectra of more disordered proteins that contain several less common secondary structure elements, including the PPII helix, and where the structural elements are not fixed in time. In such cases, the remaining secondary structure elements are lumped together and categorized as "others" or "random coils". A similar problem is also encountered when determining the secondary structure content of protein structures from simulations. Although the widely used DSSP program is able to identify and quantify a wider collection of secondary structure elements, it does not include the PPII helix. Fortunately, there are other software available that utilizes modified DSSP assignment to also include PPII structure.[6]

The development of force fields for simulating IDPs is constantly evolving to help alleviate the problem of overly collapsed structures in simulations, and to make the simulations as quick and accurate as possible. So far in force field development, the focus have mainly been on optimizing two aspects of IDP simulations.[7] The first aspect is the secondary structure propensities, which are often modified by adjusting the protein backbone dihedral parameters. The second aspect concerns the balance of the protein–solvent interactions, which is crucial to not sample too compact IDP conformations, and to accurately capture the more extended conformations. This is generally controlled by increasing and fine-tuning the interaction between the protein and the water in the simulations.

To evaluate new force fields, different properties can be considered. NMR observables, such as scalar couplings and chemical shifts, are used for assessing force field accuracy by comparing simulated and experimental values. Comparisons of scattering curves and the radius of gyration are used to evaluate compactness of the simulated proteins. In addition to such direct comparison to experimental observables, secondary structure propensities are often also assessed, and although more extensive analyses sometimes are used, they are most often restricted to the α-helix, the β-sheet, and the "random coil". Conformational clustering is also sometimes used as a tool in force field analyses.

Here we present a study where five different short peptides (7–24 residues long) with varying PPII propensities, as well as four variants of one of the peptides, have been simulated with four different force fields that are known to work relatively well for simulating IDPs. Our analyses was mainly focused on differences regarding the secondary structure content across peptides and force fields. Our findings revealed that although all the chosen force fields give rise to conformational ensembles with some level of PPII structure, they do so to different extents and with different propensities toward other secondary structure elements. Additionally, all force fields captured a trend showing that the PPII content increases with the number of proline residues in peptide chains consisting of only alanines and prolines.

2 Methods

2.1 Molecular dynamics simulations

Five different peptides known to possess PPII structure, as well as four different variants of one of the peptides, were simulated using atomistic molecular dynamics (MD) simulations. Names and amino acid sequences of the selected peptides are shown in Table 1.

Table 1: Names, number of amino acid residues (#aa), and the amino acid sequences of the peptides used in this study. Positively charged amino acids are shown in blue and negatively charged amino acids are shown in red.

Name	#aa	Sequence
A7	7	`AAAAAAA`
P-113	12	`AKRHHGYKRKFH`
P_{13}	13	`PPPPPPPPPPPPP`
V1: P_6AP_6	13	`PPPPPPAPPPPPP`
V2: $(PA)_6P$	13	`PAPAPAPAPAPAP`
V3: $(P_3A_2)_2P_3$	13	`PPPAAPPPAAPPP`
V4: $(A_3P_2)_2A_3$	13	`AAAPPAAAPPAAA`
Pep3	13	`DRALGIPELVNQG`
Hst5	24	`DSHAKRHHGYKRKFHEKHHSHRGY`

The simulations were performed using the GROMACS package (version 4.6.7),[8–10] with four different force fields used for simulating IDPs: (A) the AMBER ff99SB-disp force field with its own TIP4P-D-type water model,[11] (B) the AMBER ff99SB-ILDN force field[12] with the TIP4P-D water model,[13] (C) the CHARMM36IDPSFF[14,15] force field with the TIP3P water model,[16] and (D) the CHARMM36m force field,[17] also with the TIP3P water model. When discussed in the text the force field are referred to by their abbreviations and when described in figures and tables, they are referred to by their 1-letter code, see Table 2. A rhombic dodecahedron was used as a simulation box, with periodic boundary conditions in all directions. A minimum distance of 1 nm was set between the solute and the box edges. The initial, linear protein structures were built using PyMOL.[18]

The Verlet leap-frog algorithm,[19] with a time step of 2 fs, was used to integrate the

Table 2: Force field notations

Force field	Abbreviation	1-letter code
AMBER ff99SB-disp	A99SB-disp	A
AMBER ff99SB-ILDN	A99SB-ILDN	B
CHARMM36IDPSFF	C36IDPSFF	C
CHARMM36m	C36m	D

equations of motion. A Verlet list cut-off scheme was used for the non-bonded interactions. Short-ranged interactions were calculated using a pair list with a cut-off of 1 nm, while the long-ranged electrostatics were evaluated by using Particle Mesh Ewald summation[20] with cubic interpolation and a grid spacing of 0.16 nm. Long-ranged dispersion interactions were applied to the energy and pressure of the simulated systems. All bond lengths were constrained using the LINCS algorithm.[21] A velocity-rescaling thermostat[22] with a relaxation time of 0.1 ps was used to keep a temperature of 293 K or 300 K (see Table 3 for details), and a Parrinello-Rahman barostat[23] was used to keep the pressure at 1 bar throughout the simulations. A relaxation time of 2 ps was used, and the isothermal compressibility was set to that of water, i.e. 4.5×10^{-5} bar^{-1}.

Energy minimization was done using the steepest descent algorithm. Equilibration of the temperature and pressure was done in two steps and with position restraints on the proteins: (1) 500 ps in the NVT ensemble, and (2) 1000 ps in the NPT ensemble. Five replicates with different starting seeds were used for each simulation. The final production runs were performed in the NPT ensemble for a total of 5 µs (5×1 µs) for the majority of the peptides. Hst5 with A99SB-ILDN was run for a total of 7 µs ($1 \times 3\,\mu\text{s} + 2 \times 2\,\mu\text{s}$), and P-113 with A99SB-ILDN was run for a total of 12 µs ($1 \times 8\,\mu\text{s} + 2 \times 2\,\mu\text{s}$). The differences between all the simulations is summarized in Table 3. Simulation data of Hst5 with A99SB-ILDN and Hst5 with C36m has previously been published in the paper by Jephthah et al. (2019).[24]

Table 3: Force fields, total production run time (t), and the temperature (T) of all the simulations.

Peptide(s)	FF	t (µs)	T (K)
A7, P-113, P_{13}, V1, V2, V3, V4, Pep3	A	5	300
A7, P_{13}, V1, V2, V3, V4, Pep3	B	5	300
A7, P-113, P_{13}, V1, V2, V3, V4, Pep3	C	5	300
A7, P-113, P_{13}, V1, V2, V3, V4, Pep3	D	5	300
P-113	B	12	300
Hst5	A	5	293
Hst5	B	7	293
Hst5	C	5	293
Hst5	D	5	293

2.2 Simulation analyses

The GROMACS tool `g_cluster` was used to obtain conformational clusters, and to obtain frames for representative structures. All protein structures were visualized and rendered using PyMOL.[18]

2.2.1 Principal component analysis

Principal component analysis (PCA) is a dimensionality reduction method that makes it possible to represent a fraction of the information contained in a large set of variables (or features) in a smaller set. This is achieved by calculating the eigenvectors (or principal components) of the variables' covariance matrix. A PCA was performed, for each peptide, on an aggregated trajectory made by concatenating the trajectories resulting from the four different force fields. This, as suggested in,[25] ensures a robust comparison of the force fields by projecting the resulting trajectories onto common principal components. PCA calculations were performed with pyEMMA[26] using as features the cosine and sine of each backbone dihedral. The analyses were based on the first two principal components.

2.2.2 Secondary structure analysis

Fractions of secondary structure were estimated from the dihedral angles of the peptide backbones, which were obtained by using the GROMACS tool `g_rama`. Only the α-helix (both the right-handed and the left-handed), the β-sheet, and the PPII helix were considered for this analysis. A residue was considered to be in the right-handed α-helix (α_R) region of the Ramachandran map when $-90° \leq \phi \leq -30°$ and $-90° \leq \psi \leq 0°$, as illustrated in the Ramachandran map in Figure 1. Similarly, $30° \leq \phi \leq 90°$ and $0° \leq \psi \leq 90°$ were used for the left-handed α-helix (α_L) region, $-180° \leq \phi < -104°$ and $180° \leq \psi \leq 104°$ for the β-sheet region, and $-104° \leq \phi \leq -46°$ and $116° \leq \psi \leq 174°$ for the PPII helix region. Residues not belonging to any of the aforementioned regions were classified as "random coils". We note that this classification is based not on secondary structure elements, but simply examines which regions of the Ramachandran map the different residues populate.

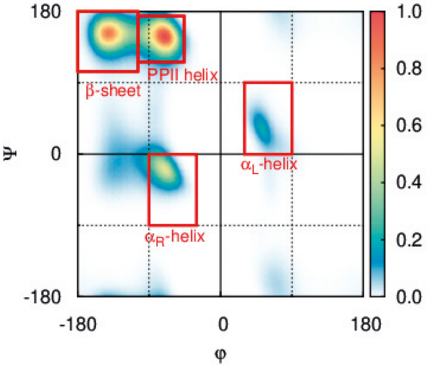


Figure 1: Example of a Ramachandran map illustrating the four different secondary structure regions analyzed in this study. The β-sheet region is defined by $-180° \leq \phi < -104°$ and $180° \leq \psi \leq 104°$; the PPII helix region by $-104° \leq \phi \leq -46°$ and $116° \leq \psi \leq 174°$; the α_R-helix region by $-90° \leq \phi \leq -30°$ and $-90° \leq \psi \leq 0°$; and the α_L-helix region by $30° \leq \phi \leq 90°$ and $0° \leq \psi \leq 90°$. Anything outside of these regions is classified as "random coil". The plot is normalized for a maximum intensity of 1.

