



MISMaP



CeNT

Natalia Ostrowska

Effects of macromolecular crowding on the activity and dynamics of the NS3/4A protease from the hepatitis C virus. The synergy between experiments and molecular dynamics simulations.

PhD Dissertation

Supervisors:

prof. dr hab. Joanna Trylska

Centre of New Technologies,
University of Warsaw

prof. dr hab. Bogdan Lesyng

Faculty of Physics,
University of Warsaw

April 2022

Acknowledgements

I would like to acknowledge my advisors, prof. Joanna Trylska and prof. Bogdan Lesyng, who helped me with the process of generating this thesis.

I would like to thank prof. Joanna Trylska, for being the greatest group leader! For the whole five years I felt lucky to be a part of her research team.

I would like to thank prof. Michael Feig, who is a co-author of nearly all my publications, for invaluable expert assistance in conducting and analysing my simulations.

I would like to thank dr Monika Wojciechowska, for her support and huge help with the laboratory experiments, and dr Agnieszka Popielec, for numerous discussions as we co-authored my first research publication. I would also like to thank the great crew of Biomolecular Machines Laboratory: Asia Miskiewicz, Julia Macyszyn, Włada Tsylents, Małgorzata Lobka Ksenia Maximova, Agnieszka Markowska-Zagrajek, Marcin Równicki, Kuba Wojtczak, Michał Burmistrz, Piotr Chyży, and my roommates Tomek Pieńko and Jan Stępień.

Specjalne podziękowania dla mojego męża Piotrka i synka Stasia!

A także dla mojej mamy Danusi,
oraz teściów Ani i Jacka – bez waszej pomocy przygotowanie tej pracy zajęłoby dziesięć lat.

Projects and funding

Studies on the effects of crowded environment on activity and dynamics of the NS3/4A protease from hepatitis C virus were carried out within the OPUS project funded by National Science Centre of Poland (grant ID: UMO-2016/23/B/NZ1/03198)



Molecular dynamics simulations described in this thesis were performed at the Centre of New Technologies (University of Warsaw), Interdisciplinary Centre for Mathematical and Computational Modelling (University of Warsaw) and at the University California San Diego.

At the Interdisciplinary Centre for Mathematical and Computational Modelling, the simulations were carried out within computational grants GB73-3 and GB77-7.

List of publications

This thesis is based on following publications:

- N. Ostrowska, M. Feig and J. Trylska, Modeling crowded environment in molecular simulations. *Frontiers in Molecular Biosciences* 6:86, 2019
- A. Popielec, N. Ostrowska, M. Wojciechowska, M. Feig and J. Trylska, Crowded environment affects the activity and inhibition of the NS3/4A protease. *Biochimie* 176:169-180, 2020
- N. Ostrowska, M. Feig and J. Trylska, Crowding affects structural dynamics and contributes to membrane association of the NS3/4A complex. *Biophysical Journal* 120(17):3795-3806, 2021

The thesis contains also unpublished results, which include analysis of the simulations with polysucrose crowders and parameterisation of the polysucrose model. Results obtained from analogical simulations with improved polysucrose model are planned to be published in 2023.

Another project in which I took part was focused on characterising substrate specificity of the Nudt16 hydrolase. I contributed by performing and analysing molecular docking experiments.

- M. Chrabaszczewska, M. Winiewska-Szajewska, N. Ostrowska, E. Bojarska, J. Stepinski, L. Mancewicz, M. Lukaszewicz, J. Trylska, M. Taube, M. Kozak, E. Darzynkiewicz, R. Grzela, Insight into the Binding and Hydrolytic Preferences of hNudt16 Based on Nucleotide Diphosphate Substrates. *International Journal of Molecular Sciences* 22(20):10929, 2021

Results regarding the effects of crowded environment on the NS3/4A protease from hepatitis C virus were presented on scientific conferences, including:

- Hands-on Workshop on Computational Biophysics, University of Pittsburgh, USA, 21-25 V 2018 (presentation)
- Fourth Polish-Korean Conference on Protein Folding: Theoretical and Experimental Approaches, Iława, 9-13 IX 2018 (poster)
- Multiscale Modeling from Macromolecules to Cell: Opportunities and Challenges of Biomolecular Simulations, CECAM Workshop in Lausanne, Switzerland, 4-6 II 2019 (presentation)
- Supercomputing Frontiers Europe 2019, Warsaw, 12-14 III 2019 (presentation)

Abstract

As a standard, both laboratory experiments and simulations of proteins, especially enzymes, are conducted in diluted solutions like buffers or water with ions. However the cytoplasm of a living cell, where the proteins function in nature, provides quite a different environment, with around 30% of the volume taken by macromolecules like RNA, proteins and lipids. Such level of crowding can affect diffusion and conformational dynamics of the proteins, and thus also indirectly affect their activity.

In this thesis I describe investigation of how does crowded environment affect the NS3/4A protease from Hepatitis C virus, which is an example of a protein containing an intrinsically disordered region (IDPr). Such proteins make up almost a half of all proteins, and even larger proportion of viral enzymes. Viruses exploit IDPr to widen a range of functionalities within a very limited number of proteins, by binding to and taking over cellular structures and proteins of the host. Unstructured protein fragments gain more ordered conformations upon approaching binding partners or under changing environment, and so they might be especially prone to be affected by crowding.

To better understand the molecular mechanisms of the effects of crowding on NS3/4A, I combined analysis of laboratory experiments and simulations conducted in analogical conditions. In both cases the protease was observed in water and in the presence of polyethylene-glycol (PEG) and polysucrose. These polymers are among the most often used crowder molecules in the laboratory assays, but hardly ever included in molecular dynamics (MD) simulations. In order to divide observed effects into those caused by volume exclusion and by protein-crowder interactions, I additionally simulated the protease surrounded with coarse-grained crowders of my design.

Thanks to combining observations in two scales - of billions of proteins in the activity assay and one protein in the simulations, I was able to propose a molecular mechanism behind the opposite effects of PEG and Ficoll on the NS3/4A activity, and discover that crowded environment may influence biological functions of the protease, by stabilising a precursor of transmembrane helix at the terminus of NS4A protein - this way facilitating the process of anchoring the protease to the host membrane. I also found that polysucrose crowders, widely thought to be an inert, non-interacting crowder, actually engage in multiple contacts and hydrogen bonds with the studied protein.

New technical methods described in the thesis include parameterisation of all-atom PEG and polysucrose models, that maintain realistic diffusion in the simulations of multiple molecules, and also a design of a spherical bead-shell coarse-grained crowder, that can be used in multi-scale MD simulations with explicit water.

Streszczenie w języku polskim

Zarówno eksperymenty biochemiczne jak i symulacje białek prowadzone są zwykle w rozcieńczonych roztworach takich jak bufory lub woda z jonami soli. Tymczasem naturalne środowisko w którym funkcjonują białka stanowi zupełnie inne środowisko; na przykład cytoplazma żywych komórek w 30-40% zajęta jest przez makromolekuły takie jak RNA i białka. Taki poziom zatłoczenia może wpływać na dyfuzję i dynamikę konformacyjną białek, a więc również pośrednio na ich aktywność enzymatyczną.

W tej pracy opisuję badania wpływu zatłoczonego środowiska na proteazę NS3/4A z wirusa zapalenia wątroby typu C (HCV), która jest przykładem białka posiadającego wewnętrznie nieustrukturyzowane fragmenty. Takie białka stanowią prawie połowę wszystkich białek, i nawet większą część enzymów wirusów, które wykorzystują te fragmenty do wiązania i przejmowania kontroli nad strukturami komórkowymi i enzymami infektowanego organizmu. Nieustrukturyzowane fragmenty białek zyskują bardziej uporządkowane konformacje w kontakcie z makromolekułami z którymi oddziałują, lub pod wpływem zmian w otaczającym je środowisku. Z tego powodu mogą być one szczególnie podatne na wpływ zatłoczenia molekularnego.

Aby lepiej zrozumieć molekularne mechanizmy wpływu zatłoczenia na NS3/4A, połączyłam analizę eksperymentów laboratoryjnych i symulacji prowadzonych w analogicznych warunkach. W obu przypadkach proteaza była obserwowana w wodzie, oraz w obecności glikolu polietylenowego (PEG) i polisacharozy. Oba polimery należą do cząsteczek najczęściej wykorzystywanych do modelowania zatłoczonego środowiska w doświadczeniach laboratoryjnych, ale prawie nigdy nie są włączane do symulacji dynamiki molekularnej (MD). Aby podzielić obserwowane efekty na te spowodowane wyłącznie przez ograniczenie przestrzeni, i te wynikające z oddziaływania z cząsteczkami modelującymi zatłoczone środowisko (crowderami), dodatkowo przeprowadziłam symulacje proteazy otoczonej gruboziarnistymi pseudo-cząsteczkami, które oddziałują z białkiem w minimalnym stopniu.

Dzięki połączeniu obserwacji układu w dwóch skalach – miliardów białek w eksperymentach biofizycznych i jednego białka w symulacjach, udało się zaproponować mechanizm molekularny odpowiedzialny za bardzo różny wpływ PEG i polisacharozy na aktywność NS3/4A. Symulacje MD w wodzie i w otoczeniu różnych crowderów wykazały również, że zatłoczone środowisko może wpływać na funkcje biologiczne proteazy. Dzięki stabilizacji prekursora helisy transbłonowej na N-końcowym fragmencie NS4A, zatłoczone środowisko wewnątrz komórki ludzkiej ułatwia proces zakotwiczenia kompleksu NS3/4A w błonie lipidowej.

Niniejsza praca opisuje jedną z pierwszych symulacji MD białka w otoczeniu pełnoatomowych crowderów PEG, oraz pierwszą symulację białka w otoczeniu polisacharozy reprezentującej wykorzystany w eksperymentach Ficoll. Ten polimer powszechnie uważany jest za nie wykazujący oddziaływań z białkami, jednakże w trajektoriach uzyskanych z modelem polisacharozy można zauważyć wiele kontaktów i wiązań wodorowych między crowderami a badanym białkiem.

Nowe techniczne metody opisane w pracy obejmują parametryzację pełnoatomowych modeli polimerów PEG i polisacharozy w taki sposób, aby zapobiec ich agregacji w symulacji wielu molekuł, a także opisanie modelu sferycznych gruboziarnistych crowderów, które mogą być wykorzystane w wieloskalowych symulacjach z wodą i białkiem w reprezentacji pełnoatomowej.

Contents

1	Introduction	11
1.1	Life inside the cell	11
1.2	The consequences of crowding – theory	14
1.3	Measuring the effects of crowding	17
1.3.1	Special case of intrinsically disordered proteins	19
1.4	The NS3/4A protease	20
1.5	Aims and Objectives	24
2	Methods	26
2.1	Computational theoretical methods	26
2.1.1	The CHARMM Force Field	31
2.1.2	Theory of trajectory analysis	34
2.2	Simulation Methods	35
2.2.1	Building the NS3/4A model	35
2.2.2	Building the peptide models	36
2.2.3	Preparation of the simulation systems	37
2.2.4	Parameterisation	41
2.2.5	Water optimization	41
2.2.6	Simulation conditions	42
2.2.7	Trajectory analysis	43
2.3	Experimental methods	48
2.3.1	Materials	48
2.3.2	Peptide substrate synthesis	49
2.3.3	Activity assay	50
2.3.4	CD - measuring peptide helicity	53
3	Crowders	55
3.1	Three models of crowder molecules	55
3.2	All-atom PEG crowders	56
3.2.1	Molecular models of PEG - topology and parameterisation	57
3.2.2	Aggregation of PEG molecules	59

3.2.3	Modification of force field parameters of water restores realistic diffusion of PEG crowders	60
3.3	The coarse-grained Crowders	62
3.3.1	Single sphere crowders	63
3.3.2	12+1 crowder model	66
3.3.3	42+1 bead-shell crowders	66
3.3.4	Artificial hydrophilicity of the CG crowders	69
3.3.5	Final model of the CG crowder	71
3.3.6	Adding crowders can actually speed up MD simulation	73
3.4	Ficoll crowders	73
3.4.1	Parameters of the model	74
3.4.2	The concentration problem	75
3.5	Dynamic properties of the crowders	76
3.6	Discussion	85
3.6.1	Non-standard crowder choice	85
3.6.2	Ficoll or polysucrose?	86
3.6.3	The coarse-grained crowders	88
3.6.4	Comparison of dynamic properties of the crowders	90
4	Effects of crowders on the NS3/4A protease	93
4.1	Impact of crowding on the activity of the NS3/4A protease	93
4.1.1	Experimental measurement of NS3/4A protease activity - the general idea . .	93
4.1.2	Activity assay results	95
4.1.3	Molecular dynamics simulations of the NS3/4A protease in water and in crowded environment	97
4.1.4	Crowding affects rotations and translational diffusion of the protease	98
4.1.5	Simulations with the peptide substrates	101
4.2	Structural dynamics of the NS3/4A protease in crowded environment	106
4.2.1	Structural flexibility of the NS3 protease	106
4.2.2	Mobility of the N- and C-terminal tails of the NS4A protein and their interactions with the NS3 surface in two water models	110
4.2.3	Mobility of the N- and C-terminal tails of the NS4A protein and their interactions with the NS3 surface in different environments	110
4.2.4	The C-terminal tail of NS4A	114
4.2.5	The N-terminal tail of NS4A	115
4.3	Crowding stabilizes a helix formed on the N-terminal tail of NS4A	121
4.4	Crowding might contribute to membrane association of the NS3/4A complex	125
4.5	Discussion	127
4.5.1	Test-tube versus MD, in numbers	127

4.5.2	Explaining experimental results through MD simulations	128
4.5.3	PEG and Ficoll as crowders, in laboratory experiments and MD	129
4.5.4	Stabilisation of the NS4A helix in a crowded environment	130
4.5.5	The role of crowding in biological function of the protease - potential consequences for the viral life cycle	131
5	Summary and future perspectives	134

Chapter 1

Introduction

1.1 Life inside the cell

The inside of a living cell is not water with occasionally occurring molecules. It looks more like a thick soup, in which it is not easy to find an unoccupied space to accommodate another macromolecule. In addition to organelles, networks of membranes and filaments, the cytoplasm is crowded with molecules and macromolecules such as ribosomes, proteins, nucleic acids, lipids and metabolites, which together occupy up to 30 – 40% of the available space¹ (Figure 1.1). The largest and most functionally diverse group of crowdors are proteins, which in the cytoplasm play the role of machines catalysing chemical reactions, transporters and transmitters of signals about the change of the state of the cell and organism.

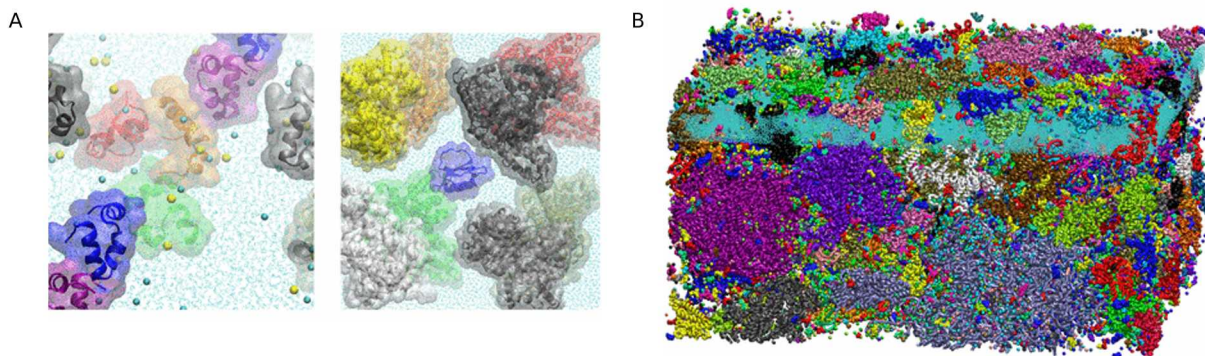


Figure 1.1: Crowded environment of bacterial cytoplasm from a computational model. Figure prepared based on the work of M. Feig et al.²

Proteins consist of several hundred to several thousand amino acids linked together in an order unique to each protein. Despite the small diversity of their elements – in case of most proteins only 20 basic amino acids – there are up to 400,000 different proteins in the human body, each performing an unique specific task.

The final structure and function of a protein in a deterministic way depends on its most basic structure, i.e. the order of the amino acids making up the protein chain. This is the so-called protein

primary structure (Figure 1.2). Aiming for an energy minimum, the amino acid chains first fold into helices and beta strands – components of the secondary structure, from which more complex structural motifs and complexes of individual protein domains arise. Formation of each of these structures results from interactions between amino acids, such as hydrogen bonds, electrostatic, hydrophobic, and even covalent interactions between the amino acid side chains.