2.2.3 CD prediction

To predict circular dichroism (CD) spectra from structural ensembles we employ SESCA.[27] The SESCA algorithm has two steps:

1. Per residue secondary structure assignment. We use DISICL[28] as secondary structure prediction algorithm as it is the one explicitly taking into account PPII conformations.
2. Spectral contributions from each secondary structure element in a conformation are combined to produce the CD spectra. In SESCA, the set of spectral contributions assigned to subsets of secondary structures are stored in the "basis sets". Different basis sets for a given secondary structure assignment are available, which differ in the resolution of the spectral contributions definition. Optionally, side chains' spectral contributions can be added. We tested several of the available basis sets, but mainly used the DS6-1SC1 one (DS6-1 with side chains contribution), as this was the one giving rise to the predicted spectra that resembled the most in shape the experimental ones.

Finally, the CD spectra from each conformation of the ensemble are linearly averaged.

3 Results and discussion

3.1 Effects of force field in simulations of five peptides

An initial comparison of the effects of the force fields was done by analyzing the resulting average radius of gyration of the five peptides, see Table 4. All force fields resulted in similar average values for the radius of gyration for each of the individual peptides, although C36IDPSFF generally seemed so sample slightly more compact conformations compared to the other force fields. This was, however, not the case for P_{13}, for which C36m sampled a slightly smaller average instead. Overall, it seemed like both A99SB-disp and A99SB-ILDN sampled similar averages for all peptides.

Table 4: The radius of gyration, R_g (nm), for the five peptides with the four different force fields.

Force field	A7[†]	P-113	P_{13}[†]	Pep3	Hst5
A	0.62	0.97±0.01	1.14	0.91±0.03	1.29±0.08
B	0.63	0.91±0.03	1.14	0.94±0.01	1.29±0.05
C	0.60	0.86±0.01	1.13	0.87±0.01	1.18±0.02
D	0.61	0.92±0.02	1.11	0.97±0.01	1.35±0.03

[†] The values of A7 and P_{13} are reported without error margins in the table since their errors are smaller than 0.005 nm.

Because the conformational ensembles of IDPs are highly heterogeneous, it is not trivial to find a set of variables that can describe the high variability of an ensemble in a low-dimensional representation. For each peptide, we thus used PCA (on aggregated trajectories over all force fields as discussed in the methods) to represent and visualize the simulations in a space of reduced dimensionality. After projecting the ensembles from the different force fields onto a common subspace we examine the free energy surfaces projected as a function of the first two principal components, and in general find relatively similar free energy surfaces. However, the relative probabilities of the conformational states may differ, with C36IDPSFF being the one giving rise to the less rough free-energy surfaces, while the others show regions poorly explored at the simulated temperature because of high conformational energies (Figure 2).

Nonetheless, it is worth highlighting that, in case of P_{13} with C36m, we observe a shift of the minimum on the second PC axis. Since P_{13} is thought to mostly populate PPII conformations, we decided to characterize and compare the free-energy minima resulting from A99SB-disp and C36m. Subtle differences were observed, both in the average radius of gyration (1.14 nm for A99SB-disp and 1.11 nm for CHARMM36m), and in the per residue average backbone dihedrals, that for both force fields reside in the PPII ranges defined in DSSP-PPII[29] (Figure 3c). At the level of local structure we find that A99SB-disp populates more PPII conformations than C36m, see Figure 3d. Additionally, the PPII helix formed in simulations with A99SB-disp is more bent in respect an imaginary helix axis, while the PPII

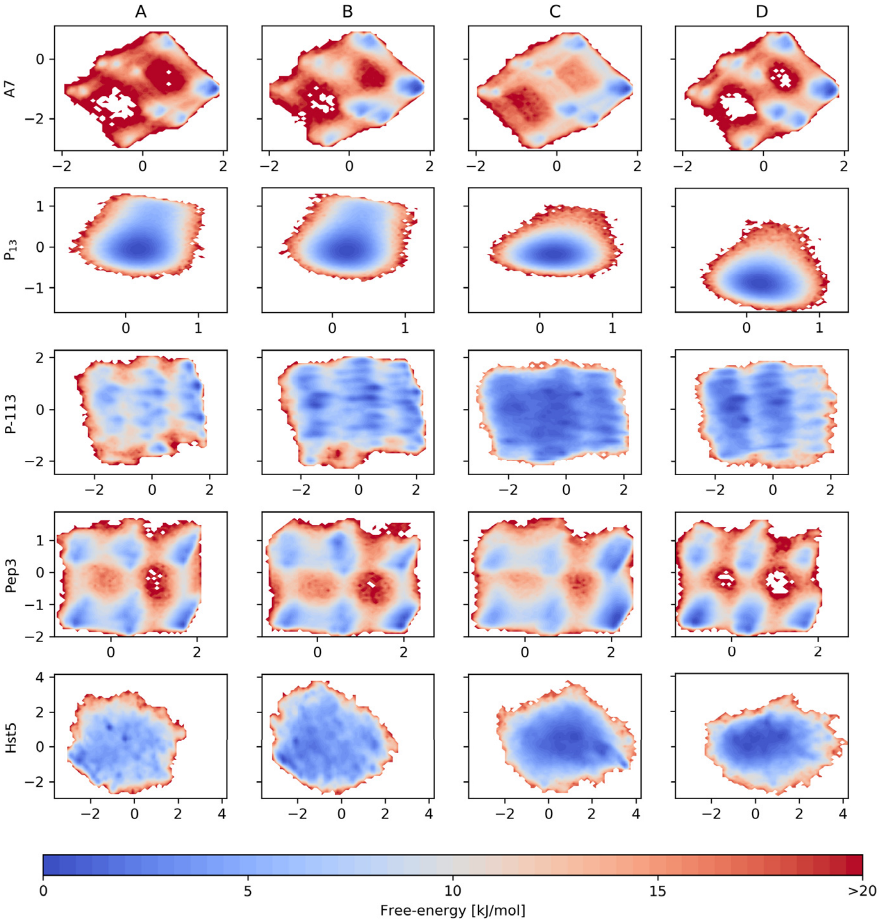


Figure 2: Free-energy surfaces as a function of the first (x-axis) and second (y-axis) principal components. As a result of performing the PCA on the aggregated trajectories, the PC coordinates for all the force fields on a row are the same.

helix formed with C36m is more straight, see Figure 3a-b.

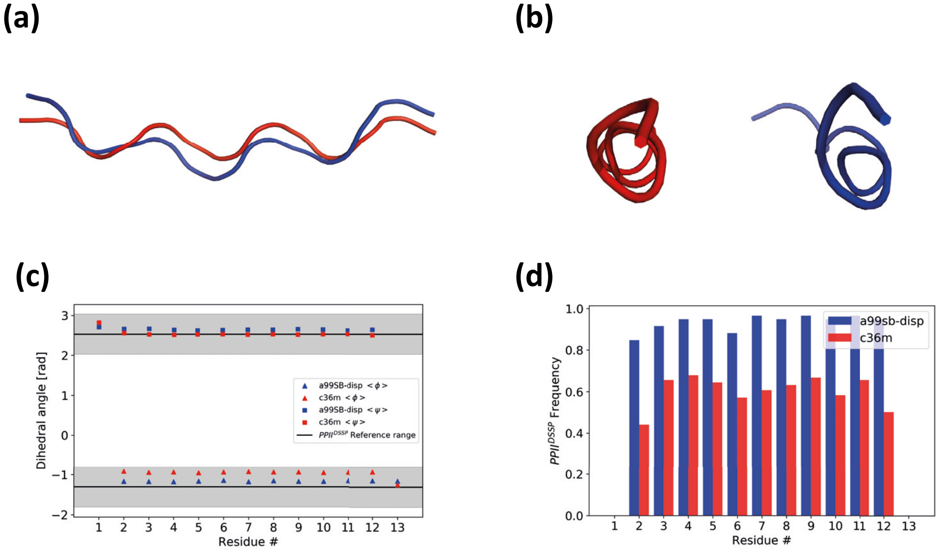


Figure 3: Structural analysis of the free-energy minima in A99SB-disp (blue) and C36m (red). (a) Side view and (b) C-terminal view of two representative P_{13} structures from A99SB-disp and C36m. (c) Average per residues ϕ and ψ dihedrals, compared to the PPII helix range defined in DSSP-PPII. (d) Per residue probability of PPII conformations as predicted by DSSP-PPII.

Average populations of different regions of the Ramachandran map (corresponding to typical dihedral angles in different seconary structure elements) was estimated from all of the simulations, see Figure 4. All simulated peptides were found to be dominated by PPII structure (31-97 %). Little β-sheet structure (13-32 %) was also observed for all peptides except P_{13}. The α-helical content was found to be low, and rarely larger than ~10–15 %, except for in Hst5 where it generally was slightly higher (13–21 %). Comparing the different force fields, A99SB-ILDN and C36m were found to sample more β-sheet structures compared to the other force fields, and the PPII content generally was slightly higher in the A99SB-disp simulations. C36IDPSFF generally sampled more "random coils", except for the case of P_{13}. For this peptide, which is expected to mainly possess PPII structure in aqueous solution,[30,31] all force fields except C36m had more than 90 % PPII content. Instead, C36m only captured slightly more than 70 % PPII content. These observations, together with the average radius of gyration, suggests that a higher content of "random coil" provides slightly more compact average conformations.

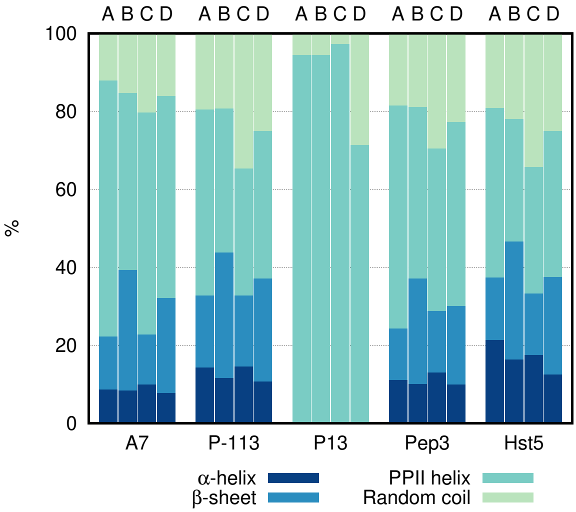


Figure 4: Stacked secondary structure histograms of the simulated peptides based on the regions defined in Figure 1. The α-helix includes both the α_R-helix and the α_L-helix.

Differences between the force fields were further analyzed for Hst5. Cluster analysis was performed for each individual force field, as well as for a concatenated trajectory in which all four force fields were included. The analysis was done using an RMSD cutoff of 0.5 nm. This value was chosen by examining the total population of the eight first clusters, and varying the cluster radius (in steps of 0.05 nm) until their total population was closest to 50 %. Additionally, using the same cutoff for all force fields made it easier to compare them.

The representative structures of the top eight conformational clusters of Hst5 with the four force fields, as well as the force field mix, are presented in Figure 5. Visual inspection of the representative structures immediately reveals that the first cluster conformers are different for the different force fields. Comparing the combined percentage sizes of the top eight conformation clusters gives some indication of the relative flexibility of the conformations sampled by the four different force fields. A higher value means that there are fewer conformations sampled in the remaining clusters, which suggests a lower degree of flexibility. By this reasoning, of the four selected force fields, A99SB-disp provides least degree of flexibility and A99SB-ILDN provides the lowest degree of flexibility, while the remaining two force fields lie in between and provide similar degrees of flexibility.

To get a more quantitative comparison between the force fields, their trajectories were concatenated, followed by a new cluster analysis where each structure could be traced back to their individual force field. The relative cluster populations of the individual force fields in the top eight clusters are illustrated in Figure 6. Although all force fields are represented in each cluster, they are not evenly distributed. For example, the first cluster is dominated by C36IDPSFF, whereas the second cluster mainly contains conformations from A99SB-disp and C36m. The fifth cluster is the most evenly distributed cluster across the force fields, and the sixth cluster is heavily dominated by A99SB-ILDN. From this analysis it is safe to say that although the average properties of different force fields may be similar, the force field's individual conformational ensembles are rather different, which naturally leads to different secondary structure content.

3.2 CD prediction using SESCA

Since the four force fields give rise different conformational ensembles, one may reasonably ask which of the force field is more representative of the real conformational ensemble in solution. To answer this question we may attempt to compare the simulations to experimental data. This ideally requires a forward model to predict the experimental observables from an ensemble of structures. We here used data from CD spectroscopy, as CD is highly sensitive of secondary structure composition, and use SESCA[27] as forward model. Experimental data for A7, P-113 and Hst5 were obtained from Graf et al. (2007),[32] Han et al. (2016),[33] and Jephthah et al. (2019),[24] respectively. Unfortunately, and as also noted for other IDPs in the papers by Fagerberg et al. (2020)[34] and Gopal et al. (2020),[35] it was

Figure 5: Representative structures of the top eight cluster conformations of Hst5 as simulated with the four different force fields (A-D), as well as from a mixture of the four force fields (Mix). The total percentage of the top eight clusters is given above the structures, and the relative size of each individual cluster is given below each structure.

not possible to obtain a meaningful agreement between the CD spectra predicted by SESCA and the experimental ones (Figure 7). This can be due to the fact that the main negative signature peak of a PPII helix may appear in experiments between 190 nm and 210 nm[36–38] because of non-secondary structure contributions, while the spectral contribution associated

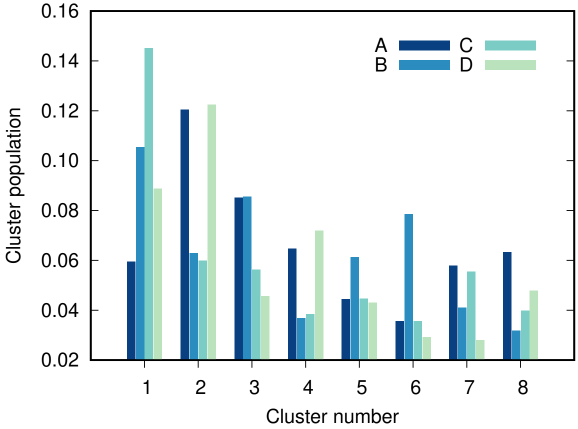


Figure 6: Weighted cluster population from the individual force fields in the top eight clusters of Hst5.

to a PPII conformation in SESCA has a fixed position. Also, a qualitative analysis based on the intensity of main negative peak does not provide a decisive suggestion of what force field may be the most reliable. This is also complicated by some intensity scaling that may be needed to take into account uncertainty in the estimate of the concentration of the sample used for the experimental CD data. At this stage it is not clear if the source of the problem may be the force fields' inaccuracy, finite sampling, or the inaccuracy of the CD calculation for this kind of systems.

3.3 The effect of proline residue content

A few variants of P_{13}, V1-V4 (see Table 1), were investigated to see how the proline content affected the PPII propensities. Figure 8 shows the PPII content as a function of the number of proline residues in P_{13} and the peptide variants. Linear regression revealed that all force fields yielded significant correlations ($p < 0.05$) between the PPII content and the number of proline residues for P_{13} and the chosen variants, where an increased number of proline residues provided a larger PPII content. The slope of these trends did however differ depending on what force field that used, with C36m having the smallest increase and

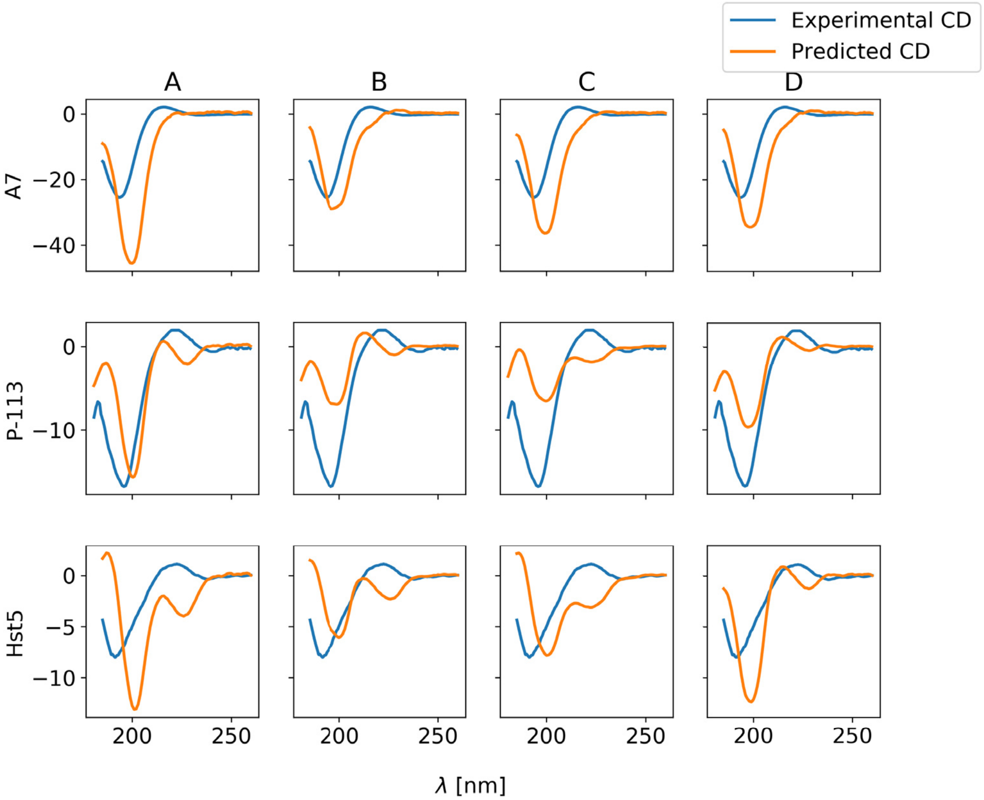


Figure 7: Comparison of the experimental CD spectra with the ones predicted with SESCA from the conformational ensembles produced employing different force fields. The y-axes show the ellipticity, θ ($\mathrm{deg\,cm^2/dmol}$).

A99SB-ILDN having the largest increase.

Although interesting, further investigation is needed to completely characterize this trend. Some questions of interest are: Is the trend affected by the peptide length? Is the trend affected by the relative position of the prolines in the amino acid sequence?