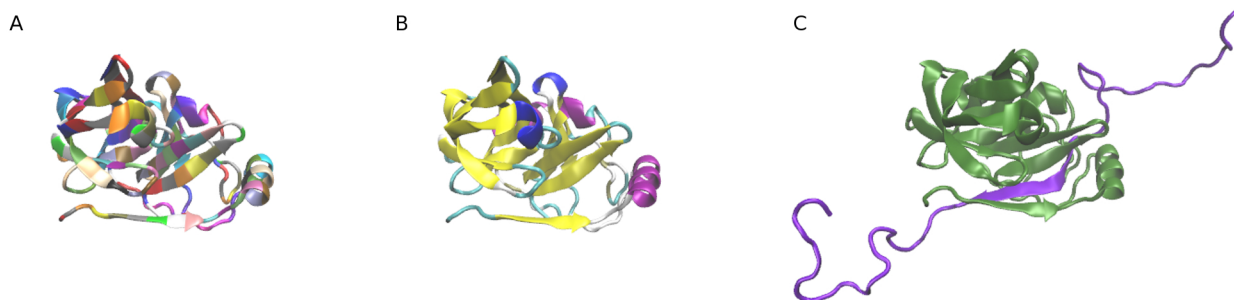


Figure 1.2: Primary, secondary and quaternary structure of a protein illustrated on an example of NS3 protease domain and NS3/4A protease from HCV. Primary structure (A) refers to a sequence of amino acids in the protein chain. Secondary structure (B) refers to basic structures like α -helices (in pink) or β -strands (in yellow), which are held together by a network of hydrogen bonds between backbone atoms. While all three proteins are folded into their specific three-dimensional structure, therefore maintaining their tertiary structure, figure C shows an additional feature, i.e. formation of a functional complex by interaction between NS3 and NS4A proteins, which constitute a quaternary structure of the NS3/4A protease.

The biological functions of proteins range from building blocks of the intracellular structures, through transporting metabolites and ions across biological membranes, or binding antigens that threaten the cell, to catalysis of metabolic processes. Within the latter, each protein catalyses a specific chemical reaction. Thanks to its unique spatial configuration and charge distribution, such proteins are able to help form or break chemical bonds in the molecules being processed, lowering the energy needed for the reaction to take place.

In each case, the ability of proteins to perform their functions depends on the interactions with other molecules – components of protein complexes, processed substrates or molecules that modify the structure and activity of the protein. Interactions with partners are strictly dependent on the characteristics of the protein's surface, especially in the ligand binding region or the active site, which includes amino acids directly involved in the catalytic process. This region, specifically dedicated to substrate processing, is often located in a pocket inside the protein structure, able to perfectly accommodate the ligand.

Each atom building a protein residue can be described by spatial coordinates x , y , z and a partial charge – in this sense mathematically they can be represented by vectors of four dimensions. The charge distribution results in another property of the protein surface – polarity or hydrophobicity determined by dipole moments of the amino acid side chains. These properties, in addition to charge-charge interactions with ligand molecules, also largely determine the overall structure of the protein, packing hydrophobic residues inwards and exposing hydrophilic ones towards the solvent.

Spatial arrangement and partial charges on the protein surface create a four-dimensional map, unique for each protein, determining interactions with its partners – molecules that, in order for the interaction to occur, must fit this map, both spatially and in terms of charge distribution. The interactions take place when the protein and ligand – both undergoing random-walk diffusion – find each other in the crowded cytoplasm. Proteins in this process, due to their mass and occupied volume (average mass in the order of tens of thousands of g/mole and the radius of 10 to 50 Å³), remain practically stationary in relation to smaller, more mobile ligands. A typical translational diffusion rate coefficient of a protein in cytoplasm is between 1 and 10 $\mu\text{m}^2/\text{s}$ ³ while for the example metabolite (lactate) it is 210 $\mu\text{m}^2/\text{s}$ ³. Relative to the macroscopic components of the cell these are huge values, therefore, despite numerous non-specific contacts with molecules in the crowded cytoplasm, ligand are observed to find its partners with surprising frequency.

Biochemical reactions in cytoplasm depend on the constant movement of all its elements. The fastest movements in the molecule are oscillations of the bond lengths between atoms present in molecules (with frequencies of the order of 10^{14}Hz ³), as well as the oscillations of angles and dihedral angles between individual pairs of bonds. At the scale of a single (macro)molecule, we observe rotational movements, i.e. rotations of the molecule around its axis, and translational diffusion, i.e. the movement of a molecule from one location to another. Most often, this movement has the character of random position fluctuations closely related to the thermal energy of the molecular system; its consequences are collisions with other molecules, including water molecules, in which kinetic energy is transferred to other elements of the system. A high degree of randomness in the directions of movement gives the impression that the interior of the cell is in chaos. However, due to the fact that the movements of molecules are extremely fast (with diffusion coefficients of the order of hundreds $\mu\text{m}^2/\text{s}$ ³) and the collisions between them occur with a high frequency (with over 100000 productive collisions per Mole per second³), the protein is able to process, and therefore also meet up to several thousand substrates during every second⁴.

In addition to random movements and collisions with neighboring molecules, Van der Waals (VdW) and electrostatic interactions help attract the substrate to the binding site of the corresponding enzyme. VdW interactions cause pairs of atoms to attract each other, the stronger the closer the distance between them; at a distance, slightly greater than the sum of the radii of the atoms, the mutual attraction is the strongest, while when this distance is exceeded, the atoms are strongly repelled as they start to overlap (Figure 1.3). The electrostatic interaction causes atoms to be attracted or repelled, depending on their charge or partial charge. In close vicinity of the active site of the enzyme, where the charge distribution is perfectly matched with the substrate, electrostatic interactions play the role of a parking support system – they help the approaching substrate to attach to the appropriate binding site in optimal conformation.

The protein as well is not a rigid structure – on the contrary, it is in constant motion, so in order to match the ligand also the active site of the protein is able to adjust the conformation of individual amino acids. In more complex systems, the interaction can take place only under appropriate conditions, for example after binding an effector adjusting the protein conformation, or if the substrate

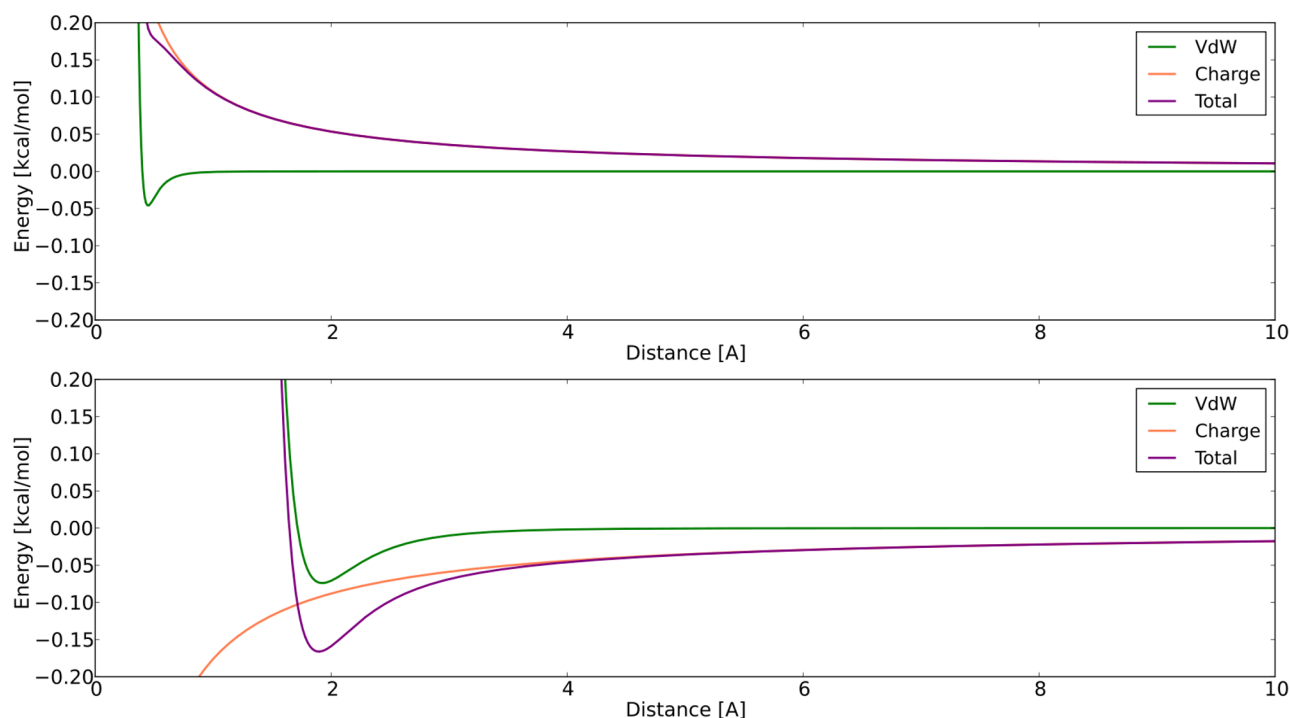


Figure 1.3: Potential energy of interaction between two pairs of nonbonded backbone atoms: oxygen and hydrogen, with partial charges of $-0.51e$ and $0.31e$ (top), and two hydrogens with partial charges of $0.31e$ (bottom).

binding site is at the interface of domains that can move relative to each other, the interaction can only take place at the appropriate configuration of protein components.

In summary, protein activity is determined by two factors: the current structure or conformation of the protein and the rate of diffusion of its own and its substrates – change in either of these factors will have consequences for that activity. As it turns out, placing a protein in a crowded environment is able to affect both of them.

1.2 The consequences of crowding – theory

A significant fraction, reaching 40% of total volume of the cytoplasm is occupied by macromolecules. In practice, it means that biomolecules evolved in an environment where the distance from neighbouring molecules is often smaller than their radius.

Meanwhile, our knowledge of biochemical processes largely relies on experiments carried out in dilute aqueous solutions, mimicking the amino acid protonation states and ionic strength present in the cytoplasm, however, completely ignoring the issue of proteins' surroundings and the extremely limited space available to them. Studying the actual impact of crowding is crucial to verify the results of such experiments and to identify cases where including crowders in the standard experimental procedure is necessary.

It is easiest to imagine the effect of crowding on the diffusion of molecules. Two to even ten-fold

decrease in translational diffusion coefficient is observed, relative to the diffusion in water, showing considerable influence on the diffusion dependent reaction rates⁵.

However in addition to diffusion, crowding can also affect the energy landscape of individual proteins and their complexes. Crowding may change the protein folding path, the size of the population of individual conformations, the stability of proteins and their complexes⁶ – these are changes that often have significant biological consequences, such as changes in the level of activity of individual proteins, and also the way they interact with other molecules, for example drug molecules.

Crowding affects proteins through two mechanisms: **excluded volume effect** and **non-specific interactions** with the crowder molecules.

In a crowded cytoplasm, a substantial fraction of volume is occupied by large and small molecules, therefore making this volume unavailable (or excluded) for other molecules. For an example metabolite, as the available volume is decreased, effective concentration of this molecule will be higher.

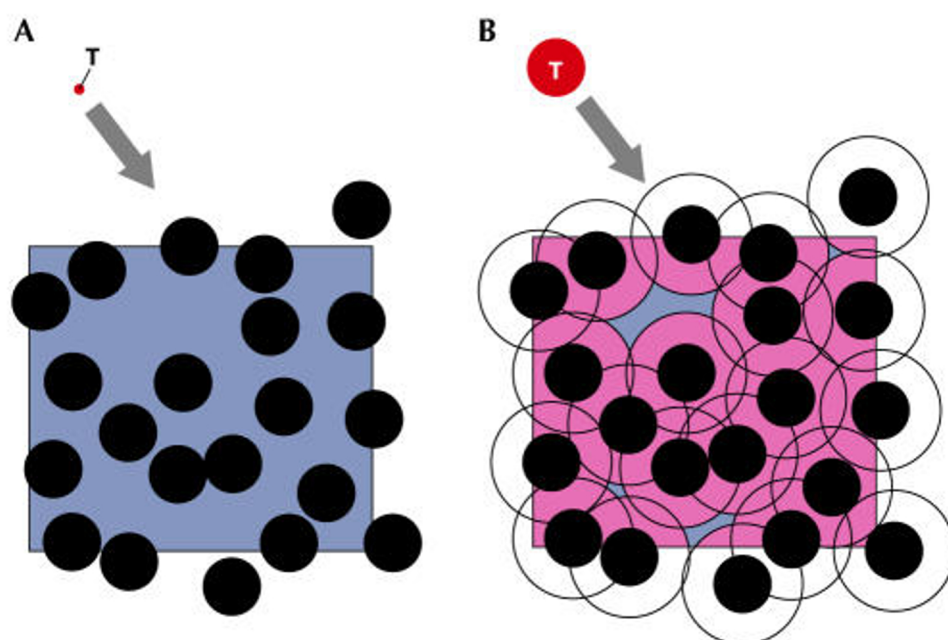


Figure 1.4: Excluded volume (shown in pink and black) and available volume (shown in blue) in a solution of background macromolecules. The two panels show the volume available to the centre of a test molecule (T) either (A) much smaller than or (B) of similar size to the background macromolecules (see the text for details). Figure reproduced from the work of Rivas et al.⁷

A new (macro)molecule can be added to the existing mixture in a very limited number of ways (Figure 1.4) – with each new molecule the entropy of the solution would decrease, to avoid overlap with the cosolutes. As a consequence, the free energy of the system will rise^{8,9}. In an effort to maximize entropy, the system will seek to minimize the excluded volume, both in terms of volume and surface to volume ratio – as a result compact, globular protein conformations would be promoted, as well as oligomeric structures that occupy less space than a number of monomeric proteins.

The increase of free energy resulting from volume exclusion can be described thermodynami-

cally as

$$\mu_i = \mu_i^0 + RT \ln(a_i) = \mu_i^0 + RT \ln(c_i \gamma_i) \quad (1.1)$$

where μ_i is partial molar free energy of the i 'th component, μ_i^0 is its free energy in a standard state and a_i is its thermodynamic activity, here expressed as effective thermodynamic concentration $c_i \gamma_i$, where c_i is the molar concentration of the i 'th component and γ_i is its "activity coefficient" that takes into account its deviation from ideal behaviour arising from volume exclusion. $\ln(\gamma_i)$ can be expanded in powers of the solute(s)' concentrations as:

$$\ln(\gamma_i) = \sum_j a_{ij} c_j + \sum_j \sum_k a_{ijk} c_j c_k + \dots \quad (1.2)$$

with the summations over excluded volume contributions of all interactions present in the system. In this form, a_{ij} can be described as two particle excess excluded volumes, $a_{ij} = U_{ij} - V_j$, where U_{ij} is the volume around particle i , where the center of particle j cannot be placed, and V_j is the molar volume of particle j . $\ln \gamma$ can be approximated as a function of ϕ , a fraction of volume occupied by spherical particles in a fluid of identical hard spheres, where γ is increasing exponentially with the volume fraction.

One of the most significant consequences of volume exclusion is its effect on the equilibria of molecular rearrangements, especially for reactions proceeding with significant conformational change of molecules or with their association. For chemical reactions, equilibrium constants relate thermodynamic activities of substrates and products:

$$K = a_B / a_A = \frac{[B] \gamma_B}{[A] \gamma_A} = K^{app} * \frac{\gamma_B}{\gamma_A} \quad (1.3)$$

K^{app} is the apparent equilibrium constant, determined by concentrations, the one that we observe in experiments

$$K^{app} = \frac{[B]}{[A]} = K \frac{\gamma_A}{\gamma_B} \quad (1.4)$$

this way, the volume excluded by products to the volume excluded by substrates ratio <1 or >1 can affect the equilibrium, shifting it towards products that occupy less space¹⁰. Major predictions based solely on volume exclusion include the preference for compact, globular conformations in crowded solutions and profound effects on association rates. Under volume occupancy between $\phi = 0.1$ and $\phi = 0.5$, that is in conditions found in the cytoplasm, the association rates are predicted to rise by several orders of magnitude, compared to the values measured in dilute solutions. This effect would be greater for association of larger molecules, and increase monotonically with volume occupied by crowder species⁸.

Indeed, it might be enough to restrict the volume available for the macromolecule to exert effect on its conformation, association and reaction rates. But molecules present in the cytoplasm are not inert space-fillers – each of them can interact with the molecule of interest by VdW interactions – (each background molecule weakly attracts the tracer molecule), and also charge-charge and

hydrophobic interactions, being either attractive or repulsive, depending on the charge distribution and hydrophobicity of each molecule's surface and mutual orientation of each pair of interacting molecules.

While effects related to the limitation of available space are associated with changes in the entropy of the system, chemical interactions, called soft interactions, are enthalpic in nature. Crowder-protein repulsion interacts with entropic effects, acting to reduce the excluded volume and in effect stabilize compact structures. At the same time attractive interactions can act in the opposite direction, causing stabilization of extended or unfolded conformations. This way, enthalpic contribution to crowding effect can alter the activity of a protein, increasing the proportion of active or inactive conformations¹¹.

Crowding components also affect diffusion of macromolecules in a different way. While diffusivity reduction caused by volume exclusion affects mainly translational diffusion (except for sterically blocking a non-uniformly shaped molecule), crowder interactions can reduce both translational and rotational diffusion. Weak non-specific interactions can result in forming transient clusters of molecules. Such clusters will diffuse as a group, while significantly reducing rotations of individual molecules^{12,13}.

The effect of crowding on a protein's structure and reactions is a sum of many components: excluded volume effect and non-specific interactions affecting molecule's conformation and association and multi-component effects on diffusion rates. The sum of these effects can be very complex and different for each protein-environment combinations. This is why we lack standardised corrections to the dilute-solution experiments that would include predicted effects from crowding, and each system needs to be studied separately.

1.3 Measuring the effects of crowding

Including the crowd into any experimental procedure, both computational and in vitro, first requires a decision on which model of crowded environment to choose. Although much depends on it, a compromise is often needed between the realism and practicality of a given solution. The natural environment inside the living cell is a complex composition of proteins, RNA and metabolites – all of them fully interacting particles, some of them heavily charged as well. Using equally diverse solution, like for example a cell lysate as the crowded environment would rank highest in terms of realism, but the presence of many types of macromolecules in solution could make it impossible to experimentally measure the properties of individual proteins or reactions of interest. It is more practical to place the studied protein in a solution of homogeneous particles, later called the crowders. A popular practice is to choose crowders that interact with proteins as little as possible, to minimize the contribution of non-specific interactions to the effects on protein behaviour. Such crowders can be well-soluble non-protein polymers such as Ficoll and Dextran (polysucrose and polyglucose) or Polyethylene Glycol (PEG). A more realistic approach involves the use of small proteins, whose properties are similar to molecules found in the cytoplasm, but a significant drawback of this method

is the need to obtain significant amounts of purified proteins.

In case of simulations, on the other hand, the options are virtually unlimited. While it is difficult to find a non-interacting polymer, creating a non-interacting crowder model is not problematic. Such a particle can be obtained by creating pseudo atoms or pseudo particles of any size and shape, from a protein-sized sphere to a chain of beads modelling a polymer. More realistic models of crowded environments include using full-atom or coarse-grained models of crowders used in laboratory experiments, or even placing a protein of interest within a modelled fragment of the cytoplasm.

Crowded environment can be incorporated into a range of biochemical and biophysical experiments, as well as into molecular simulations, to study numerous aspects of protein behaviour:

Enzymatic activity under crowding is typically described with kinetic parameters, measured by monitoring substrate or product content in the reaction mixture. Choice of methods include fluorescence assays and Förster Resonance Energy Transfer (FRET) measurements, that allow to assess changes in reactants' fluorescence during the course of the reaction. They often require attaching fluorophore or chromophore tracking tags to interacting molecules. Alternatively, with the Isothermal Titration Calorimetry (ITC) method, the reaction of unlabelled molecules can be observed by measuring heat changes that occur during the biochemical process.

A flagship example of the effect of crowding on protein activity is DNA ligase¹⁴, which becomes activated when surrounded by molecules such as Ficoll, PEG or Bovine Serum Albumin (BSA). Crowding also leads to enhancement of RNase T1¹⁵, when surrounded by dextran molecules, while both proteins remain inactive in dilute solutions. Another example relates to chaperonins GroEL and GroES, whose efficacy is enhanced upon crowding¹⁶. On the contrary, experimental data show decrease in enzymatic activity of trypsin¹⁷ and HIV I protease¹⁸, in the presence of PEG crowders. As crowding is generally expected to decrease reaction rates due to altered diffusion constants, increased activity is often thought to be implicated by changes in conformational dynamics and stability of the interacting enzyme.

Protein stability is an example of a property that is inversely affected by individual crowding components. This is evidenced by the results of MD simulations in which exclusively repulsive crowder particles can be used to approximate crowding by volume exclusion. In agreement with theoretical predictions, surrounding a protein with repulsive crowders can promote compact conformations¹⁹, stabilise proteins²⁰ and protein-protein complexes^{21,22,23}. In one of the studies¹⁹, compaction of apoflavodoxin under crowding observed in the simulation has been confirmed experimentally with circular dichroism (CD) spectrometry, fluorescence correlation spectroscopy and FRET, measuring changes in secondary structure content, volume and radius of gyration of the protein of interest. However simulations with all-atom crowders show examples of reverse effects. Simulations studying proteins surrounded with protein crowders has shown that the presence of realistic crowder molecules modelled with atomistic detail can actually destabilise the native conformation of some proteins^{24,25}, or even promote local unfolding²⁶. This behaviour can be attributed to the presence of attractive interactions between the studied proteins and protein crowders, which confirms the importance of including the enthalpic component of crowding in experiments.

Structural change, as a relatively rapid process, is well suited to observation through MD simulations that allow monitoring the conformational dynamics processes and observe the mechanics behind it. Laboratory experiments that can study structural dynamics under crowding conditions include, on the other hand, CD spectrometry and NMR, which capture structural information at a specific point in time. The spectrum of possibilities for representing molecular systems in MD simulations range from fully coarse-grained simulations with single pseudo-atom crowders to short experiments with all-atom proteins simulated in atomistic model of a cytoplasm. Simulating a long process of HIV-I protease subdomain movements with spherical crowders and CG protein model²⁷ has shown that the presence of crowders is able to affect the frequency of adopting the active conformation, related to opening and closing of flap-shaped elements of the enzyme. Similar changes in the subdomain positions have been observed for the PDHA enzyme, in the all-atom simulation of the cytoplasm²⁵ and directly attributed to protein-protein interactions with the crowders.

Another category of experiments investigating crowding-induced structural changes are short protein simulations, which are discussed in more detail in the next subsection.

1.3.1 Special case of intrinsically disordered proteins

Intrinsically disordered proteins (IDP) and Intrinsically Disordered Protein Regions (IDPr) comprise of extremely heterogeneous ensemble of conformations that lack stable secondary and/or tertiary structure. Some IDPr transition to more ordered structures upon interaction with their binding partners, like RNA or other protein, other may act as flexible linkers between protein domains or mediate protein-protein interactions.

IDP change structure or fold in response to changing environmental properties such as pH, temperature, polarity or hydrophobicity of the immediate surroundings²⁸, and so they are particularly sensitive to the effects of crowding. Based on theoretical predictions, they should fold upon volume exclusion, yet some remain completely unfolded in crowded conditions, needing another fold-inducing factor like the presence of their natural binding partner²⁹. In vitro and in silico experiments with crowding report wide spectrum of observations^{28,29}. Although compaction is frequently observed, the overall effects of crowded environment highly depend on the sequence of particular IDP, the volume fraction and size of used crowders³⁰.

A good example of IDPs affected by crowding are peptides and their aggregates present in the course of Alzheimer's disease whose structure is often altered and whose aggregation accelerates in crowded milieu³¹. Experiments investigating their structure inside mammalian cells confirm that α -synuclein undergoes some level of compaction, but remains largely unstructured in the cytoplasm, with the aggregation prone component hidden inside the structure³². Compaction of this peptide is greater under Ficoll crowding, than in the presence of BSA – interacting protein crowders³³. Simulations of a fast-folding trp-cage peptide in the presence of two types of protein crowders show another example of diverse effects on conformational dynamics of IDP. The peptide, forming compact structure in dilute solutions, in the presence of interacting crowders adopt broad range of non-native

conformations^{34,35}, with stabilised helical fragment³⁵.

A particularly widespread group of proteins containing IDPr are viral proteins, especially of RNA viruses. One of the strategies, adopted by those viruses to take over components of the host cell, is based on using unstructured fragments of its proteins to interact with host proteins, or to anchor viral proteins to host membranes. In this work, the object of observation is one of such proteins from the Hepatitis C Virus.

1.4 The NS3/4A protease

The Hepatitis C Virus (HCV) genome encodes only ten proteins: three structural proteins, including the building blocks of viral envelope and capsid, and seven nonstructural proteins responsible for production of new virion components and silencing of the cellular immune response (p7, NS2, NS3, NS4a, NS4b, NS5a, NS5b), Figure 1.5. At the very beginning of the HCV infection, viral RNA enters the cell and in the region of endoplasmic reticulum (ER) it is directly transcribed into a polyprotein – all ten proteins linked to one another. To perform further steps of infection, each protein has to be "freed" – its junctions with neighboring proteins have to be cleaved by specific enzymes, one of which is the NS3 protease.



Figure 1.5: Schematic illustration of the genome of Hepatitis C Virus, with the junctions cleaved by the NS3 protease marked.

Some of the freed proteins, like NS3, NS4A, NS4B, NS5 or the HCV RNA polymerase, anchor to the ER membrane, forming membrane embedded replication complex³⁶ with numerous protein-protein interactions³⁷. With access to membranes, human ribosomes, and stores of amino acids and nucleic acids, the endoplasmic reticulum creates excellent conditions for the production of elements of new viral particles. In addition, to protect itself from the cell's immune system and to protect fragile RNA molecules from the activity of endonucleases, the virus replication machinery is encapsulated by a network of membrane vesicles³⁸ formed as a result of viral proteins action. made in a factory like this, viral proteins and replicated RNA copies are assembled into new viral particles.

NS3/4A is a bifunctional protein, a part of which acts as a protease. After performing enzymatic cleavage, the protein switches to helicase activity (Figure 1.6). Those two functions are performed by two spatially separated domains, connected by a linker³⁹. In this work, only the protease domain is considered.

The protease domain of NS3 is a globular protein, with two large beta-barrels in its core. It is a serine protease, cleaving protein-protein junctions with three catalytic aminoacids: His57, Asp (?)

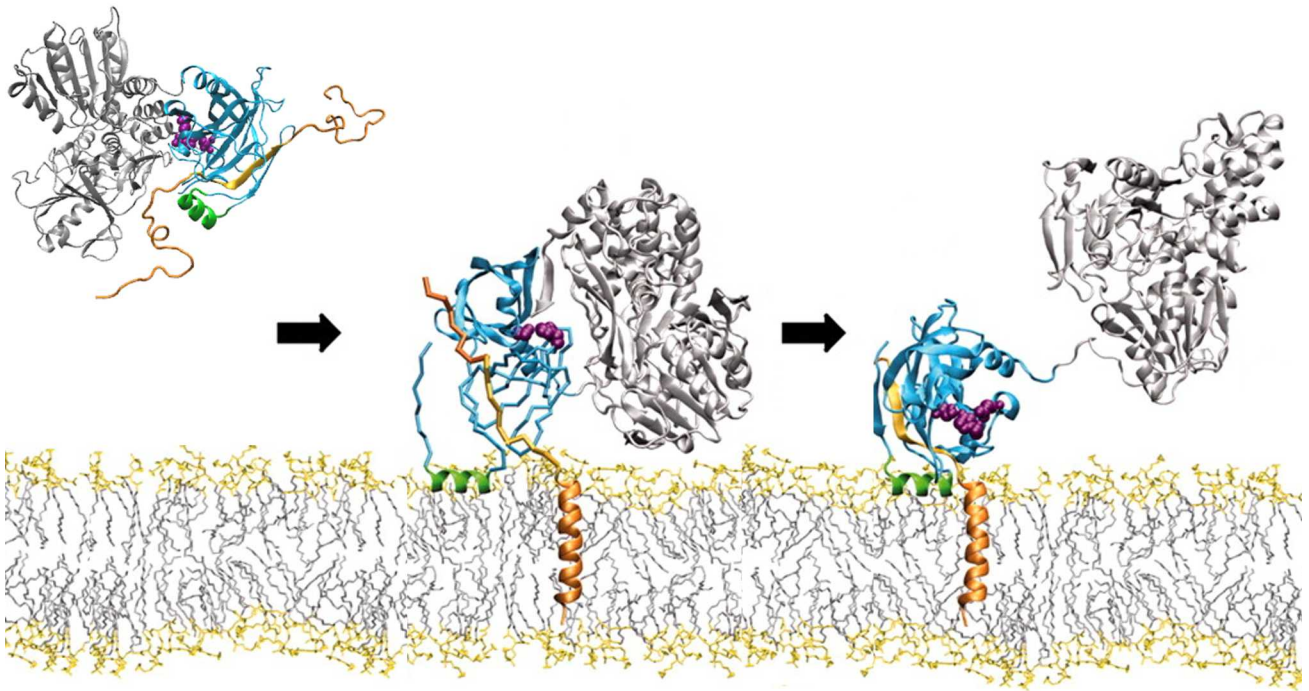


Figure 1.6: Membrane binding process of the NS3/4A protease, as proposed in the work of Brass et al.⁴⁰, illustrated with the structures of the whole protein, with both protease and helicase domains. First, the protease approaches the membrane, with the NS4A cofactor in the partially unstructured conformations, then the complex anchors to the membrane, inserting a 12-24 helix of NS3 and transmembrane helix formed on the N-terminal tail of NS4A. After membrane anchoring, the helicase domain rotates along the domain linker to move away from the membrane interface. The helicase domain of NS3 is shown in gray, the protease domain in blue, with the active site side chains highlighted in purple. The NS4A cofactor is shown in orange, with the central part bound to NS3 highlighted in yellow. Orientation of the membrane is shown with the schematic model of the lipid bilayer. The figure on the left is prepared based on the crystallographic structure from PDB (ID:1CU1), with the NS4A conformation selected from the molecular dynamic simulation of the NS3/4A protease in crowded environment. Middle and right figures are prepared based on the figure published by Brass et al.⁴⁰

81 and Ser139, located on the brink of two barrels⁴¹. A characteristic element of the structure is a zinc ion, coordinated by histidine and three cysteine residues (Figure 1.7). Removal of the cation leads to inactivation of the protein, associated with nearly complete loss of its tertiary structure⁴². Another structure defining effector of NS3 is its protein cofactor – NS4A. Without it, N-terminal fragment (aa 1 to 30) of the protein remains unstructured and its proteolytic activity is severely limited⁴³ – the cofactor enables cleavage at NS4A/NS4B and NS4B/NS5A site, (it's able to free NS4A only when bound to NS4A), whereas unbound NS3 only cleaves NS5A/NS5B junction (Figure 1.5). NS3/4A complex is formed by hydrogen bonding between the eighth beta-strand of the NS3's beta barrel and the central part of NS4A (aa 21 to 34). Upon NS4A binding, the N-terminal of NS3 folds into ninth beta strand and a short helix (Figure 1.7), and the active site of the protease is stabilised, enhancing its activity towards all substrates⁴³.

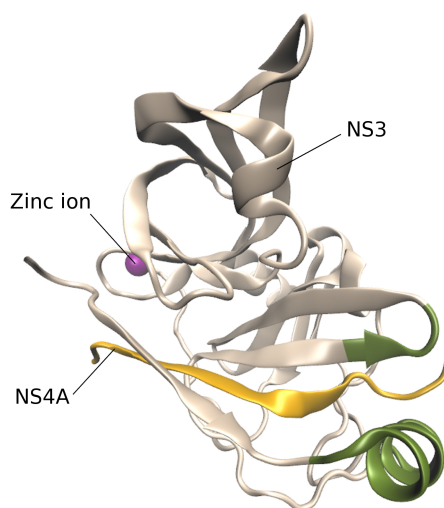


Figure 1.7: The ribbon model of the NS3/4A protease structure (PDB ID 4JMY). The central part of the NS4A protein cofactor is shown in yellow and the NS3 membrane contact sites are in green. The β -sheets include a β -barrel shape in a horizontal view and a long β -sheet interacting with the NS4A cofactor. Figure prepared based on the published work⁴⁴.

NS4A is a short 54 aa protein. While its central part becomes an integral component of the NS3 structure, the N- and C-terminal tails remain unstructured in water. Typically for IDPs, the structure of the tails strongly depends on the surrounding environment. The N-terminal tail contains amino acid pattern typical for transmembrane helices – 15 mostly aliphatic residues flanked with membrane interface residues – Thr2, Trp3 and Thr19, Thr20 on each side. As confirmed by CD and NMR experiments conducted in hydrophobic conditions, this region is able to fold into an α -helix – it is predicted to form transmembrane helix, which anchors the NS3/4A complex to the ER membrane (Figure 1.6). The tail is located in the vicinity of other membrane contacting elements of NS3 – the 38-40 loop and 10-24 helix folded upon NS4A binding. Additionally, the TM helix region also contains several polar amino acids, that might be engaged in protein-protein interaction with TM helices of other HCV proteins⁴⁰. The C-terminal tail of NS4A following predicted helix

kink at aa 33 to 39 contains an acidic segment, that might in some conditions fold into a helix as well – with protonated acidic groups, which can occur in high or inside another protein. This region likely connects NS3/4A with other members of the replication complex. The function of the C-terminus is unclear, however through mutational analysis it has been associated with virus replication and assembly processes, as well as hyperphosphorylation of NS5A protein⁴⁵.

As a partially unstructured protein with tails that change their structure according to conditions and the surrounding environment, NS3/4A is a perfect candidate for a case-study investigating the effects of crowding on the structural dynamics of IDPr. As time scales of potential movements and folding events match time scales available in molecular dynamics simulations, this method was chosen for the study as a way to observe the structural changes themselves at the atomic level and to analyse in detail the mechanisms of potential crowding effects. Experimental methods were used to confirm the simulation results and to investigate the effects of crowding on protein activity. An important part of the study was an interdisciplinary approach to reconstruct a test-tube environment in a molecular dynamics simulation, with atomistic accuracy. This way simulation results can be confirmed with lab experiments, and vice versa, laboratory results can be explained through simulation of a matching system.

1.5 Aims and Objectives

Biochemical experiments and molecular dynamics simulations studying proteins are normally conducted in dilute solutions, such as water or buffers. At the same time, the interior of living cells, that is the natural environment in which the proteins function, is largely crowded with macromolecules and metabolites, which occupy about 30% of the space in the cytoplasm¹. Scientific studies in recent years have shown that the presence of a crowded environment can affect the activity, reaction rates, diffusive properties and structure of proteins^{46,47}. However, there still remain open questions in the field, concerning the mechanism of crowding-related effects, optimal modelling of the crowded environment in experiments and estimation of how it will affect individual macromolecules.