4 Conclusions

In this study we have evaluated the differences between four different force fields in simulations of five short peptides with varying PPII propensities. All force fields gave similar

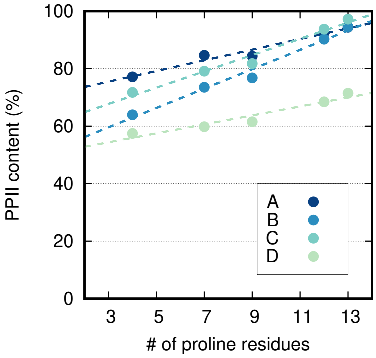


Figure 8: PPII content based on the regions defined in Figure 1 as a function of the total number of proline residues for the four different force fields. Values from P_{13} and its variants were used for this figure.

ensemble averages of the radius of gyration, although the averages by the C36IDPSFF force field were generally slightly smaller compared to the other force fields. All force fields gave rise to similar conformational spaces (as probed by the first two principal components) for each individual peptide, although with slightly different probabilities. Similarly, all force fields sampled PPII structure, but to different degrees. Additionally, some force fields were more prone to sampling other secondary structure elements. For example, A99SB-ILDN and C36m sampled more β-sheet structures, C36IDPSFF sampled more "random coil", and A99SB-disp often had the highest PPII content. There also seemed to be a vague correlation between higher "random coil" content and more compact conformations. Direct comparison by conformation clustering revealed that the force fields have a bias toward different conformational clusters. CD prediction using SESCA was performed to examine which force field provided a more accurate conformational ensemble. Unfortunately, the method was not able to match the predicted spectra to any experimental spectra. Finally, the effect of proline residue content on the PPII content of short peptides containing only alanine and proline was investigated, and observed a correlation between the number of prolines in the amino acid sequence and the PPII content. We conclude by highlighting that we need better methods to calculate experimental observables that are sensitive to secondary structure preferences for flexible peptides.

Acknowledgement

This research was funded by the NordForsk's Nordic Neutron Science Programme (grant number 81912) and the Lundbeck Foundation BRAINSTRUC initiative in structural biology (R155-2015-2666, lundbeckfonden.com). Computer resources for the simulations were provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC). We also acknowledge access to computational resources from the ROBUST Resource for Biomolecular Simulations (supported by the Novo Nordisk Foundation grant no. NF18OC0032608).

References

(1) Krimm, S.; Tiffany, M. L. The circular dichroism spectrum and structure of unordered polypeptides and proteins. *Israel Journal of Chemistry* **1974**, *12*, 189–200.

(2) Kakinoki, S.; Hirano, Y.; Oka, M. On the stability of polyproline-I and II structures of proline oligopeptides. *Polymer Bulletin* **2005**, *53*, 109–115.

(3) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chemical reviews* **2006**, *106*, 1877–1897.

(4) Adzhubei, A. A.; Sternberg, M. J.; Makarov, A. A. Polyproline-II helix in proteins: structure and function. *Journal of molecular biology* **2013**, *425*, 2100–2132.

(5) Uversky, V. N.; Dunker, A. K. *Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools*; Springer, 2012.

(6) Chebrek, R.; Leonard, S.; de Brevern, A. G.; Gelly, J.-C. PolyprOnline: polyproline helix II and secondary structure assignment database. *Database* **2014**, *2014*.

(7) Zerze, G. H.; Zheng, W.; Best, R. B.; Mittal, J. Evolution of All-atom Protein Force Fields to Improve Local and Global Properties. *The journal of physical chemistry letters* **2019**,

(8) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.

(9) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.

(10) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

(11) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **2018**, *115*, E4758–E4766.

(12) 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: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(13) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *Journal of Physical Chemistry B* **2015**, *119*, 5113–5123.

(14) Liu, H.; Song, D.; Lu, H.; Luo, R.; Chen, H.-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Chemical biology & drug design* **2018**, *92*, 1722–1735.

(15) Liu, H.; Song, D.; Zhang, Y.; Yang, S.; Luo, R.; Chen, H.-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Physical Chemistry Chemical Physics* **2019**, *21*, 21918–21931.

(16) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(17) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *nature methods* **2016**, *14*, 71.

(18) Schrödinger LLC, The PyMOL Molecular Graphics System, Version 1.2r1. 2009.

(19) Berendsen, H.; Van Gunsteren, W. Practical algorithms for dynamic simulations. *Molecular-dynamics simulation of statistical-mechanical systems* **1986**, 43–65.

(20) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. *Journal of Chemical Physics.* **1993**, *98*, 10089.

(21) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

(22) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **2007**, *126*, 014101.

(23) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182.

(24) Jephthah, S.; Staby, L.; Kragelund, B.; Skepo, M. Temperature dependence of intrinsically disordered proteins in simulations: What are we missing? *Journal of chemical theory and computation* **2019**, *15*, 2672–2683.

(25) Martín-García, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing molecular dynamics force fields in the essential subspace. *PLoS One* **2015**, *10*, e0121114.

(26) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *Journal of chemical theory and computation* **2015**, *11*, 5525–5542.

(27) Nagy, G.; Igaev, M.; Jones, N. C.; Hoffmann, S. V.; Grubmüller, H. SESCA: predicting circular dichroism spectra from protein molecular structures. *Journal of chemical theory and computation* **2019**, *15*, 5087–5102.

(28) Nagy, G.; Oostenbrink, C. Dihedral-based segment identification and classification of biopolymers I: proteins. *Journal of chemical information and modeling* **2014**, *54*, 266–277.

(29) Mansiaux, Y.; Joseph, A. P.; Gelly, J.-C.; de Brevern, A. G. Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PloS one* **2011**, *6*, e18401.

(30) Tiffany, M. L.; Krimm, S. Circular dichroism of poly-L-proline in an unordered conformation. *Biopolymers: Original Research on Biomolecules* **1968**, *6*, 1767–1770.

(31) Dukor, R. K.; Keiderling, T. A. Mutarotation studies of poly-L-proline using FTIR, electronic and vibrational circular dichroism. *Biospectroscopy* **1996**, *2*, 83–100.

(32) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *Journal of the American Chemical Society* **2007**, *129*, 1179–1189.

(33) Han, J.; Jyoti, M. A.; Song, H.-Y.; Jang, W. S. Antifungal activity and action mechanism of histatin 5-halocidin hybrid peptides against Candida ssp. *PLoS One* **2016**, *11*, e0150196.

(34) Fagerberg, E.; Månsson, L. K.; Lenton, S.; Skepö, M. The Effects of Chain Length on the Structural Properties of Intrinsically Disordered Proteins in Concentrated Solutions. *The Journal of Physical Chemistry B* **2020**, *124*, 11843–11853.

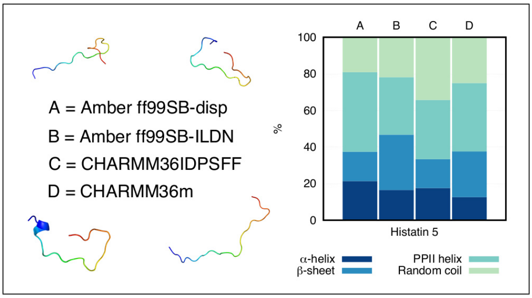
(35) Gopal, S. M.; Wingbermühle, S.; Schnatwinkel, J.; Juber, S.; Herrmann, C.; Schäfer, L. V. Conformational Preferences of an Intrinsically Disordered Protein Domain: A Case Study for Modern Force Fields. *The Journal of Physical Chemistry B* **2020**, *125*, 24–35.

(36) Lopes, J. L.; Miles, A. J.; Whitmore, L.; Wallace, B. A. Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. *Protein Science* **2014**, *23*, 1765–1772.

(37) Rucker, A. L.; Creamer, T. P. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein science* **2002**, *11*, 980–985.

(38) Woody, R. W. Circular dichroism spectrum of peptides in the poly (Pro) II conformation. *Journal of the American Chemical Society* **2009**, *131*, 8234–8245.

Graphical TOC Entry



Paper v

The effect of histidine protonation on the conformational ensemble of histatin 5

Stéphanie Jephthah,* Marco Polimeni, Mikael Lund, and Marie Skepö*

Division of Theoretical Chemistry, Lund University, Lund, Sweden

E-mail: stephanie.jephthah@teokem.lu.se; marie.skepo@teokem.lu.se

Phone: +46-46-222 17 54; +46-46-222 33 66

Abstract

1 Introduction

Histatin 5 (Hst5) is an intrinsically disordered protein (IDP) that is important and interesting for several different reasons. Firstly, as a natural constituent of saliva, Hst5 aids in preserving the oral health by inhibiting of growth and germination of *C. albicans*,[1–4] and because Hst5 is a naturally occurring antimicrobial peptide, there is a great interest in utilizing it as a therapeutic agent for treating and preventing candidiasis and other fungal infections.[4] Secondly, because Hst5 is a good example of a polyampholytic IDP that is relatively easy to work with in the lab, it is also a very convenient alternative to use as a model IDP when investigating different simulation models.

Almost 30 % of the 24 amino acid residues of Hst5 are charged at physiological pH, and with seven positively charged residues and two negatively charged residues, the entire peptide has a net charge of $+5e$. However, another 30 % of the amino acids are histidines

(HIS, $\mathrm{pK_a}$~6) which are known to be able to charge regulate when the pH of the system is close to the $\mathrm{pK_a}$ of the HIS residues, or if interacting with other charged species in the system. Thus, it is common for HIS residues to be partially charged, even at physiological pH. A plot showing how the capacitance of Hst5 changes with pH is presented in Figure 1. A high charge capacitance also means a good charge-regulating ability. Thus, in a salt-free system, the ability of Hst5 to charge regulate is best around pH 5, whereas in a system with 150 mM salt it is shifted up to pH 6.

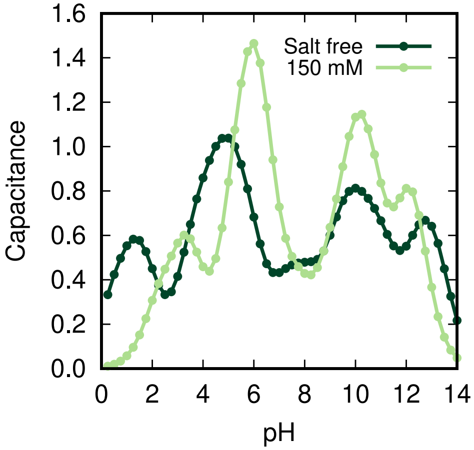


Figure 1: Capacitance of Hst5 as a function of pH in (i) a salt-free system (dark green) and (ii) in a system with 150 mM 1:1 salt (light green). The data that was used to create this plot originates from the paper by Kurut et al. (2014).[5]

Previous studies have shown that Hst5 adsorbs to negatively charged surfaces, such as hydroxyapatite[6,7] and poly(methyl methacrylate).[8] A computational study of Hst5 adsorption from 2014[5] investigated the role of HIS for charge regulation and found that Hst5 adsorption to a negatively charged surface was highest at pH 6 and that charge regulation promotes adsorption. The study concluded that the charge regulation of HIS residues might be a key feature that is responsible for the antimicrobial properties of Hst5. Adsorption of Hst5 to hydrophilic silica surfaces has also been studied both experimentally and computationally to investigate the effect of pH.[9] The study found that Hst5 adsorption increased with the pH, but since the surface charge density of hydrophilic silica also is pH dependent[10] the increased adsorption could also be caused by the decreased surface charge density.

Normally when performing simulations of Hst5 at pH 7, the HIS residues are assumed to be neutral, and in a bulk simulation of a single peptide, this is usually a reasonable assumption. However, Hst5 is a biologically active peptide that interacts with negatively charged biomembranes, and for simulations of such systems, it might be unreasonable to assume that the HIS residues are neutral. By instead assuming that all of the HIS residues are protonated, Hst5 would obtain a net charge of $+12e$, which in theory should alter the conformational properties of Hst5 significantly. Thus, it is important to be aware of the conformational differences that arise due to the charge regulation, because that will aid in explaining the biological functions of Hst5.

In the study presented here, molecular dynamics (MD) simulations of two Hst5 variants have been performed: (i) Hst5 with all HIS residues being neutral (Hst5_HIE), and (ii) Hst5 with all HIS residues being protonated (Hst5_HIP), that is, with a charge of $+1e$. These two extreme cases were chosen to determine and evaluate the largest conformational differences possible that can arise upon protonation of the HIS residues in Hst5. By illustrating these differences, we also emphasize the importance of considering the protonation state of the HIS residues, which does not only depend on the pH but also on other charged species in the simulated system.

2 Methods

2.1 The Hst5 variants

Two variants of Hst5 were included in this study: (i) Hst5 with all HIS residues neutral (Hst5_HIE), and (ii) Hst5 with all HIS residues protonated (Hst5_HIP). A schematic illustration of the charge distribution of the two variants is shown in Figure 2. Hst5_HIE has a net charge of $+5e$ and Hst5_HIP has a net charge of $+12e$.

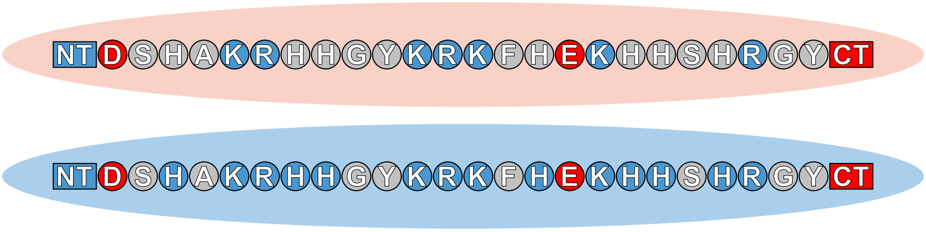


Figure 2: Charge distribution of the two different Hst5 variants simulated in this study with Hst5_HIE at the top (red background) and Hst5_HIP (blue background) at the bottom. Blue residues have a charge of $+1e$, red residues have a charge of $-1e$, and gray residues are neutral. The protein termini are explicitly included with their own charges.

2.2 MD simulations

MD simulations of the three Hst5 variants were performed using the GROMACS package (version 4.6.7),[11–13] with the AMBER ff99SB-ILDN force field[14] in combination with the TIP4P-D water model.[15] The initial, linear protein structure was built using PyMOL,[16] and the charges of the HIS residues were set when building up the simulation systems in GROMACS. Chloride ions were added to neutralize the systems. A rhombic dodecahedron was used as a simulation box, with periodic boundary conditions in all directions, and a minimum distance of 1 nm was set between the solute and the box edges.

A Verlet leap-frog algorithm,[17] with a time step of 2 fs, was used to integrate the equations of motion, and a Verlet list cut-off scheme was used for the non-bonded interactions. The short-range interactions were calculated using a pair list with a cut-off of 1 nm, and the long-ranged electrostatics were managed by using Particle Mesh Ewald[18] with cubic interpolation and a grid spacing of 0.16 nm. Long-range dispersion interactions were applied to the energy and pressure of the simulated systems. The bond lengths were constrained using the LINCS algorithm.[19] The temperature was kept at 300 K by a velocity-rescaling thermostat[20] with a relaxation time of 0.1 ps, and a Parrinello-Rahman pressure coupling[21] was used to keep the pressure at 1 bar. The isothermal compressibility was set to that of water, i.e. 4.5×10^{-5}

bar^{-1}, and the relaxation time was set to 2 ps.

Energy minimization was performed by using the steepest descent algorithm. After the energy minimization, five replicates with different starting seeds were created for each system. The temperature and pressure were equilibrated with position restraints on the proteins; initially by a 500 ps simulation per replicate in the NVT ensemble, followed by a 1000 ps simulation per replicate in the NPT ensemble. The final production run of each replicate was 1 µs, giving a total simulation time of 5 µs per system.

2.3 MC Simulations

The atomistic trajectories of Hst5_HIE and Hst5_HIP were reduced into two libraries by saving every 100th structure (frame). The resulting libraries, composed of 500 structures each, were coarse-grained at the amino-acid level. Each amino acid, including the C and the N terminals, was represented as a spherical bead centered around its center of mass. These libraries were then used in Metropolis Monte Carlo simulations to investigate the behavior of Hst5_HIE and Hst5_HIP at the interface of a negatively charged surface. The simulation box had a z-dimension of 300 Å and a base x–y of 182 Å×158 Å, see Figure 3. The surface had a net charge of $-36e$ and was placed parallel to the x–y plane at $z = 0$ Å. The surface was modeled as in the recent work by Hyltegren et al. (2020),[22] that is, as a smooth charge distribution with a charge density of $-2.0\,\mu\text{C/cm}^2$. Periodic boundary conditions were only applied in the x–y direction. All MC simulations were performed with Faunus[23,24] (git revision: 715a484), which allows for different MC moves: the protein could rotate, translate, and its conformation was swapped along with the simulation with a conformation picked up from the library. Each structure was selected with the same initial probability (1 over 500) for being used with the move, then the final occurrence of a given structure was only driven by the interactions in solution. Salt was accounted for explicitly using a grand-canonical scheme for particle insertion. A specific salt concentration, c, was reproduced by fixing the salt activity according to the relation $a = \gamma c$ where a is the salt activity and γ is the salt activity coefficient, which was extrapolated by fitting the data from Robinson & Stokes (1949).[25] The whole box was made electroneutral by inserting the appropriate number of counter-ions to neutralize the net charge.

The interactions between the protein, the salt particles, and the surface were purely electrostatic and were sampled according to the following potential energy:

$$U = \sum_i \sum_j \frac{e^2 z_i z_j}{4\pi\epsilon_0\epsilon_r r_{ij}}, \tag{1}$$

where e is the electron charge, z_i and z_j are the charges of the i^{th} and j^{th} interaction site, ϵ_0 is the vacuum permittivity, $\epsilon_r = 78.7$ is the relative dielectric constant of the water at room temperature (25º C), and r_{ij} the distance between the i^{th} and j^{th} charges.

Each simulation comprised 50000000 MC steps and the id conformation was registered as a function of the z-distance from the surface every 1000^{th} step. From this id, the conformations were constructed back the corresponding atomistic trajectories, on which further analysis was performed. Therefore, although data was obtained from MC simulations, the analysis was always based on the all-atom model.