In a broadest context, this thesis is part of an attempt to answer the questions of how to estimate in the case of which proteins standard experimental conditions are sufficient, and in the case of which it is necessary to conduct experiments in an environment that mimicks cellular level of crowding.

The specific aim of this study is to investigate the effect of a crowded environment on the activity and dynamics of a particular protein: the NS3/4A protease from the hepatitis C virus (HCV). It is an example of a protein that possesses intrinsically disordered regions (IDPr) in its structure - i.e. fragments of protein chain that lack fixed three-dimensional structure, which in many cases is recovered when interacting with the binding partner⁴⁸. The hypothesis behind the choice of this model for the study is that proteins carrying unstructured fragments might be a group particularly susceptible to the crowding effect. This is quite a large group, comprising a third of eukaryotic proteins⁴⁹, and an even larger proportion of viral proteins - about 20-40% of viral protein residues are part of IDPr, with the most characteristic IDPr-containing proteins found in RNA viruses⁵⁰. Among them is the studied NS3/4A protease from HCV.

In an interdisciplinary study, I combined analysis of laboratory experiments and simulations to analyse the effects of crowding on the NS3/4A protease in aspects such as:

- rate of the enzymatic reaction catalysed by the protease, differently affected by the presence of two types of crowders
- internal dynamics of the NS3/4A protein complex, with particular emphasis on the biologically relevant unstructured fragment of NS4A
- diffusion of the protein and its peptide substrates, as well as their interactions with crowder molecules

The effects of crowding were studied using both experimental and computational methods, in an attempt to construct one experimental setup in two scales: the test-tube scale, where the results of collective action of billions of proteins are observed, and the simulation, where a single protein is viewed with atomistic detail. In order to best replicate the experimental conditions, crowded environment models normally used in laboratory experiments were incorporated into molecular dynamics (MD) simulations: the protease was surrounded by crowders in the form of non-protein polymers,

which provide a restriction of the volume available for the protein, minimising the protein-crowder interactions.

Polymers in the all-atom representation although minimally but nevertheless show some interactions with the protein residues. In order to divide the effects observed in the simulations into those arising from protein-crowder interactions and those caused solely by the limitation of the space available for the protein, an additional crowder model was designed to exclusively provide volume exclusion. In this model, each polymer molecule was represented by a bead-shell, constructed of carbon-sized pseudo atoms distributed on the surface of the sphere, with a total size corresponding to the average size of a polymer crowder.

By transferring the experimental conditions to simulations, and using various models of crowded environment, it was possible to investigate the molecular mechanisms behind the observed effects, which I focused on to the greatest extent in this work.

The main objectives of my thesis were to investigate how various crowders influence the diffusion, activity and internal dynamics of the NS3/4A protease and its substrates. To achieve them, I carried out research according to the following plan:

- design and test the coarse-grained model of a crowder that would reduce the volume available for NS3/4A, without showing interactions with the protein
- perform molecular dynamics simulations of the protease in water and in the presence of crowders, as well as simulations including the peptide substrates (of the protease surrounded with substrates with and without the crowders)
- determine how the crowders affect the secondary structure of the NS4A cofactor via MD simulations and CD experiments
- analyse obtained trajectories of the protease, substrates and crowders in terms of diffusive properties and inter-molecular interactions
- explain the differences in the activity of the protease in the presence of various crowders shown in the experiments

Chapter 2

Methods

2.1 Computational theoretical methods

This study includes a variety of molecular models, from molecules represented in atomistic detail to reduced models. In this chapter, I describe the techniques used in performing and analysing molecular dynamics (MD) simulations, and in terms of the models used, I focus on the models that were included in every one of conducted simulations - the all-atom model of the NS3/4A protease and explicit water. Specific information about crowder models that I designed, as part of obtained results, are presented in a following chapter.

All-atom molecular model

After one of my conference talks I was asked why should we trust the results obtained with simulations that are using coarse-grained models - models in which a group of atoms is represented by a single pseudo atom. By raising such question we have no choice but to set the all-atom model as an example of the well-known and trusted one. But looking closer into the model where each atom is represented individually, it turns out this is largely a coarse-grained model itself.

In the all-atom model, a molecule is constructed of classical spheres on springs. Each atom is represented as a point mass with assigned charge and radius. Atoms are connected by springs modelled with harmonic potential, representing bonds between them.

In real-life atom, electric charge is spatially distributed between heavier positively charged nucleus and negatively charged electron cloud, where electron locations are only given by probability of finding an electron in a given location. The state of an electron is described with a wave function $\Psi(t)$, whose square of the absolute value at a given point defines a probability density function - a function that describes probability of finding the system in a state described by $\Psi(x, t)$ (Eq 2.2).

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle \quad (2.1)$$

$$P(x, t) = |\Psi(x, t)|^2 \quad (2.2)$$

Evolution of such system over time would be described by Schrödinger equation (Eq 2.1) - a quantum counterpart of Newton Law of dynamics, used to characterize the state of the spheres-on-springs molecule. In both equations, change in the spatial coordinates of the system is strictly dependent on its energy, here described by the Hamiltonian operator $\hat{H}$.

In the case of proton and electron, for the building parts of each of our balls on springs, the word heavier has a very specific meaning; the proton is over 1800 times more massive than the electron. Following this observation, it could be assumed, that nuclei and electronic movements can be treated separately and we can describe electron density in relation to immobilized nuclei - as results from the Born-Oppenheimer approximation^{51,52} - an approximation crucial for our all-atom model. The model focusses on the movements of atomic nuclei, providing information about the electron cloud in the model parameters, i.e. in the partial charges of the particular atom types.

One of the ways to describe the electronic structure of a multi-atom molecule is the molecular orbital theory, in which electrons are assumed to move through the whole molecule, influenced by atomic nuclei, and they are not assigned to particular chemical bond. The states of electrons are then described as linear combinations of atomic orbitals (LCAO)⁵³:

$$\Psi_j = \sum_{i=1}^n c_{ij} \chi_i \quad (2.3)$$

where Ψ_j is an j 'th molecular orbital being a sum of n χ_i atomic orbitals with corresponding weights c_{ij} . χ_i are one electron functions, called basis functions. In quantum-chemical calculations, these functions are often chosen to be linear combinations of Gaussian functions, such as for example:

$$\begin{aligned} g_s(\alpha, r) &= (2\alpha/\pi)^{3/4} \exp(-\alpha r^2) \\ g_{px} &= (128\alpha^5/\pi^3)^{1/4} x \exp(-\alpha r^2) \end{aligned} \quad (2.4)$$

where s and px indexes refer to electronic orbitals. Such functions satisfy the Schrödinger equation and permit integration of the Hamiltonian operator describing the energy of a molecule⁵⁴.

Using quantum chemistry methods, molecular orbitals structures can be obtained for molecules such as amino acids building the structure of a protein, including the studied protease. As shown in Figure 2.1, such structure is far more complex than the balls-and-springs model; moreover, the shape of these orbitals would change upon any fluctuation of the bond length or inter-atomic angles.

For the purposes of the balls and springs models, the structure of atoms is simplified, to be suitable for classical mechanics treatment: the masses are centered on the nuclei, and more importantly, charges are centered as well. In the CHARMM force field used here, charges are calculated based on molecular orbitals - Mulliken population analysis is used for fitting initial summaric partial charge on each atom⁵⁶. Thus, although the model uses significant simplifications in the representation of atoms, the effective mechanics of the molecular system strongly relies on quantum mechanical analyses.

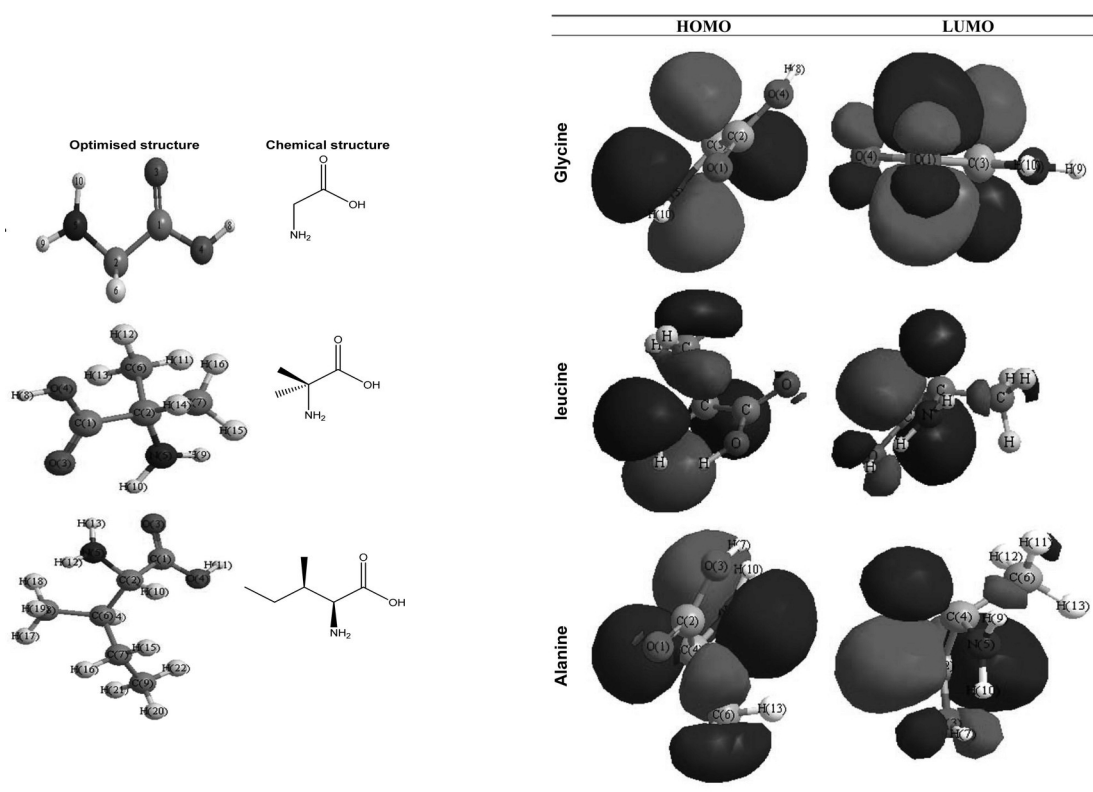


Figure 2.1: The structures of glycine, alanine and leucine, represented as balls-and-springs and showing the highest occupied molecular orbitals (HOMO) and lowest unoccupied molecular orbitals (LUMO). Figure is from the work of N. Eddy et al.⁵⁵.

Molecular Dynamics Simulations

By simulating the movement of biological macromolecules over time, it is possible to get a close-up view of the life of a molecule under selected conditions and surroundings. In full-atom simulations, the motion of each atom is calculated in successive time steps based on the energy landscape of its surroundings, thus depending on the spatial distribution and properties of the atoms in the molecular system.

Dynamics of molecular systems can be described as spatial coordinates and velocities of each atom changing over time - such set of data will be later called a trajectory. Molecular dynamics simulations compute this trajectory using equations of motion, together with empirical equations for the potential energy of each atom. The relation between the motion and force on example particle A is given by Newton's equation of motion:

$$\vec{F}_A = m_A \vec{a}_A = m_A \frac{d^2 \vec{r}_A}{dt^2} \quad (2.5)$$

where F_A is a force on particle A, whose mass is m_A and position vector is r_A . This equation is a second order differential equation, that can be written as:

$$\begin{aligned} \vec{F}_A &= m_A \frac{d\vec{v}_A}{dt} \\ \vec{v}_A &= \frac{d\vec{r}_A}{dt} \end{aligned} \quad (2.6)$$

where r_A is the particle's position and v_A is its velocity. In NAMD software, used in this work to run MD simulations, those equations are integrated using a velocity Verlet algorithm⁵⁷; for x, y and z directions it takes a form:

$$\begin{aligned} \vec{r}(t + \Delta t) &= \vec{r}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 \\ \vec{v}(t + \Delta t) &= \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2}\Delta t \end{aligned} \quad (2.7)$$

It is implemented in three steps:

$$\begin{aligned} \vec{v}(t + \frac{1}{2}\Delta t) &= \vec{v}(t) + \frac{1}{2}\vec{a}(t)\Delta t \\ \vec{r}(t + \Delta t) &= \vec{r}(t) + \vec{v}(t + \frac{1}{2}\Delta t)\Delta t \\ \vec{v}(t + \Delta t) &= \vec{v}(t + \frac{1}{2}\Delta t) + \frac{1}{2}\vec{a}(t + \Delta t)\Delta t \end{aligned} \quad (2.8)$$

where $\vec{a}(t + \frac{1}{2}\Delta t)$ is derived from force calculated for $\vec{r}(t + \Delta t)$. Initial conditions for the algorithm are supplied as starting coordinates, e. g., experimentally resolved protein structure, and starting velocities are drawn from the Maxwell-Boltzmann distribution in a given temperature.

The force on particle A ($\vec{F}_A$) can be found by differentiating the equation for the potential energy,

with respect to its coordinates:

$$\vec{F}_A = -gradU_A \quad (2.9)$$

In molecular mechanics the potential energy of the molecular system is described by a force field - an equation describing the energy of each atom, as a function of current coordinates and parameters of other atoms present in the system. In this work, the CHARMM force field is used, with the potential energy function in the following form:

$$\begin{aligned} U_{TOTAL} = & \sum_{bonds} K_b(\underline{b} - b_0)^2 + \sum_{angles} K_\theta(\underline{\theta} - \theta_0)^2 + \\ & \sum_{dihedrals} K_\varphi(1 + \cos(n\underline{\varphi} - \delta)) + \sum_{improper} K_\phi(\underline{\phi} - \phi_0)^2 + \\ & \sum_{Urey-Bradley} K_{UB}(\underline{r}_{1,3} - r_{1,3;0})^2 + \\ & \sum_{nonbonded} \frac{q_i q_j}{4\pi D_{r_{ij}}} + \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \end{aligned} \quad (2.10)$$

In this formula, components marked in colour are the parameters specific to particular atom types - not just elements but elements that occur in different contexts; the carbon in a peptide bond will be assigned a different type than the carbon in the alanine side chain. Those atom types and their respective parameters are also an essential component of the force field, in CHARMM determined experimentally, with a large contribution from quantum mechanical methods⁵⁶. The underlined components are variables calculated from the current position of the atoms present in the system. In the energy function calculated for each atom, first five terms account for energies resulting from intra-molecular dynamics, such as for example bond stretching described in the first term. In this term, energies from stretching all bonds, in which the particular atom part-takes are summed up, with energies calculated as harmonic potential with a force constant of K_b , specific for this bond, and $\underline{b} - b_0$ being the distance from equilibrium length b_0 , that the atom has moved to. The last term of the energy function, on the other hand, sums up interactions that do not result from bonded connections within a molecule, but from the van Der Waals (VDW) and electrostatic energies of interactions with surrounding atoms. Individual terms of the equation are further discussed in the following section.

Performing a molecular dynamics simulation consists of a production step preceded by system equilibration, and requires obtaining an energetically balanced molecular system. Equilibration is carried out in several key steps:

- Minimization, which corresponds to simulation at 0K, and serves to obtain the optimal conformation of the studied molecules. In this work, the standard minimisation of the NAMD program was used, i.e. according to the conjugate gradient and line search algorithm. In the consecutive (out of about 2000) steps of this stage, the algorithm analyses the energy landscape of the system and finds such a direction of changes in the positions of atoms that would reduce the energy of the system as much as possible. In this way the most energetically favourable

starting structure of the system is obtained.

- Thermalization, which corresponds to gradually raising the temperature and slowly setting the atoms in motion. Here, the division of this stage into several steps has been used - starting with the gradual heating of the solute with restriction of the solute movement, then gradually reducing the restriction and setting the solute in motion as well.
- Introducing pressure control in the system, i.e. changing the thermodynamic system from NVT to NPT, in which constant temperature and pressure are maintained in the system.

A further simulation called a production simulation is run at constant pressure and temperature, usually for a time on the order of nanoseconds or microseconds. Simulations are usually run with periodic boundary conditions, which means that although the simulation box has given dimensions, when an atom encounters a box wall it passes through it to appear on the opposite side of the box. In this way the simulation box mimics the continuity of the simulated solution, as the system behaves as if the box units could be infinitely tiled in all directions. In the simulations described in this work, the simulation box is filled with water in the TIP3P model⁵⁸ - this is an explicit solvent model in which every water molecule is present in the full-atom model. However, by applying the SHAKE algorithm⁵⁹, the hydrogen-oxygen distances are constrained to the nominal length and angle, making the water molecules completely rigid.

In the simulation energies of the atoms, and thus the directions of movements and velocities are calculated based on a given force field, with nonbonded interactions calculated up to a certain cutoff distance. However for greater accuracy calculation of the electrostatic energy of a system is enhanced by adding the component describing long-range interactions. In NAMD software, this is achieved by introducing multiple time step, and dividing the total force acting on each atom into the fast varying local component, including all the terms defined in the force field, and a slower varying long term component, approximating the long range electrostatic interactions according to the Particle mesh Ewald (PME) method⁶⁰. These interactions are not varying as fast as local ones, and so they are not evaluated every timestep.

2.1.1 The CHARMM Force Field

The motion of individual atoms in the MD simulation is dictated by their potential energy, calculated from the position of other atoms in its surroundings. Potential energy of the atoms is described by a force field, which consists of a detailed potential energy equation and a set of parameters describing the properties of individual atom types and molecular geometry.