2.4 Analyses

The average radius of gyration (R_g) was obtained from the simulations by using the GROMACS tool `g_polystat`. MD simulation convergence was assessed by computing autocorrelation functions and block error estimates of R_g using the GROMACS tool `g_analyze` (see Figure S1 in the Supporting Information). To ensure that the simulated peptides did not interact with their periodic images in the MD simulations, the minimum distance between the periodic images in the simulations was monitored by using the GROMACS tool `g_mindist`. Estimations of the full width half maximum, FWHM, of the distance distribution of the radius of gyration, $p(R_g)$, were obtained by fitting the $p(R_g)$ curves with a Gaussian function before calculating the FWHM by the method previously reported by Cragnell et

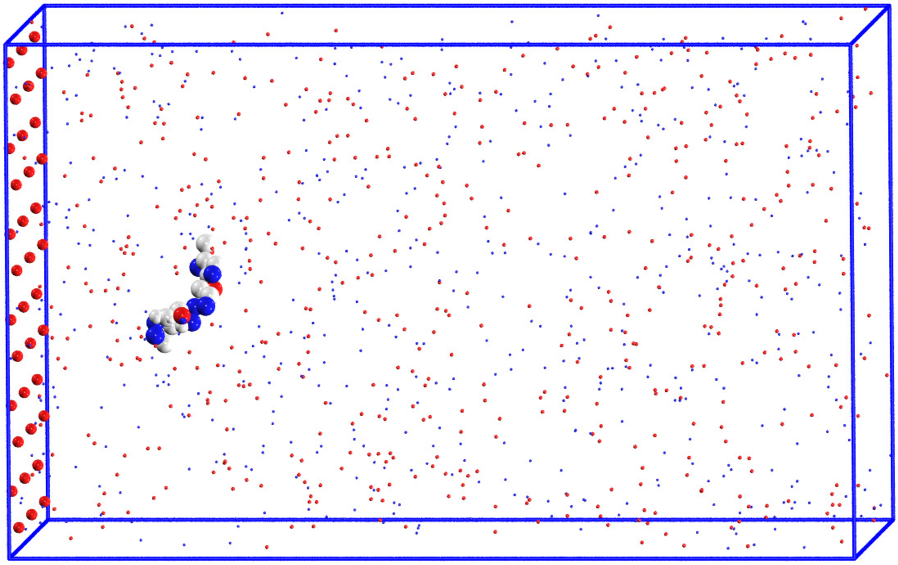


Figure 3: Simulation box for MC simulations: a parallelepiped with a x–y basis of 182 Å×158 Å and z-dimension was 300 Å. The box contains a Hst5 peptide, the neutral silica surface (left side), salt particles, and counter-ions. Blue and red colors indicate positive and negative charges, respectively.

al. (2018).[26] Conformation clustering analysis was performed by using the GROMACS tool `g_cluster` with the GROMOS method.[27] Secondary structure estimates were obtained by computing the backbone dihedral angles with the GROMACS tool `g_rama`. For this analysis, only the α-helix (both the right-handed and the left-handed), the β-sheet, and the PPII helix were considered. A residue was considered to belong to the right-handed α-helix (α_R) region of the Ramachandran map when $-90^\circ \leq \phi \leq -30^\circ$ and $-90^\circ \leq \psi \leq 0^\circ$, as illustrated in Figure 4. Similarly, $30^\circ \leq \phi \leq 90^\circ$ and $0^\circ \leq \psi \leq 90^\circ$ was used for the left-handed α-helix (α_L), $-180^\circ \leq \phi < -104^\circ$ and $180^\circ \leq \psi \leq 104^\circ$ for the β-sheet, and $-104^\circ \leq \phi \leq -46^\circ$ and $116^\circ \leq \psi \leq 174^\circ$ for the PPII helix. Residues not belonging to any of the aforementioned regions were classified as "random coils". This classification is not based on secondary structure elements, but simply examines which regions of the Ramachandran map the different residues populate.

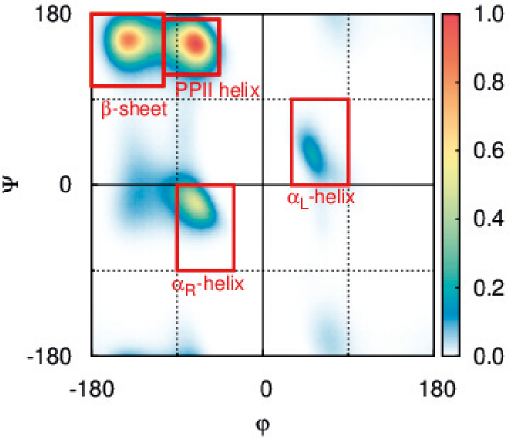


Figure 4: Example of a Ramachandran plot illustrating the four different secondary structure regions analyzed in this study. The plot is normalized for a maximum intensity of 1.

3 Results and discussion

The results of this study are divided into two sections. In the first part, the two Hst5 variants have been simulated as single chains using atomistic MD simulations, with the aim of examining any conformational differences between the two variants. The second part of this section concerns the coarse-grained MC simulations of the Hst5 variants in the vicinity of negatively charged surfaces. For these simulations, the trajectories of the previous MD simulations were used to create structure libraries so that the coarse-grained MC conformations could be reconstructed to atomistic conformations. By doing this, it was possible to obtain and analyze the backbone dihedral angles of the peptides from the coarse-grained MC simulations. However, the main goal of this part was to investigate and highlight the effect of HIS protonation on the surface adsorption of Hst5.

3.1 Single chain properties from MD simulations

The MD simulations revealed that there are obvious differences between the Hst5 variants. The average radius of gyration was found to be 13.5±0.2 Å for Hst5_HIE, whereas it was found to extend to 17.7±0.2 Å for Hst5_HIP. Figure 5a shows the distance distributions of the radius of gyration, which have similar appearance and are partially overlapping, although being centered around different values. No obvious difference between the appearance of the peaks could be identified by visual inspection only, but by performing a Gaussian fit (see Figure S2 in the Supporting Information), the FWHM of the curves were estimated to be 5.4 Å for Hst5_HIE and 5.1 Å for Hst5_HIP. This analysis of the distance distributions and the FWHM discloses that HIS protonation definitely has a large effect on the conformational ensemble as it causes it to become more extended, with a significantly larger radius of gyration. The difference in FWHM could indicate that the flexibility decreases as well, but because the observed difference was quite small, and because the Gaussian fit is not perfect for these distributions, the FWHM difference should be taken with a grain of salt.

Aligned representative cartoon structures of the two Hst5 variants are depicted in Figure 5b. The RMSD of the aligned atoms was 8.88 nm. These structures represent the largest conformation cluster of each individual trajectory, which comprised 14.1 % of the structures of Hst5_HIE, and 40.1 % of the structures of Hst5_HIP while using the same cutoff of 0.50 nm for both peptides. Different cutoffs were tested in intervals of 0.5 Å, and 5.0 Å was chosen based on that the top eight conformation clusters of Hst5_HIE together comprise values as close to 50 % as possible. Thus, with the chosen cutoff, the top eight conformation clusters of Hst5_HIE comprised 56.6 % of the simulated structures, whereas the top eight conformation clusters of Hst5_HIP comprised 85.2 % of the simulated structures. These results strongly indicate that the flexibility of Hst5 decreases upon protonation of the HIS residues.

To further investigate the difference between the two states of Hst5, the secondary structure was analyzed through the simulated backbone dihedral angles. Figure 6 shows the

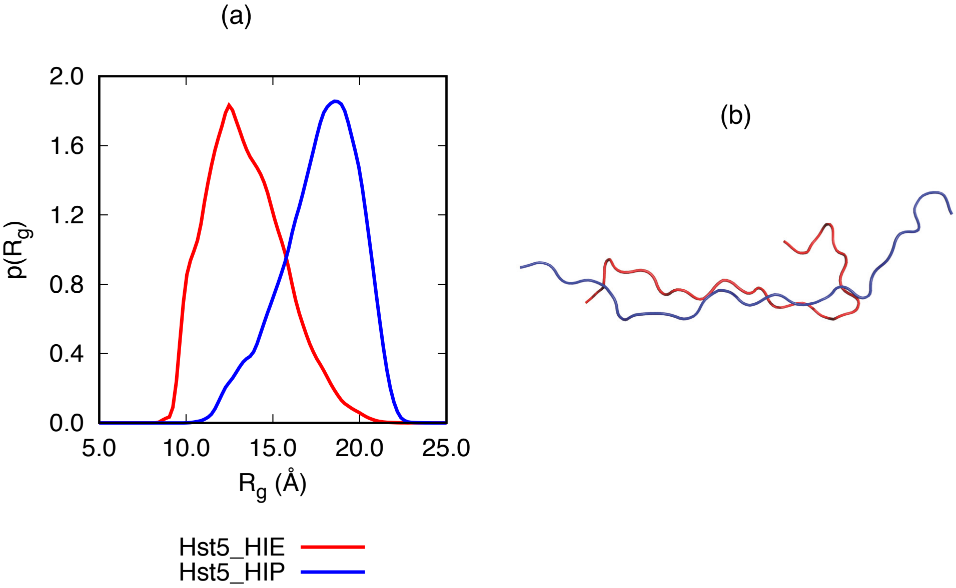


Figure 5: (a) Distance distribution of the radius of gyration, and (b) cartoon representations of the largest cluster conformer of the two Hst5 variants. Hst5_HIE is shown in red and Hst5_HIP in blue.

average secondary structure content of the two Hst5 variants, as obtained from each sampled backbone dihedral angle from the simulations. While both Hst5 variants displayed significant amounts of PPII structure, the PPII content was found to be more than 10 percentage points larger in Hst5_HIP. Additionally, a large difference was observed for the α-helical content, which was found to be almost three times larger when the HIS residues were neutral compared to when they were protonated. Hst5_HIE also had slightly higher random coil content and lower β-sheet content than Hst5_HIP. Thus, by protonating the HIS residues, it seems like the protein can adopt more PPII-helical structures at the expense of α-helical and random coil structures. An increase of PPII structure and a decrease of random coil structure might indicate more extended conformations with less flexibility, which was also implied by the previous results.