The equation for the energy of an atom in the CHARMM force field used in this work is the sum of the energies resulting from interactions with neighbouring atoms bound within the same molecule and from electrostatic and Van der Waals interactions (see eq. 2.10). Bonded terms in the force field equation (Eq 2.11) describe fluctuations of bond lengths and angles as well as rotations around bonds within the molecule:

$$U_{bonded} = \sum_{bonds} K_b(\underline{b} - b_0)^2 + \sum_{angles} K_\theta(\underline{\theta} - \theta_0)^2 + \sum_{dihedrals} K_\varphi(1 + \cos(n\underline{\varphi} - \delta)) + \sum_{improper} K_\phi(\underline{\phi} - \phi_0)^2 + \sum_{Urey-Bradley} K_{UB}(\underline{r}_{1,3} - r_{1,3;0})^2 \quad (2.11)$$

Those include the most basic terms: **bonds** and **angles** modelled by harmonic potential, with force constants (K_b , K_θ) and equilibrium values (b_0 , θ_0) determined for each atom combination occurring in molecules covered by the force field; **dihedral** angles defined for each group of four atoms bound to each other, describing torsional rotation of atoms about a central bond, modelled as a cosine function taking positive values, with force constant K_φ , n factor multiplying the cosine function by a natural number from 1 to 4 and δ being a phase constant, typically equal to 0 or 180 (Figure 2.2).

Improper and Urey-Bradley terms don't have to be defined for every group of atoms. Both terms take a form of harmonic potential, with force constants K_ϕ and K_{UB} and equilibrium values ϕ_0 and $r_{1,3;0}$. Improper term describes out of plane deflection, enforcing planar conformations of selected atom groups, and UB term refers to harmonic potential between atoms located two bonds away - these are additional constraints imposed on certain angles between atoms.

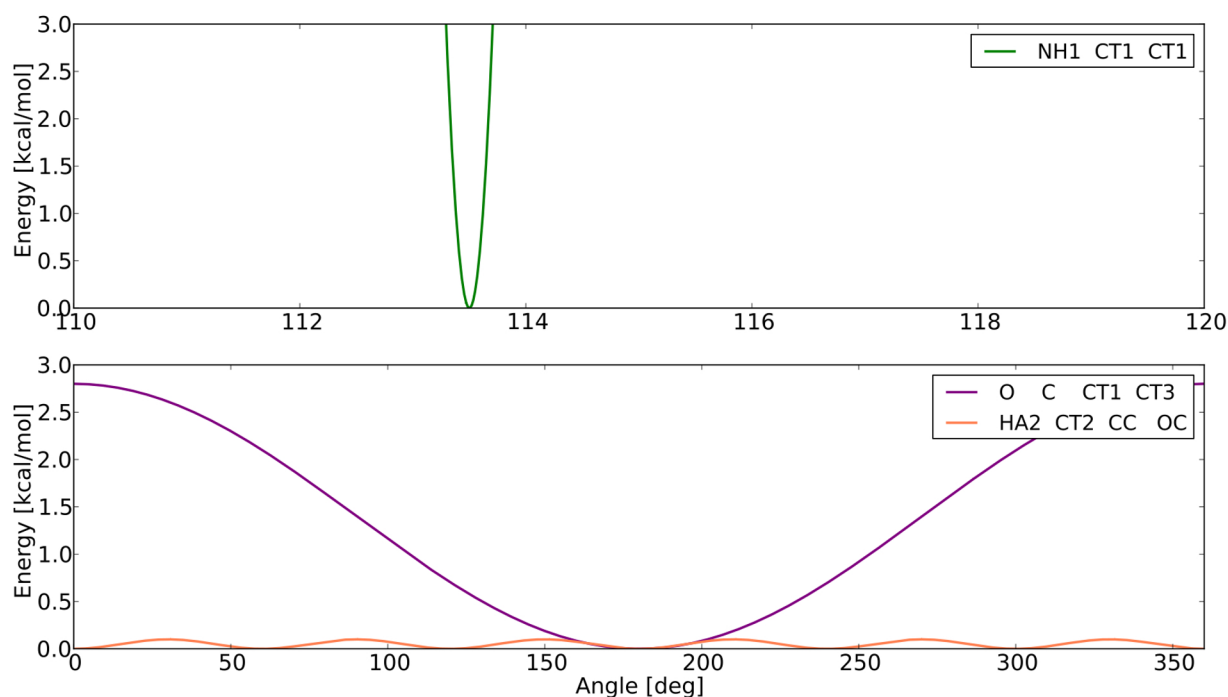


Figure 2.2: Example graphs illustrating the force field terms corresponding to bending of angles and dihedral angles. Labels correspond to CHARMM36m entries, based on which the graphs have been generated.

Nonbonded terms (Eq 2.12) describe interactions between each pair of atoms - both belonging to the same molecule as well as to different ones - these terms would describe the interactions between

the proteins and water molecules, as well as interactions between for example the side chains of neighbouring amino acids. These terms are calculated for each pair of atoms within cutoff distance, except for excluded pairs, which are pairs of very closely spaced atoms bonded together.

$$U_{nonbonded} = \sum_{Electrostatic} \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} + \sum_{Lennard-Jones} \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \quad (2.12)$$

The first term, in the form of Coulomb equation, describes electrostatic interactions between pairs of atoms. Interacting charges, q_i and q_j , except in the special case of ions and ionized particles, are not integers. They are partial charges determined with quantum mechanical methods, resulting from asymmetric distribution of electron cloud along chemical bonds present within a molecule. Due to the potential values, electrostatic interactions are critical for the attractive or repulsive interactions between molecules - the distribution of charges around a given atom will largely determine the direction of changes in its position (Figure 2.3).

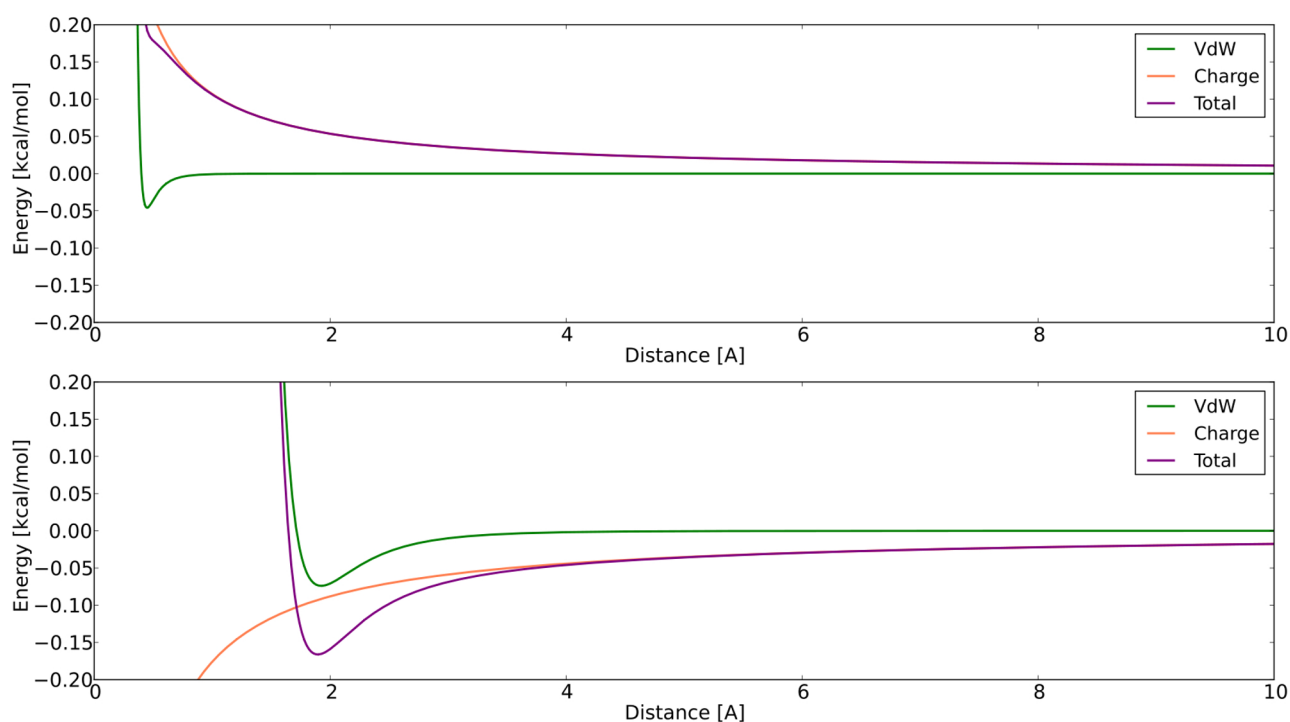


Figure 2.3: Potential energy of interaction between two pairs of nonbonded backbone atoms: oxygen and hydrogen, with partial charges of -0.51e and 0.31e (top), and two hydrogens with partial charges of 0.31e (bottom).

The second term describes Van der Waals interactions modelled with Lennard-Jones potential; it says that a pair of atoms i, j would repel each other at very close distances and when the distance is larger than a certain value they would attract each other. At large distances the LJ potential approaches zero - this is why on the upper graph at Figure 2.3 the total energy line is almost indistinguishable from the one representing electrostatic potential. In the equation for LJ potential,

$R_{min,ij}$ is the distance between two atoms, at which their potential energy equals to zero - this distance is smaller than the one at which energy reaches the minimum value of $-\epsilon_{ij}$. Thus while ϵ_{ij} characterizes the maximum strength of attraction between the atoms, $R_{min,ij}$ value does not try to approximate the atomic radii, but rather describe the shape of the LJ function. Both ϵ_{ij} and $R_{min,ij}$ are pair specific - they are calculated based on parameters specified for two interacting atoms.

In the simulations described here, scaled 1-4 exclusion was used. This means, that interactions between 1-3 atoms (atoms located up to two bonds away) are excluded from nonbonded interaction calculations, and for atoms separated by three bonds the LJ potential is scaled by a factor given in a parameter file, while electrostatic interactions are counted with a scaling factor of 1.

2.1.2 Theory of trajectory analysis

The trajectories obtained from MD simulations are large data sets that often require more than a visual inspection to analyse, especially for complex molecular systems. In order to quantitatively assess the dynamics of the studied molecules, parameters characterising this dynamics are calculated on the basis of the spatial coordinates of molecules in consecutive frames of obtained trajectories.

The most basic and commonly used parameters quantifying trajectory observations include:

- root-mean-square deviation (RMSD) of atomic positions characterising the extent to which conformations of the same molecule differ from one another, or in other words, how much a given structure differs from the reference structure. This measure is calculated based on the coordinates of two structurally aligned conformations as:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (2.13)$$

where δ_i is the distance between atom i in two aligned structures and N is the number of compared atoms.

In this work RMSD calculations are used to select representative conformations of a peptide by grouping conformations obtained from MD trajectories by structural similarity. In the first step of this procedure, an $N_f \times N_f$ RMSD matrix is generated, where N_f is the number of conformations in the trajectory, containing the RMSD values for comparing each conformation with each other. The resulting matrix is then subjected to hierarchical clustering, i.e. clustering of conformations into groups of similar ones. Within this step, an algorithm starts by classifying each conformation as distinct cluster, and then iteratively identify the two clusters that are closest to each other, and merges the two most similar clusters until desired number of clusters is reached.

- Root Mean Square Fluctuation (RMSF) of a molecular structure is calculated as a time averaged RMSD, in which the reference structure is the average conformation of the simulation.

RMSF is calculated for individual atoms, typically the $C\alpha$ atoms of each protein residue as:

$$RMSF_i = \sqrt{\langle (\vec{r}_i(t) - \langle \vec{r}_i \rangle)^2 \rangle} \quad (2.14)$$

where $\vec{r}_i$ are the spatial coordinates of atom i and $\langle \vec{r}_i \rangle$ are the ensemble average coordinates of this atom. While RMSD quantifies how much a structure of a molecule diverges from the reference or starting structure, RMSF characterises the mobility of consecutive amino acids - thus helping to determine which parts of the molecule are stable and which are prone to structural fluctuation.

- Radius of Gyration (R_g) is an example of a measure used to describe the spatial dimensions of a studied molecule, defined as:

$$R_g = \sqrt{\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \rangle} \quad (2.15)$$

where $\vec{r}_i$ is a position of i 'th atom of the molecule and $\langle \vec{r}_i \rangle$ is the mean position of atoms, calculated as an average over coordinates of all atoms of the molecule. This measure is used to describe compactness of a molecule at a given time. Radius of gyration is a very rough estimate of the dimensions of the molecule, but is a measure suitable for analysing changes in these dimensions over time.

This work uses these measures and a range of other analyses to transform trajectory data into easily comparable and interpretable values. In each case, the trajectory analysis scripts take as input data on the position of atoms in successive frames of the trajectory, and from this they calculate parameters characterising processes such as diffusion or conformational changes occurring within molecules. As custom analysis scripts and methods were used in several cases, more detailed information is provided in the following subchapter (Simulation Methods: Trajectory Analysis).

2.2 Simulation Methods

This section describes the technical aspects of preparing the NS3/4A protease model and assembling systems for the simulations. It also provides a detailed description of the conditions of the simulations performed and the methods used to analyse the resulting trajectories.

2.2.1 Building the NS3/4A model

The NS3/4A protease is a complex of a 182 amino acid NS3 protease and 54 aa NS4A protein cofactor. Structures available in the Protein Data Bank (PDB) all comprise of the protease complexed with the central part of NS4A - the part that is one of the building blocks of a β -barrel located in a core of NS3 (Figure 2.4). Unstructured tails on the C- and N-termini of NS4A therefore had to be built based on the amino acid sequence of the protein.

As a model of NS3 protease, the 4JMY model⁶¹ from PDB has been chosen. This is a crystallographic structure obtained with good resolution (1.95 Å). In selecting the structure, the decisive factor was the presence of the Zn^{2+} ion, which is essential for the stability of the NS3 protease⁴².

To match the amino acid sequence of the HCV 1b genotype⁶² used in the experimental studies, a series of point mutations has been carried out. Mutations involving D30E, L36V, G66A, A87K, M94L, S147F, V150A, I170V, A181S, and S182P were introduced using the Chimera software⁶³, and mutated amino acid side chains were modelled automatically using Dunbrack rotamer library⁶⁴.

In the structure obtained from PDB, only residues 21-32 of NS4A were included⁶¹. The N- and C-terminal tails, amino acids 1-20 and 33-54 were added in a way that did not affect the structure of experimentally resolved central part. Template free modelling function included in MODELLER⁶⁵, ver. 9.19, was used to generate five models of full NS4A, including unmodified central part and added tails. Dope algorithm⁶⁶ was used to score the obtained models and select the best one to be used in the NS3/4A simulations.

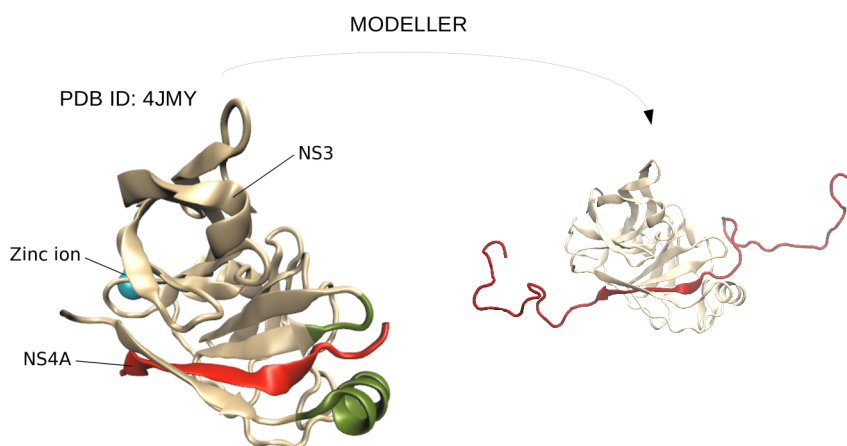


Figure 2.4: Left: NS3/4A structure as deposited in the Protein Data Bank, ID:4JMY. Right: Molecular model of full NS3/4A protease, which was used as the starting structure in all MD simulations. All-atom model was built, however for clarity only simplified backbone structure was shown.

The obtained model of the complete NS3/4A protease was used as a starting structure (Figure 2.4) in the simulations of this protein complex in water (Figure 2.5) and in various types of crowded environments, as well as in the simulation including substrate molecules.

2.2.2 Building the peptide models

Biological function of the NS3 protease is cutting junctions in HCV polyprotein - it means that the protease is able to cleave proteins or peptides in specific positions, recognising their amino acid patterns. Thus a perfect candidate for experimental studies investigating the effects of crowding on enzymatic activity of NS3, would be a short peptide containing one of those specific sequences.

For this particular study, a peptide of a sequence $Ac - DEDEEAASK - NH_2$ was chosen. A peptide of analogical sequence has been used in activity assay, experimentally measuring enzymatic

activity of the NS3/4A protease (See Experimental Methods: Activity Assay). For the purposes of the experiment, FRET tags were added to the terminal residues of the peptide, and one of the peptide bonds was replaced with an ester bond - however, these modifications were not included in the model used for MD simulations. Model of the peptide substrate of NS3 protease was built using Chimera⁶³, along with optimization with Dunbrack Rotamer Library⁶⁴. Conformation used as a starting structure for the NS3/4A - PEG - substrates systems has been obtained based on 20 copies of molecular dynamics simulation, each containing a peptide surrounded with explicit water and ions. Substrate simulations generated 20 production trajectories of 200 ns each, preceded with minimization and equilibration of the systems.