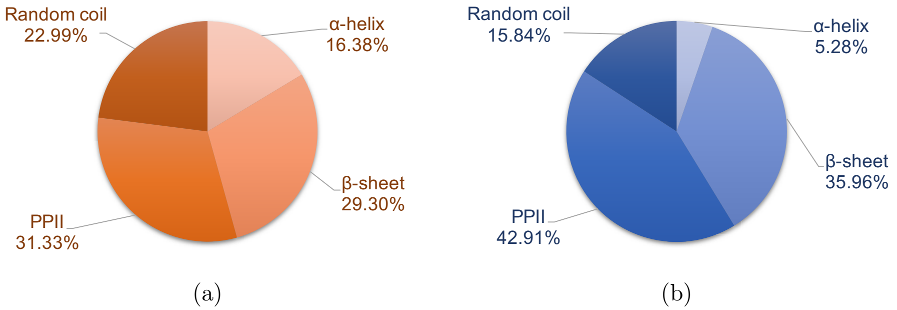


Figure 6: The secondary structure content of (a) Hst5_HIE and (b) Hst5_HIP from MD simulations.

3.2 The effect of HIS protonation on surface adsorption

MC simulations of Hst5_HIE and Hst5_HIP in the vicinity of a negatively charged surface were performed in systems with two different salt concentrations: 10 mM and 100 mM. Figure 7 shows the distance distribution of the radius of gyration of the two peptides from the MC simulations at 10 mM salt compared to the previous distributions from the MD simulations. As clearly visible in the figure, the sampled sizes of the MC simulations correspond well to those sampled by the MD simulations. The average radius of gyration of the adsorbed Hst5_HIE in 10 mM salt was 12.9 Å, and the corresponding value for Hst5_HIP was 17.4 Å, which is slightly smaller than what was obtained from the MD simulations. The FWHM of the curves were estimated to 5.2 Å and 6.3 Å for Hst5_HIE and Hst5_HIP respectively. Thus, the latter variant seems to be more flexible, which is contrary to what was found in the MD simulations.

The secondary structure content of Hst5_HIE and Hst5_HIP in the vicinity of a negatively charged surface is depicted in Figure 8. These two figures display basically the same values and differences that were observed when comparing the two peptides from the MD simulations (Figure 6).

Three regions were defined for determining adsorption: (A) bulk region ($A > 30$ Å from

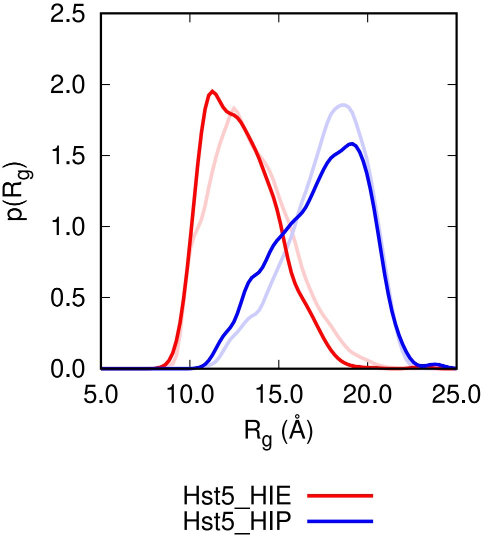
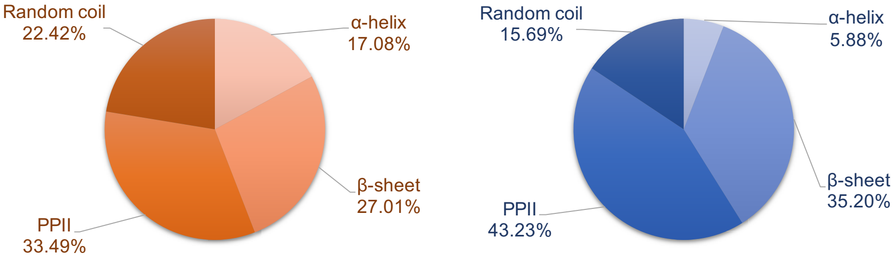


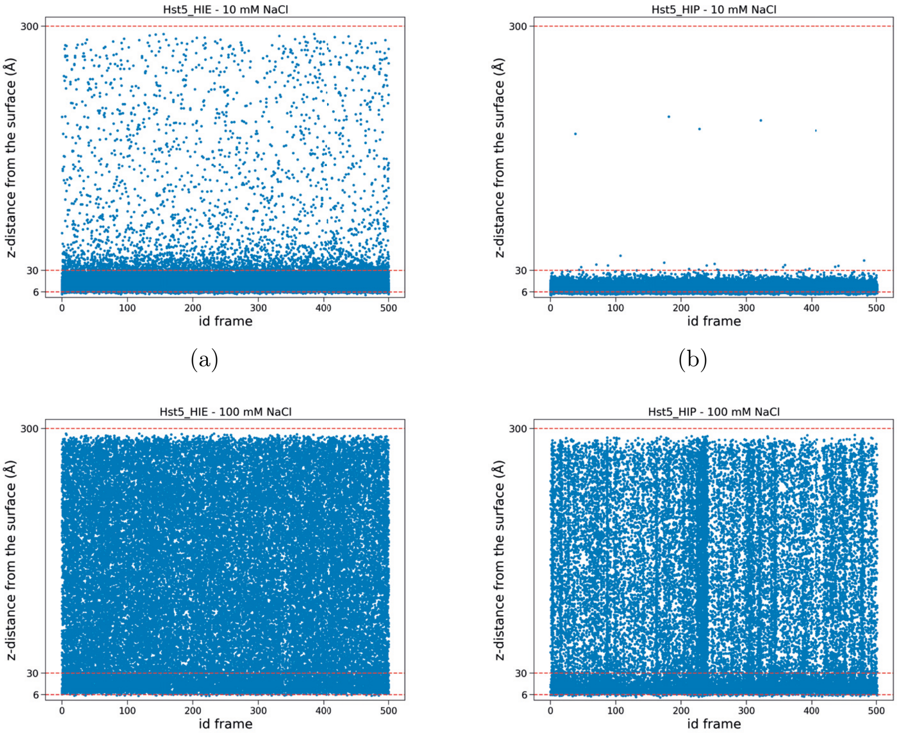
Figure 7: Distance distribution of the radius of gyration of Hst5_HIE (red) and Hst5_HIP (blue) from the MC simulations with 10 mM salt. The curves are compared to their corresponding MD curves from Figure 5, which are depicted in lighter shades.



(a) (b)

Figure 8: The secondary structure content of (a) Hst5_HIE and (b) Hst5_HIP from MC simulations with 10 mM salt.

the surface), (B) intermediate region ($6\,\text{Å} < B \leq 30\,\text{Å}$ from the surface), and (C) adsorbed region ($C \leq 6\,\text{Å}$ from the surface). A peptide was considered to belong to a region if its center of mass was found within the limits of that region. This analysis was done for each registered frame in the simulations, see Figure 9.



(a) (b)

(c) (d)

Figure 9: The distance, z, from the surface of the center of mass of individual frames from the MC simulations (blue dots). The upper row depicts (a) Hst5_HIE and (b) Hst5_HIP at 10 mM salt, and the lower row depicts (c) Hst5_HIE and (d) Hst5_HIP at 100 mM. The limits defining the three different regions are included as red lines in each figure.

By comparing the peptide properties at different salt concentrations, no significant conformational differences were observed for Hst5_HIE. For Hst5_HIP, the average radius of gyration decreased by 1.1 Å when the salt concentration was increased 10 mM from to 100 mM. Additionally, the secondary structure content did not change significantly based on the salt concentration. The effect of salt concentration is summarized in Table S1 in the Supporting

Information.

The number of frames belonging to the different regions was found to be affected by the salt content, see Table 1. For both Hst5 variants, more frames were found in region B and C at the lower salt concentration, and the content of region A was higher when the salt content was high. This is no surprise since an increased ionic strength leads to screening of the electrostatic interactions between the positive peptides and the negative surface. Interestingly enough, both Hst5 variants most often found in region B when the salt concentration was low, although the content of this region was still fairly high at the higher salt concentration.

Table 1: Percentage content of frames of each region for the two Hst5 variants in systems with salt concentrations of both 10 mM and 100 mM.

Region	10 mM		100 mM	
	Hst5_HIE	Hst5_HIP	Hst5_HIE	Hst5_HIP
A	7.10	0.06	84.03	63.10
B	91.94	92.34	15.89	36.57
C	0.96	7.60	0.08	0.33

The conformational properties were also compared over the three regions at 10 mM salt. Considering the average radius of gyration and the secondary structure content per region, no differences were found for Hst5_HIE. However, a small increase of the radius of gyration was observed for Hst5_HIP when going from bulk toward the surface: (A) 17.1 Å, (B) 17.3 Å, and (C) 17.9 Å. Because there are only three average-based data points, this apparent trend is not statistically significant. Additionally, the average for region A is only based on 29 frames (= 0.06 % of the total number of frames), which makes it even less reliable. At this salt concentration there was also a small difference in the secondary structure content. Again, these small differences cannot be considered statistically reliable since there were only 29 frames sampled in region A. Thus, this study is somewhat limited by the sampling of the simulations, as more frames are needed to provide more statistics from the less populated regions of the simulation box.