Trajectories were merged and clustered, to obtain structures representing most frequently occurring conformations, later used as starting conformations in the simulations including the NS3/4A protease and its substrates. In the first step, RMSD matrix was generated with ProDy⁶⁷, by comparing all-vs-all RMSD measurements. This matrix was used as an input to `scipy.cluster.hierarchy` module from SciPy, which allows performing hierarchical clustering based on distance matrices. Within `scipy`, complete clustering algorithm was used to divide all substrate conformation into five groups, each representing a distinct type of conformation adopted by the substrate. From those groups, five representative conformations were randomly selected for further study.

Another short protein used in this work was a peptide representing the N-terminal tail of NS4A, of a sequence $N_{ter} - STWVLVGGVLAALAYCLTTGS - C_{ter}$. This peptide has been modelled in two different conformations: helical and extended. It was used for additional simulations aiming to confirm the results of the full protein simulations - longer, 1000 ns trajectories were generated for four systems: extended peptide in water and in water with the coarse-grained crowders, and helical peptide in water and in water with the coarse-grained crowders. A fragment of the full NS4A model was adapted as the extended conformation of the N-terminal tail, whereas the helical conformation was obtained with template-free modelling using Chimera⁶³, this time enforcing helical conformation along the entire length of the peptide.

2.2.3 Preparation of the simulation systems

All simulated systems contained an all-atom model of a protein or peptide located in the center of a simulation box. Another property linking all simulations was the presence of water modelled as explicit solvent, in the full-atom TIP3P model, as well as Na^+ and Cl^- ions, included to neutralise the system and create an ionic strength of around 20 mM. Neutralisation was needed in case of the systems containing full NS3/4A protease, bound to a Zn^{2+} ion and the systems containing peptide substrates, whose total charge, due to the presence of negatively charged residues equals -4e per substrate.

In case of crowded systems (Figure 2.6), the first step of assembly process was choosing the points around the protein, in which the centers of crowder molecules would be located. Locations were chosen randomly, taking into account the crowder radius and desired simulation box size. For

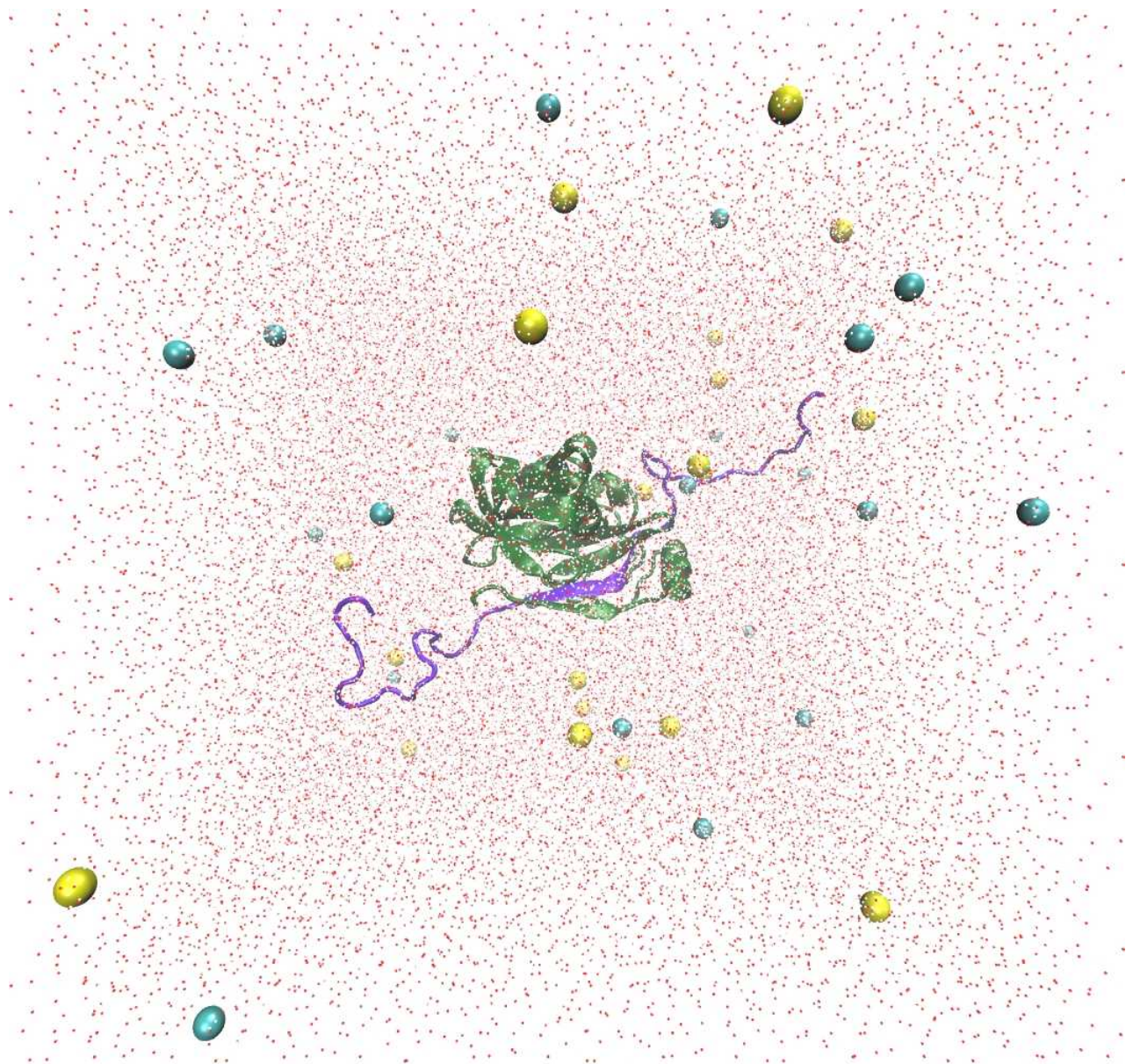


Figure 2.5: Control system containing the NS3/4A protease, water and ions. NS3 protease is shown in green, the NS4A cofactor in violet and ions in yellow and cyan. Explicit water molecules are shown in simplified representation.

each full-protease system, 130 crowder locations were obtained. Then, using VMD⁶⁸ and a tcl script, a chosen crowder molecule has been added in each point location. In case of PEG crowders, before adding each crowder to the system, one of five available conformations was randomly selected.

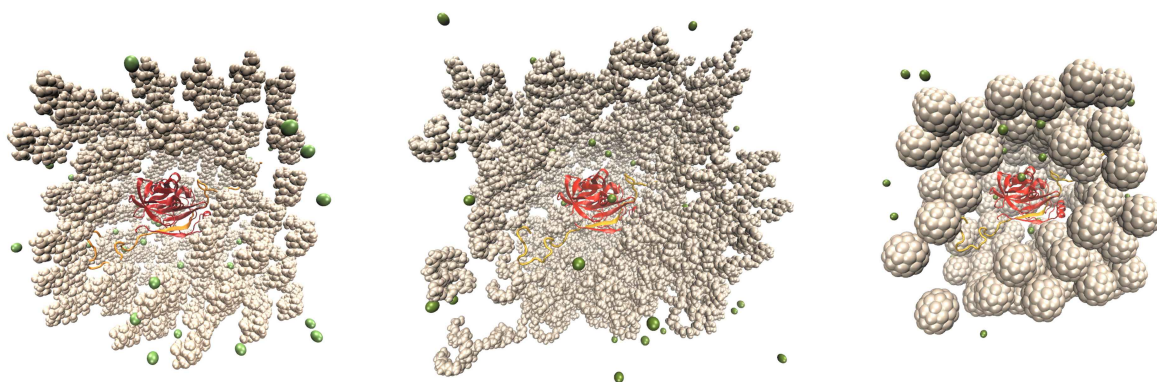


Figure 2.6: Simulated systems containing the NS3/4A protease surrounded by three types of crowders - from the left: polysucrose, PEG and the coarse-grained crowders. The NS3 protease is shown in red, the NS4A cofactor in orange and ions in green. Water molecules are not shown, but they were present in all presented systems.

Building the systems containing peptide substrates (Figure 2.7) comprised of one additional step of adding 10 substrate molecules per system. The substrates were added using VMD, by manually moving the substrates to free spaces between the crowder and protein molecules. Conformations of inserted peptides were chosen based on initial simulations of the peptide in water; the number of peptides of a given conformation was chosen in proportion to the frequency of its occurrence in the initial simulations.

Each crowded system has been checked for clashes - the spots where atoms overlap or bonds cross with each other. The presence of such clashes would result in simulation errors, that is why the systems were carefully checked using VMD⁶⁸ scripts and visually. In case of polysucrose crowders, the clashes were fixed manually by moving individual atoms, while the systems containing PEGs were subjected to 2000 steps of steepest-descent energy minimization in vacuum, before adding water molecules.

Prepared systems were solvated with the MMTSB software⁶⁹, up to a cut-off distance of 9 Å for the uncrowded NS3/4A system and 3 Å for the systems with all-atom crowders. Due to the diverse and not necessarily spherical structures of the crowders, a large cutoff reduction was needed to prevent systems with large margins of water at the edges. MMTSB was also used to add desired amount of Na^+ and Cl^- ions, by replacing randomly chosen water molecules.

Each simulation was run in three copies. A summary of all simulated systems, along with the number of atoms and the number of copies ran can be found in Table 2.1. For each type of the systems with crowders, three copies were generated, differing by the starting locations of crowder molecules. For the systems without crowders, copies of the simulation were differing by the random seed, based on which the starting velocities were generated.

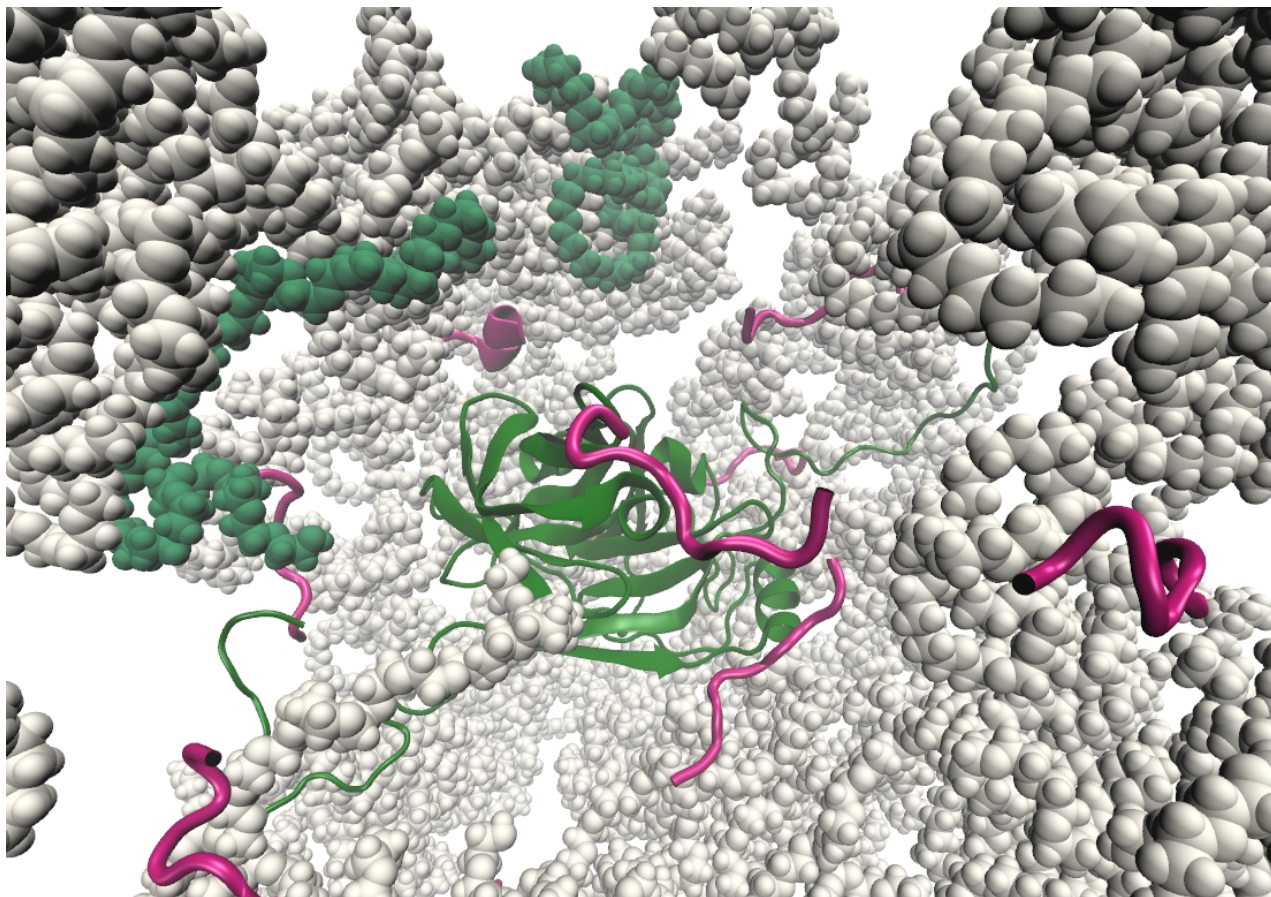


Figure 2.7: NS3/4A protease surrounded with PEG crowders and peptide substrates in starting conformations. The protease is shown in green, PEG crowders in grey and peptide substrates in magenta. Two PEG crowders are shown in sage green, to illustrate various PEG conformations included in this system. For clarity, water molecules and ions are not included in the figure.

System	Mean number of atoms	Number of generated trajectories
Control systems - protease in water	128 726	5
Protease with PEG crowders	250 840	5
Protease with polysucrose crowders	148 154	3
Protease with coarse-grained crowders	120 658	5
N-terminal tail of NS4A in water	20 827	6
N-terminal tail of NS4A with CG crowders	25 045	6
Protease and substrates	119 825	3
Protease with PEG crowders and substrates	241 211	3

Table 2.1: Numbers of atoms in all simulated systems, and number of trajectories generated for each one. There were slight differences in the number of atoms between the copies of the simulations, due to the fact that in each copy of a given system I used different crowder positions, while maintaining their number. Therefore, when solvating systems, there were differences in the amount of water molecules. In the table, for the sake of simplicity, I have given the average numbers of atoms.

2.2.4 Parameterisation

All simulated systems were parameterised using CHARMM force field⁷⁰. For protein and peptide substrate residues, standard protonation at pH 7 were assumed. Histidine residues were represented as neutral τ tautomers, with the hydrogen bond acceptor located at their N ϵ position. A modified patch was used for deprotonated cysteine residues (Cys97, Cys99 and Cys145) coordinating the zinc ion⁴³. The N- and C-terminal parts of NS4A, together making almost 40 amino acids, form highly mobile tails that remain unstructured in water. These fragments are a key component of the simulated systems - their correct parameterisation can affect the results of the study. For this reason, the choice was made for a force field specifically adapted to the simulation of IDP and IDPr⁷¹. While simulating unstructured proteins and peptides, standard force fields have been found to generate increased content of left-handed α -helices relative to experimental measurements⁷². To improve accuracy in modeling IDP conformational ensembles, CHARMM36 force field has been refined and tested against wide range of available experimental data. The resulting CHARMM 36m force field⁷¹ is suitable for both folded and disordered protein chains, therefore it has been used to parameterise both NS3 and NS4A proteins.

For parameterising all-atom PEG and Ficoll molecules used as crowders in presented simulations, elements of CHARMM36 force field for carbohydrates were used^{73,74}, with torsion angles modifications optimised for PEG polymers⁷⁵. Details on design, assembly and parameterisation of all three types of crowders are presented in the next section.

Structural files needed for conducting MD simulation in NAMD, containing information on atomic coordinates and partial charges, bonds and angles in the simulated molecules (pdb and psf files) were generated using the MMTSB toolset⁶⁹.

2.2.5 Water optimization

A problem with many of the available force fields is the overestimation of hydrophobicity of molecules, including the protein chains. This problem does not interfere with simulations of single globular proteins, but when simulating several proteins it leads to excessive association and aggregation of molecules. In the case of unstructured fragments, excessive hydrophobicity leads to promotion of compact structures and can also increase the content of helical fragments. The first work that raised the question of whether modern force fields are suitable for simulating multiple proteins was a 2014 publication by D. Petrow et al.⁷⁶; its authors noted, that while there are examples of simulations of several proteins in explicit solvent, from which conclusions are drawn on their aggregation and interactions, there is no negative control, i.e. simulation of non-aggregating peptides. They performed such simulation, using a protein called villin, and its carbonylated counterpart, that forms aggregates in vitro. Non-carbonated villin is an example of a small (36 aa) protein that doesn't aggregate even at very high concentrations - that's why it later became a model system for studying aggregation in MD simulations. The first simulation of villin ensemble, conducted in the GROMOS force field, showed that the non-aggregating villins under natural conditions aggregate as much or even more

than their modified counterpart. All villins present in the system formed a single aggregate in 22/26 simulations in GROMOS, and 65% of the proteins did not dissociate from the aggregate at all. The simulation was repeated for Amber, CHARMM22 and OPLSA force fields, all with exactly the same results. In each case, formation of the aggregate proceeded with a decrease in potential energy of the system. Another study published in the same year highlights a slightly different aspect of inaccuracies in the force fields - observations showing that proteins in simulations are too poorly solvated relative to experimental data⁷⁷. A factor that can affect both protein association and solvation energy is the hydrophobic effect, and the authors concluded this might be where the problem lies. The only place where we can attempt to model hydrophobic effect in traditional force fields is in the parameters of interactions with water molecules, specifically the epsilon parameter for water atoms, which is an input to calculating the 'strength' of LJ potential of interactions between the water and solute atoms. Upscaling this parameter resulted in agreement of the simulation result with experimental data on hydrodynamic radius of the protein, solvation energy and protein complex formation. In the simulations presented here the VdW interactions between solute and water molecules were up-scaled by the factor of 1.09, based on optimisation conducted for the CHARMM force field¹³. In the optimization process, such a scaling factor allowed the elimination of excessive aggregation of the test peptides, and was low enough not to destabilize their native structure. Technically, the LJ epsilon parameter of the water oxygen atom was increased from -0.1521 to -0.1807 and the epsilon parameter of water hydrogen atoms from -0.046 to -0.05466. This way the calculated water-solute LJ potential is upscaled by the factor of 1.09. Epsilon values for water atoms were calculated taking into account the equation for epsilon parameter for a pair interaction:

$$\epsilon_{i,j} = \sqrt{\epsilon_i \epsilon_j} \quad (2.16)$$

where ϵ_i and ϵ_j are values specific for a pair of interacting atoms. The parameters were upscaled for the interactions with all atom types present in the system, then for the water-water interactions it was reduced back to the original values using NBFIX entries, that allow modifying the nonbonded LJ parameters of selected atom pairs. Such small modification allowed to restore realistic diffusion of all-atom crowders, observed to aggregate when simulated in standard TIP3P water model. The crowder aggregation process itself and the effects of modifying water parameters on the conformational dynamics of the NS3/4A protease are described in the following chapter.

2.2.6 Simulation conditions

One uniform system preparation and simulation protocol has been used in all simulations described in this thesis. The software used to conduct most of MD simulations was NAMD 2.11⁷⁸. Six copies of the simulations with full NS3/4A protease in water and in the presence of PEG and CG crowders were run with OpenMM 7.4⁷⁹. In the first step, the systems were minimized for 3000 steps with the steepest descent method. Then, in the thermalization step, the system was simulated in gradually increasing temperature, with harmonic constraints on protein with the zinc ion and

crowder molecules, with force constant $k = 10 \frac{\text{kcal}}{\text{mol}\text{\AA}^2}$. Each thermalization round lasted 25 ps, with the temperature rising by 10 K per round, from 10 to 310 K. In the next step, the force constant k was gradually decreased, with k values of 10, 5, 2, 1 and $0.1 \frac{\text{kcal}}{\text{mol}\text{\AA}^2}$, 25 ps per round. Both thermalization and equilibration (force constant reduction) were carried out with an integration time step of 1 fs and in NVT ensemble. In an additional equilibration step, the time step was increased to 2 fs, and the ensemble switched to NPT, where the pressure of 1 atm was maintained with a Langevin piston⁸⁰. Production simulations were also conducted in NPT, with temperature of 310 K controlled with a Langevin thermostat. In all of the simulations, nonbonded interactions were calculated up to a cut-off distance of 12 Å. For the systems containing coarse-grained crowders, a 'pairlistdist' parameter was increased from 14 to 17 Å. This parameter specifies the distance, up to which the program writes up atom pairs, for which nonbonded interactions will be considered. Increasing it prevented overlapping between neighboring CG crowders. Electrostatic interactions were calculated using particle mesh Ewald summation⁶⁰. To allow for a 2 fs time step, SHAKE algorithm⁵⁹ was used constraining all bonds with hydrogen atoms.

2.2.7 Trajectory analysis

Most trajectory analysis calculations were carried out for truncated 500 ns trajectories in which frames were saved every 0.5 ns. Mdconv and catdcd were used to assemble full trajectories from partial dcd files and to generate trajectory dcd files for analysis by cutting selected frames or trajectories of selected atoms. For some analyses trajectories were modified by centering them on the NS3 protease or unwrapping the trajectories by removing periodic boundary conditions - those procedures were conducted using mdconv software.

Basic calculations were performed using built-in VMD functionalities:

RMSD (Root Mean Square Deviation of atomic positions) measuring the average distance between atoms in two superimposed conformations of a given molecule, was calculated using RMSD Trajectory Tool. RMSD was calculated for backbone atoms of the NS3 protease, and together with other measures, it was used to assess stability of the protein. In this work, RMSD is given as an average over values comparing conformations from subsequent simulation frames with the starting conformation.

RMSF (Root Mean Square Fluctuation) values, describing the mobility of individual protein residues throughout the simulation, were calculated for the $C\alpha$ atoms using 'measure' command from the VMD Tcl console. RMSF was measured for both NS3 and NS4A proteins, and visualised as plots and protein structures colored according to RMSF values. The structures were generated using a custom python script, that takes a structural pdb file and a RMSF data file as input, and substitutes the values in a pdb's beta column with RMSF values calculated for a given amino acid. The structures obtained in this way were later coloured in VMD according to the values in the beta column.

Radius of gyration (Rg) was calculated using the 'measure' command from the VMD Tcl con-

sole. Rg of the NS3 protein was used to assess the stability of the protease; it is given as an average Rg values calculated for subsequent trajectory frames. Rg of the NS4A cofactor is used to describe conformational changes of this partially unstructured protein. It is visualised as histograms, showing the fraction of conformations occurring in the trajectory that are characterized by a given Rg value.

Secondary Structure Plots were generated using VMD Timeline extension, that assigns secondary structure based on STRIDE⁸¹ algorithm. STRIDE recognizes secondary structure elements in proteins, like α -helices and β -strands based on comparing hydrogen bond energy and mainchain dihedral angles of a given structure with the database-derived and knowledge based recognition patterns. Secondary structure plots generated using VMD were used to visualise conformational changes occurring in the NS4A protein over the course of the simulation. They were also used to calculate helicity parameters, used to compare simulations performed in different environments with respect to the time of stable helix occurrence. The parameter was determined as the fraction of total number of trajectory frames in which the N-terminal tail of NS4A formed either an α -helix of a 3-10 helix. The fraction of trajectory frames was calculated for three copies of each simulation summed up together, considering only the second half of each trajectory, a time between 250 and 500 ns.

To describe the conformation of the C- and N-terminal tails of the NS4A protease a custom measure was introduced. The measure was called the wrapping coefficient, as it allows you to determine the extent to which the tail is wrapped around the protease. The measure assigns each conformation a number from zero to one, where one is a conformation that is maximally extended and zero is the conformation in which the tail maximally adheres to the NS3 surface. The wrapping coefficient W is defined as:

$$W_{tail} = \frac{\left(\sum_{i=1}^N x_i\right) - s_{min}}{s_{max} - s_{min}} \quad (2.17)$$

where x_i is the distance between the $C\alpha$ atom of the i^{th} residue of a given NS4A tail and a fixed point close to the center of mass of the NS3 protease (here the $C\alpha$ atom of GLY141), s_{max} and s_{min} are the minimal and maximal values of the distance sums that were found for the tail in all production simulations. The wrapping coefficient is visualised as plots of W versus simulation time and histograms showing the fractions of conformations with a given W value found in a given trajectory.

RMSD based conformation **clustering** was used to classify and group NS3/4A conformations and to select peptide substrate conformation to be used as starting structure in MD simulations. All-against-all conformation similarity was assessed with ProDy⁶⁷, by generating a RMSD matrix in which each pair of conformations from the trajectory was assigned an rmsd value. The `scipy.cluster.hierarchy` module from Python Scipy library was used to cluster conformations based on the RMSD matrix. Complete clustering algorithm was chosen as most appropriate for molecular trajectory analysis⁸².

Translational **diffusion** coefficients were calculated for the NS3/4A protease, the substrates and all types of crowder molecules, based on the Mean Square Displacement (MSD) of their center of masses. In case of dimeric NS3/4A, to eliminate the effects of high NS4A flexibility, locations of

the centre of mass of the stable globular NS3 were collected. Centers of mass were collected in vmd using the measure center function. MSD was calculated using a custom script as:

$$MSD(\tau) = \left\langle (\vec{r}(t + \tau) - \vec{r}(t))^2 \right\rangle \quad (2.18)$$

where r is the location of the molecule in time t and τ are the lag times between the positions. Diffusion coefficients D_0 were calculated based on the slopes of a linear fit to $MSD(\tau)$, according to the Einstein relation:

$$D_0 = \frac{MSD(\tau)}{6\tau} \quad (2.19)$$

D_0 was corrected to eliminate the periodic boundary condition artefacts⁸³ by implementing the $D_{t,PBC}$ correction:

$$D_{t,PBC} = \frac{Tk_B}{6L\pi\eta} \left(\zeta - \frac{4\pi R_h^2}{3L^2} \right) \quad (2.20)$$

$$\eta = \eta_w(1 + 2.5\phi) \quad (2.21)$$

where $\zeta = 2.837$, k_b is the Boltzmann constant, T is the temperature, here equal to 310 K, L is the length of the periodic box, η is the shear viscosity of the solvent. R_h is the hydrodynamic radius of a molecule calculated with HYDROPRO⁸⁴, with $R_{NS3/4A} = 29 \text{ \AA}$ and the average radius of a PEG crowder R_{PEG} equal to $7.7 \text{ \AA}$. Shear viscosity of the solvent in crowded systems was calculated based on the volume fraction of PEG crowders, ϕ and the value of pure water viscosity, η_w . To account for the effect of crowding on the viscosity of solvent, the diffusion coefficients were further scaled with the factor of $3.08/8.9$. In this scaling term, 3.08 corresponds to the shear viscosity of the TIP3P water model, which is altered relative to the experimental value of 8.9, calculated for bulk water. Final corrected diffusion coefficients are calculated as:

$$D_{corrected} = (D_0 + D_{PBC}) \cdot \frac{3.08}{8.9} \quad (2.22)$$

A method based on rotational correlation functions⁸⁵ was used to calculate rotational diffusion coefficients D_r and rotational correlation times τ_r of the NS3/4A protease. Again, to eliminate the impact of NS4A mobility, NS3 rotations were considered as an approximated diffusion of the protein dimer. The chosen method requires first to generate a trajectory with random vectors, and merge it with the trajectory of a centered protease. The vectors are then rotated along with the rotations of NS3, obtained by fitting the protease to its starting conformation. The rotational correlation times are calculated from the average correlation function generated for the random vectors, fitted to the exponential function:

$$f = \exp\left(\frac{-x}{\tau_r}\right) \quad (2.23)$$

Correlation functions were generated using MMTSB Toolset⁶⁹, and the fitting procedure was carried

out with Gnuplot. Calculated correlation times are related to the rotational diffusion coefficient by

$$D_r = \frac{1}{6\tau_r} \cdot \frac{3.08}{8.9} \quad (2.24)$$

where $\frac{3.08}{8.9}$ is a scaling term included to account for altered viscosity of the TIP3P water model. The PBC correction was not applied in this case, as its effect has been found negligible for rotational diffusion⁸⁶. Rotational diffusion analysis carried out for NS3 was used to assess the extent to which crowder molecules interact with the protease. However, it was not possible to use this measure to describe the behavior of the crowders themselves. The high conformational variability of the PEG and polysucrose molecules did not allow it - a problematic part of the procedure in this case would be fitting the polymer conformations onto the starting structure. Artefacts generated this way would make the calculated value not correspond to the actual diffusion of the crowder molecules.

Obtained translational diffusion coefficients and rotational correlation times were compared with theoretical values calculated for the hard sphere model (HS)⁸, to assess coupling and decoupling of those measures. By analysing whether the effect of crowders of a given type on diffusion parameters matches the theoretical effect of non-interacting HS crowders, it is possible to assess to what extent the molecules forming the crowded environment interact with the protein under study⁸⁷. For example, a crowder that affects the rotational and translational diffusion of a protein in a way that corresponds to how HS would affect it would be considered a molecule that does not interact with the protein. In contrast, when rotational diffusion is restricted to a greater extent than would be restricted by the presence of HS, it may mean that the protein's rotation is blocked by crowders interacting with its surface. Comparison of the crowders with the HS model is done by counting the proportionality factor k for both types of diffusion - these tell us how much the effect of the crowders on a given type of diffusion matches the theoretical effect of the HS. By comparing the obtained coefficients, we can assess whether a given crowder affects translational and rotational diffusion to a comparable or different extent - this is the so-called coupling or decoupling of the calculated diffusion values.

Theoretical diffusion parameters in the HS model were calculated based on effective hard sphere volume fractions ϕ_{HS} defined as

$$\phi_{HS} = k\phi \quad (2.25)$$

where k is the scaling factor and ϕ is the volume fraction of crowders obtained based on the trajectory. The translational diffusion coefficient and rotational correlation time in the HS model are described with following equations^{88,89}:

$$\begin{aligned} \frac{D^L(\phi_{HS})}{D_0^L} &\cong \frac{(1 - \phi_{HS})^3}{1 + (3/2)\phi_{HS} + 2\phi_{HS}^2 + 3\phi_{HS}^3} \\ \frac{\tau_r(\phi_{HS})}{\tau_{r,0}} &\cong [1 - 0.631\phi_{HS} - 0.762\phi_{HS}^2]^{-1} \end{aligned} \quad (2.26)$$

where D_0^L and $\tau_{r,0}$ are the diffusion parameters calculated from the trajectories of the protein in water, $D^L(\phi_{HS})$ and $\tau_r(\phi_{HS})$ are diffusion parameters in crowded environment, derived theoretic-

cally or calculated from MD, dependent on the effective HS volume fraction ϕ_{HS} . The equations have been used in two ways: (a) substituting D_0^L , $\tau_{r,0}$, $D^L(\phi_{HS})$ and $\tau_r(\phi_{HS})$ with the values of the diffusion parameters for proteins in water and in the crowded environment obtained from the trajectories allows us to calculate the k-factors for a given type of crowder; (b) substituting D_0^L and $\tau_{r,0}$ with the values of the diffusion parameters for the protein in water and assuming scaling parameter $k=1$ allows to calculate the diffusion parameters for crowders in the HS model.

Protease-crowder contacts and interactions were characterized based on how the presence of crowders affected diffusion of the protein, but also on detailed analysis of the distribution of crowder molecules around it. One of key methods used for this purpose was determining the Radial Distribution Function (RDF), which describes the density of molecules of a given type as a function of distance from the protein surface. For example, a visible layer of crowder molecules formed around the protein surface would indicate the presence of protein-crowder interactions, while a roughly uniform distribution would suggest their absence. In this work, RDF was calculated using a custom program written by prof. Michael Feig. The program counts the number of specific crowder atoms found within a given distance from the closest heavy atom of the protease. Given that the purpose of the calculations was to compare the distributions of three different crowders that differ significantly in the number of atoms, the distributions were calculated only for atoms chosen in such a way that their numbers were comparable. For the CG crowders, RDF was calculated for 42 surface pseudo atoms, and for PEG and polysucrose molecules only the oxygen atoms or the atoms having 1 in their name were taken into account (30 and 31 atoms respectively). The RDF values were also normalized with respect to the number of atoms, so that their plots could be easily compared. In addition to analysing protein-crowder interactions, RDFs were used to monitor aggregation of crowder molecules.

To visualize the distribution of crowders in the vicinity of each element of the protease structure, data were collected on the number of crowder atoms at a fixed distance from each amino acid of the protein. VMD 'atomselect' and 'num' commands were used to calculate the number of PEG and polysucrose atoms found within 5 Å from a given residue, averaged over trajectory frames. To visualise the data, pdb files were generated with calculated number of crowder atoms written in the beta column. Obtained structures were then colored by the beta column value, to show around which residues crowders are most often found.

Based on correlation analysis, protein-crowder encounters were examined in terms of the average time scale of contact occurrence. The contact function was used to describe, whether a contact that is present at time t_0 is still present at $t_0 + \Delta t$:

$$P(\Delta t) = \frac{1}{N-k} \frac{1}{N_p} \sum_{j=1}^{N-k} \sum_{i=1}^{N_p} \delta_i(t_j) \delta_i(t_j + \Delta t) \quad (2.27)$$

where N is the total number of trajectory frames, Δt is the k -th time interval and N_p is the number of molecule pairs. The summation includes j trajectory frames and i pairs of molecules. The function $\delta_i(t)$ takes the value of 1 when the i -th pair of molecules is in contact at time t , and the

value of 0 when the molecules are not in contact, that is when the distance between them is larger than the cutoff value of 5 Å. To obtain correlation times of contacts, the contact function is fitted to a double exponential function:

$$f(\Delta t) = S_c \cdot \exp\left(\frac{-\Delta t}{\tau_c^s}\right) + (1 - S_c) \cdot \exp\left(\frac{-\Delta t}{\tau_c^l}\right) \quad (2.28)$$

where τ_c^s and τ_c^l are the estimated contact times on shorter and longer time scales and S_c is the weight corresponding to the shorter contact time. For additional analysis of crowder distributions, average minimal distances between the protein surface and crowder atoms were calculated. Minimal distances were identified using the processDCD.pl program from the MMTSB Toolset⁶⁹.