This study has shown the differences between the two extreme states of Hst5, as well as how the protonation state can affect the surface adsorption. Nevertheless, it would be interesting to further investigate the surface adsorption of Hst5 by introducing titrating moves, that is, by allowing Hst5 to change its protonation state during the simulation. Because the simulated surface is based on silica and because silica also can change its charge depending on the environment, the next step would then be to study the effect of titrating charges on both the amino acids as well as the surface.

next step would be to study the effect of titrating charges on both the amino acids as well as the surface

4 Conclusions

This study was done to elucidate the conformational differences that arise upon protonation of the HIS residues of Hst5. Comparison between the two Hst5 variants Hst5_HIE and Hst5_HIP) from atomistic MD simulations revealed that protonation causes Hst5 to assume more extended conformations. Additionally, it was discovered that the population of the PPII region and the β-sheet region of the Ramachandran map grew larger at the expense of the population of the α-helix region and the remaining uncategorized regions of the Ramachandran map.

Simulations of the two Hst5 variants with coarse-grained MC simulations at the interface of a negatively charged surfaced provided slightly smaller structures as compared to the MD simulations, although the distributions and the secondary structure estimates were basically the same as in the MD simulations. By examining the distances between the surface and the center of mass of the peptides, Hst5_HIP was found closer to the surface to a higher degree than Hst5_HIE, although both variants seemed to prefer to remain within the intermediate region when the salt concentration was low. At high salt concentration, the surface adsorption was diminished, which is attributed to electrostatic screening. Furthermore, there seemed to be a trend of increasing radius of gyration when going across the regions from bulk toward the surface. Further investigation is required to determine if the trend is of statistical significance.

Acknowledgement

This research was funded by the NordForsk's Nordic Neutron Science Programme (grant number 81912). Computer resources for the simulations were provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC).

Supporting Information Available

MD simulation convergence (ACF and BSE); Gaussian fitting to the distance distribution functions; salt effects.

References

(1) Oppenheim, F.; Xu, T.; McMillian, F.; Levitz, S.; Diamond, R.; Offner, G.; Troxler, R. Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on Candida albicans. *J. Biol. Chem.* **1988**, *263*, 7472–7477.

(2) Raj, P. A.; Edgerton, M.; Levine, M. Salivary histatin 5: dependence of sequence, chain length, and helical conformation for candidacidal activity. *J. Biol. Chem.* **1990**, *265*, 3898–3905.

(3) Tsai, H.; Raj, P. A.; Bobek, L. A. Candidacidal activity of recombinant human salivary histatin-5 and variants. *Infect. Immun.* **1996**, *64*, 5000–5007.

(4) Tsai, H.; Bobek, L. Human salivary histatins: promising anti-fungal therapeutic agents. *Crit. Rev. Oral Biol. Med.* **1998**, *9*, 480–497.

(5) Kurut, A.; Henriques, J.; Forsman, J.; Skepö, M.; Lund, M. Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 657–667.

(6) Richardson, C.; Johnsson, M.; Raj, P.; Levine, M.; Nancollas, G. The influence of histatin-5 fragments on the mineralization of hydroxyapatite. *Archives of oral biology* **1993**, *38*, 997–1002.

(7) Yin, A.; Margolis, H.; Grogan, J.; Yao, Y.; Troxler, R.; Oppenheim, F. Physical parameters of hydroxyapatite adsorption and effect on candidacidal activity of histatins. *Archives of oral biology* **2003**, *48*, 361–368.

(8) Yoshinari, M.; Kato, T.; Matsuzaka, K.; Hayakawa, T.; Inoue, T.; Oda, Y.; Okuda, K.; Shimono, M. Adsorption behavior of antimicrobial peptide histatin 5 on PMMA. *Journal of Biomedical Materials Research Part B: Applied Biomaterials: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials* **2006**, *77*, 47–54.

(9) Hyltegren, K.; Nylander, T.; Lund, M.; Skepö, M. Adsorption of the intrinsically disordered saliva protein histatin 5 to silica surfaces. A Monte Carlo simulation and ellipsometry study. *Journal of colloid and interface science* **2016**, *467*, 280–290.

(10) Samoshina, Y.; Nylander, T.; Shubin, V.; Bauer, R.; Eskilsson, K. Equilibrium aspects of polycation adsorption on silica surface: how the adsorbed layer responds to changes in bulk solution. *Langmuir* **2005**, *21*, 5872–5881.

(11) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(12) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(13) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

(14) 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., Bioinf.* **2010**, *78*, 1950–1958.

(15) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(16) Schrödinger LLC, The PyMOL Molecular Graphics System, Version 1.2r1. 2009.

(17) Berendsen, H.; Van Gunsteren, W. Practical algorithms for dynamic simulations. *Molecular-dynamics simulation of statistical-mechanical systems* **1986**, 43–65.

(18) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(19) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(20) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(21) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. (Melville, NY, U. S.)* **1981**, *52*, 7182.

(22) Hyltegren, K.; Polimeni, M.; Skepö, M.; Lund, M. Integrating All-Atom and Coarse-Grained Simulations - Toward Understanding of IDPs at Surfaces. *Journal of Chemical Theory and Computation* **2020**,

(23) Lund, M.; Trulsson, M.; Persson, B. Faunus: An object oriented framework for molecular simulation. 2008.

(24) Stenqvist, B.; Thuresson, A.; Kurut, A.; Vácha, R.; Lund, M. Faunus-A flexible framework for Monte Carlo simulation. *Molecular Simulation* **2013**,

(25) Robinson, R. A.; Stokes, R. H. Tables of osmotic and activity coefficients of electrolytes in aqueous solution at 25° C. *Transactions of the Faraday Society* **1949**,

(26) Cragnell, C.; Rieloff, E.; Skepö, M. Utilizing coarse-grained modeling and monte carlo simulations to evaluate the conformational ensemble of intrinsically disordered proteins and regions. *Journal of molecular biology* **2018**, *430*, 2478–2492.

(27) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.

Supporting Information:

The effect of histidine protonation on the conformational ensemble of histatin 5

Stéphanie Jephthah,* Marco Polimeni, Mikael Lund, and Marie Skepö*

Division of Theoretical Chemistry, Lund University, Lund, Sweden

E-mail: stephanie.jephthah@teokem.lu.se; marie.skepo@teokem.lu.se

Phone: +46-46-222 17 54; +46-46-222 33 66

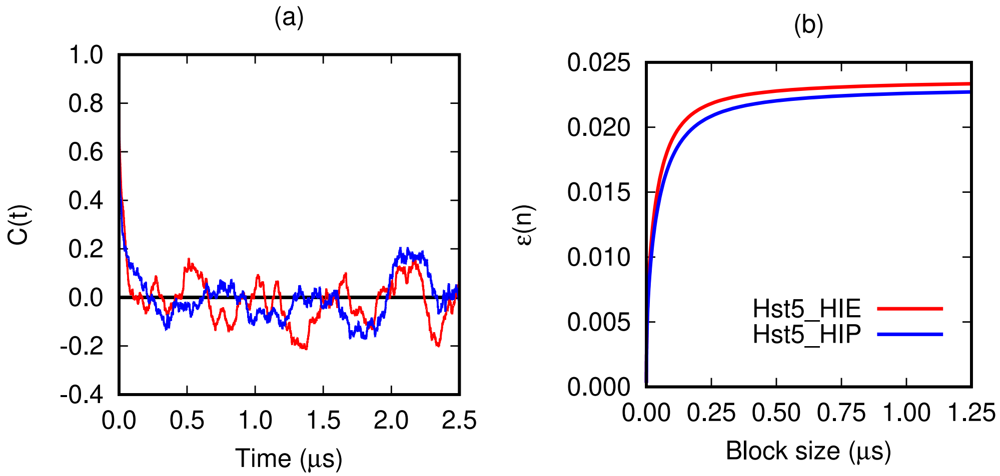


Figure S1: (a) ACF and (b) BSE of Hst5_HIE (red) and Hst5_HIP (blue).

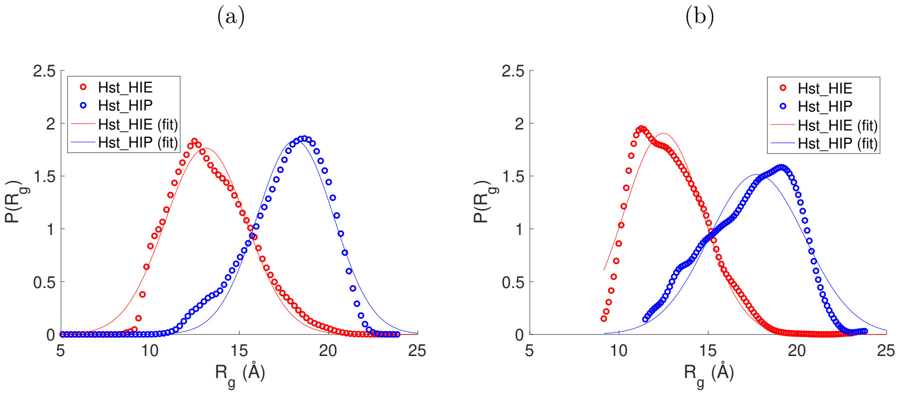


Figure S2: Gaussian fits of the distance distributions of Hst5_HIE (red) and Hst5_HIP (blue) from (a) the MD simulations and (b) the MC simulations simulations.

Table S1: The radius of gyration, as well as secondary structure content of Hst5_HIE and Hst5_HIP at salt concentrations of 10 mM nd 100 mM.

Peptide	C (mM)	R_g (Å)	α-helix (%)	β-sheet (%)	PPII (%)	RC (%)
Hst5_HIE	10	12.9	17.08	27.01	33.49	22.42
	100	13.0	16.97	27.18	33.58	22.27
Hst5_HIP	10	17.4	5.88	35.20	43.23	15.69
	100	16.3	7.30	34.15	42.22	16.33