Formation of **hydrogen bonds** between protein and crowders was analysed using VMD Hydrogen Bond extension, allowing to generate a list of hydrogen bonds between two groups of atoms, based on given distance and angle criteria. The maximum donor-acceptor distance was set to 3.5 Å and the donor-hydrogen-acceptor angle criteria were set to no less than 120 degrees, which extends the default VMD angle criteria by 20 degrees.

All **errors** calculated for presented results and included in the graphs are the standard errors of the mean (SEM). The **p-values** used to compare the mean helicity of the N-terminal tail of NS4A between the simulations conducted in water and crowded environment were calculated using a two-sample t-test.

Figures presenting molecular structures and simulated systems were prepared using VMD⁶⁸; the graphs presenting computational results were generated using Python with the PyPlot library.

2.3 Experimental methods

2.3.1 Materials

The NS3/4A protease from HCV genome 1b was purchased from Sigma Aldrich, ID: HCV-NS4A/NS3-1b. Though the protease was only bound to a fragment of NS4A, which allowed it to properly fold its N-terminal part and retain full enzymatic activity.

The peptide substrate used in the activity assays was synthesized in our laboratory, and the peptide substrate of analogical sequence was used in MD simulations. The peptides used in these two experiments had minor differences due to the specificity of each experiment. For the activity assay, EDANS and Dabcyl fluorescent groups were added. To maximize proteolytic cleavage efficacy, the peptide bond at the cleavage site was exchanged for an ester bond and a non-standard aminobutyric acid (Abu) residue was used⁹⁰. Substrate modifications enhancing NS3/4A activity enabled detection of activity with smaller enzyme concentrations. For the purpose of molecular dynamics simulations studying substrate diffusivity in various environments, the Asp-Glu-Asp-Glu-Glu-Ala-Ala-Ser-Lys peptide was used to simplify the system build-up and eliminate potential artefacts arising from the presence of large fluorescent groups.

CD experiments focus on the structural characteristics of the peptide resembling the N-terminal part of the NS4A protein cofactor. For this purpose, the HPLC-grade peptide was purchased from Lipopharm.pl (Gdańsk, Poland). 22-amino acid chain of the sequence identical to this used in MD was additionally flanked by KKGG solubility tags, which were not used in the simulations.

PEG crowders, including PEG 600, 1200 and 6000 were purchased from Alfa Aesar (USA) and Ficoll 400 from Sigma-Aldrich (USA).

For substrate synthesis, analytical grade residues were used, including Fmoc-Asp(EDANS)-OH and Fmoc-Lys(DabcyI)-OH purchased from Nova-biochem, Merck.

5-((2-Aminoethyl)amino)naphthalene-1-sulfonic acid (EDANS) purchased from AnaSpec Inc. and also BSA and Tenta Gel S RAM resin obtained from Sigma Aldrich.

2.3.2 Peptide substrate synthesis

Substrate synthesis has been performed by Natalia Ostrowska, with the methodological help from dr Monika Wojciechowska and dr Ksenia Maximova⁹¹.

The NS3/4A protease substrate used in the FRET activity assay was synthesized by the solid-phase peptide synthesis (SPPS) method^{92,90}. The first step of the SPPS method is immobilizing the C-terminal residue on the solid state resin - here the Tenta Gel S RAM resin (loading 0.24 mmol/g) was used. The first residue, as well as other amino acids used to synthesise the peptide are N-protected with 9-fluorenylmethyloxycarbonyl group (Fmoc). The reactive groups of amino acid side chains are also protected with groups such as tert-butyl (tBu) or tert-butoxy (OtBu). Consecutive steps of the synthesis process include the Fmoc deprotection and coupling step where carbonyl group of the next N-protected amino acid reacts with the last amino acid in the existing chain. After coupling, the excess of the amino acid is rinsed off with DMF and DCM, and the reaction progress of each coupling step is confirmed by the Kaiser test⁹³, which helps to make sure that the coupling step was completed. In this test, free amine groups of the un-protected N-terminus of peptide-resin are detected by color-changing reaction with potassium cyanide, phenol and ninhydrin - this way uncompleted coupling step can be identified.

Protected amino acids used in the synthesis described here were Fmoc-Lys(DabcyI)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Asp(EDANS)-OH and Fmoc-Asp(OtBu)-OH. For the purpose of the FRET based assay, two of the amino acids were connected with donor and quencher fluorescent tags - EDANS and DabcyI.

The synthesis procedure, as well as the substrate sequence was designed based on the protocol described for the NS3/4A protease by Taliani et al.⁹⁰. The synthesis steps are described below.

Deprotection of the Fmoc group was done with 20% piperidine in dimethylformamide (DMF), for 2 cycles (5 and 15 minutes).

Coupling of amino acids bound with protective groups was carried out by dissolving Fmoc-AA-COOH in DMF/ N-Methyl-2-pyrrolidone (NMP) (1:1) solution for 1.5 h with a mixture consisting of excess of coupled residue and peptide coupling reagents:

- 2.5 eq. of Fmoc-AA-COOH,
- 2.5 eq. of PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate),
- 2.5 eq. of HOAt (1-Hydroxy-7-azabenzotriazole), and
- 5 eq. of DIPEA (N,N-diisopropylethylamine).

Esterification of **Fmoc-Abu-OH** to the free hydroxyl group of lactic acid was done by using 12 eq. of Fmoc-Abu-OH, 6 eq. of *N, N'*-Diisopropylcarbodiimide (DIC) and catalytic amount of DMAP (4-dimethylaminopyridine) - the reaction mixture was stirred for 1 h.. Coupling of L-lactic acid was carried out using a different mixture:

- 5 eq. of L-lactic acid,
- 5 eq. of DIC and
- 5 eq. of HOAt.

In case of each residue, the reaction mixture was stirred for 1.5 h.

The steps performed after the desired sequence of amino acids was obtained included:

- **acetylation** of the N-terminus of the obtained peptide with 10 eq. of Ac_2O and 10 eq. of DIPEA, stirred for 2 hours,

- **deprotection** of the tBu and OtBu protecting groups and **cleavage** of the peptide from resin with a mixture of trifluoroacetic acid (TFA), water and triisopropylsilane (92.5, 2.5, 2.5) for 2 h at room temperature,

- **lyophilization** of the crude peptide

- **purification** done by Reversed Phase High-Performance Liquid Chromatography (RP-HPLC).

It is a method used to separate the components of a mixture on the basis of their hydrophilicity: the analysed mixture is injected into the column containing stationary phase such as silica beads, and then washed with liquid phase of gradually changing polarity. Analytical and semi-preparative RP-HPLC of the peptide was performed with the SYKAM system, using Knauer C18 columns (4.6 250 mm, 5 mm particle size 250 mm, 5 mm particle size). The analysis was conducted with the mobile phase gradient from 30% MeCN/H₂O + 0.1% TFA to 50% MeCN/H₂O + 0.1% TFA in 30 min

- **Confirmation** of the presence of desired peptide with mass spectrometry (MS) (Figure 2.8). This kind of analysis allows to separate the components of a studied solution based on their mass and charge - the results of MS are displayed as each component abundance as a function of mass-to-charge ratio. It is achieved by first ionizing the sample with an electron beam, and then accelerating and deflecting obtained ions in an electric and magnetic fields⁹⁴.

- purified peptide was dissolved in TFA, that has been exchanged to HCl by dissolving the peptide in 100 mM HCl. The solution was then freeze-dried and lyophilized to obtain solid product.

2.3.3 Activity assay

The enzymatic activity assay was performed by Monika Wojciechowska and Agnieszka Popielec⁹¹

Detecting the NS3/4A activity was based on Förster resonance energy transfer (FRET) between two fluorophores (light sensitive molecules) attached to the peptide substrate. In this method, light-

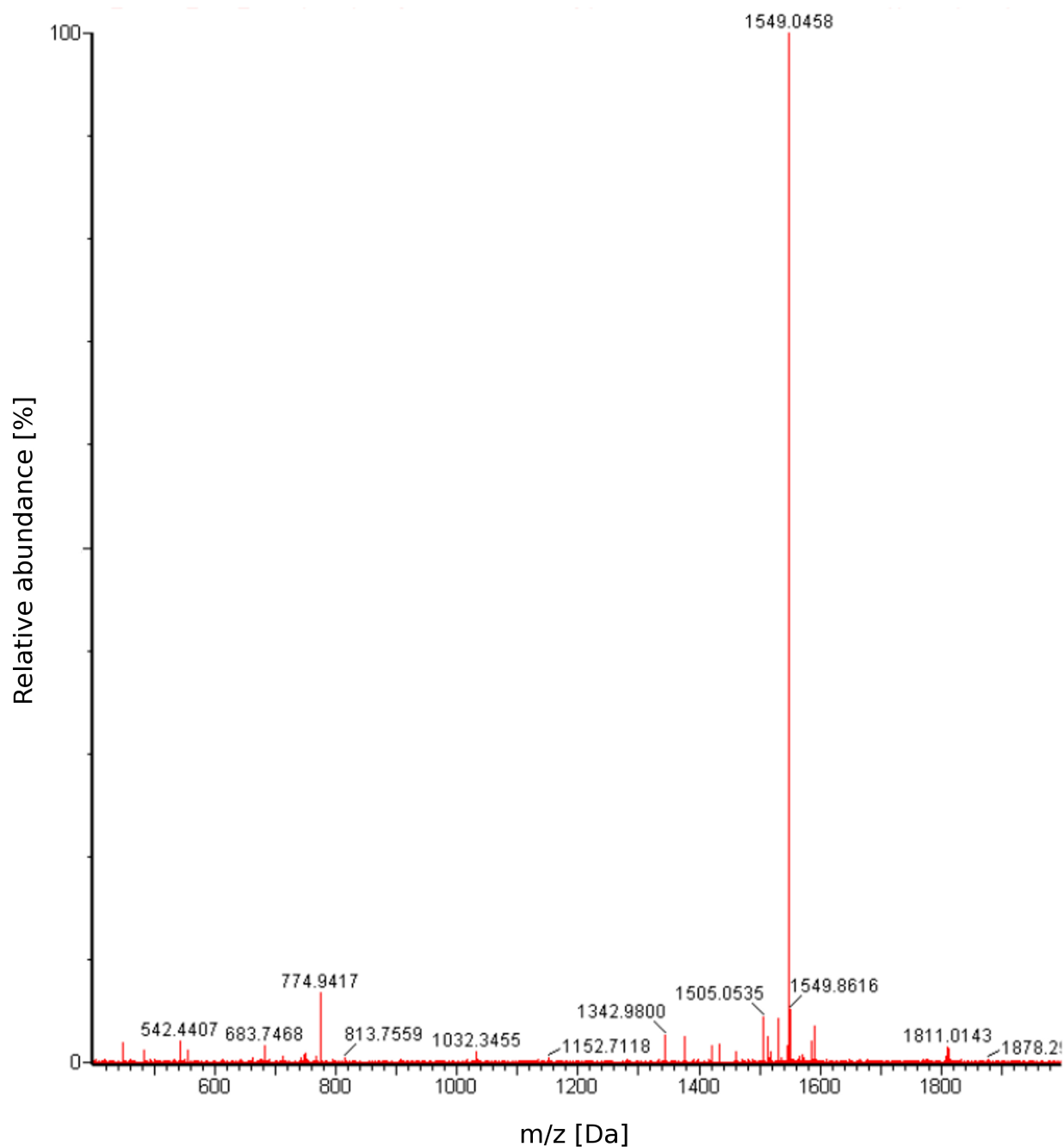


Figure 2.8: Mass spectrum of the NS3/4A substrate tagged with FRET donor and acceptor, with a sequence of $N_{ter} - Ac - Asp - Glu - Asp(EDANS) - Glu - Glu - Abu - \Psi[COO] - Ala - Ser - Lys(Dabcyl) - NH_2 - C_{ter}$. Calculated mass of the substrate equals 1548.58 and the MS determined m/z value was 1549.05 - this result consists of the mass of peptide and the H^+ ion.

waves are used to induce electronic excited state in one of the fluorophores (a FRET donor). The excitation energy can be emitted as fluorescence, or it can be transferred to another chromophore (the FRET acceptor) through nonradiative dipole-dipole coupling. This type of energy transfer decays with the donor-acceptor distance - this way FRET can be used to detect changes in this distance, by measuring changing fluorescence of the donor⁹⁵.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad (2.29)$$

where E is the FRET efficacy, that is the probability of FRET energy transfer per single donor excitation, r is the distance between donor and acceptor and R_0 is the donor-acceptor distance, on which E equals 50%, also called the Förster distance, that can be calculated for each donor-acceptor pair as:

$$R_0^6 = \frac{9000 \ln(10)}{128 \pi^5 N_A} \frac{\kappa^2 Q_D}{n^4} J \quad (2.30)$$

The Förster distance depends on Q_C , the quantum yield of the donor when the acceptor is absent, n , the refractive index of the medium, κ^2 , being the orientation factor of the emission and absorption dipole moments of donor and acceptor molecules and J , the integral describing the overlap between the donor emission spectrum and the acceptor absorption spectrum⁹⁵.

In the substrate used here, EDANS absorbs light of a given wavelength, adopts its electronic excited state and then transfers the energy either to Dabcyl, through FRET, when EDANS and Dabcyl are located close to each other, or it emits energy mostly as fluorescence, when the fluorophores are far enough apart. This way, with FRET donor and acceptor located on two sides of enzymatic cleavage site, FRET can be used to detect cleavage events, as with cutting each substrate its EDANS tag fluorescence is increased⁹⁰.

To prepare the experiment, preliminary tests were performed, such as EDANS calibration and determination of the EDANS inner filter effect in the presence of the FRET substrate. The calibration procedure was aimed to check the linearity of EDANS fluorescence growth with increasing EDANS concentration, in conditions used in the protease activity assay. EDANS fluorescence was measured for its concentrations of 1,000, 500, 250, 125, 62.5, 31.25 and 15.63 nM and in the presence of crowders - PEG 600, 6000 and Ficoll 400. Neither high concentration nor the presence of crowders has been found to disturb linear growth of EDANS fluorescence. Another problem that can affect the spectral measurements in concentrated solutions is the inner filter effect - it happens when the excitation beam is mostly absorbed by the surface part of the sample, and not the centre of the cuvette, or when the light emitted in the central part is then absorbed by the chromophores closer to the cuvette walls. Both effects may result in artefacts in the experimental results. To ensure that the inner filter effect does not occur for EDANS in high FRET substrate concentrations, its fluorescence was measured for EDANS and in different concentrations of the substrate. Then EDANS fluorescence was compared with the difference between the fluorescence of EDANS and substrate+EDANS solutions. Statistical significance of the observed differences was calculated with GraphPad Prism,

using ANOVA test. Tested concentration of EDANS was 0.5 mM and the concentrations of substrate were 2.5, 5, 10, 20 and 40 mM. The experiment did not detect any significant effect of increasing substrate concentration on EDANS fluorescence.

As mentioned before, the activity of the NS3/4A protease was detected based on the fluorescence of EDANS-tagged peptide substrate, whose fluorescence increased upon enzymatic cleavage. The enzyme concentration was 2 nM, the concentrations used for the peptide substrate were 40, 20, 10, 5, and 2.5 mM and the concentrations of crowder molecules (PEG 600, 6000 and Ficoll 400) were either zero or 50, 100 or 200 g/L. The buffer used to make enzyme-substrate solutions containing 50 mM HEPES (pH 7.8), 100 mM NaCl, 20% glycerol, 5 mM dithiotreitol.

The EDANS fluorescence was recorded every 50 s for 2 h, and the measurements were carried out using a microplate reader from BioTek (Winooski, United States) with Corning Thermowell PCR 96 well plates. The excitation wavelength was 340 nm, and the recorded EDANS emission wavelength was 490 nm. The experiments were carried out in 37°C. Each measurement was performed with three repetitions, and the results were reported as average values with errors calculated as standard deviation.

EDANS fluorescence was recorded every 50 seconds for 2 hours. The results were used to calculate kinetic parameters of the NS3/4A catalysed hydrolysis. Initial values of V_{max} were obtained based on initial phase the of reaction, and expressed as fluorescence units per second. Final values of V_{max} and K_M were calculated by fitting the experimental results to the Michaelis-Menten equation:

$$v = \frac{d[P]}{dt} = V_{max} \frac{[S]}{K_M + [S]} \quad (2.31)$$

In this wide-used model, V_{max} represents the maximum catalysis rate achieved by an enzyme, and the Michaelis constant K_M equals the concentration of the enzyme's substrate, at which the enzymatic reaction achieves 50 % of V_{max} . Experimental data was fitted using GraphPad Prism.

2.3.4 CD - measuring peptide helicity

Circular dichroism spectroscopy (CD) was used to experimentally determine secondary structure of the N-terminal tail of NS4A protein. The structure of this fragment was one of the main objects of interest during the analysis of molecular dynamics simulations of the NS3/4A protease. Results of those simulations pointed at increased helical content in the N-terminal tail of NS4A in the presence of all analyzed crowders - the CD experiments were conducted by Natalia Ostrowska to confirm those findings.

Circular Dichroism is the difference between the extent to which the left-handed and right-handed polarized light is absorbed by a chiral molecule. In studying protein structure, CD spectroscopy relies on characteristic absorption spectra exhibited by α -helices and β -strands (Figure 2.9) - structural elements in which residues adopt repeating pattern. In such structures, characteristic configuration of peptide bonds affects the far-UV spectra such a way, that allows to use them to determine the content of particular structure types in the studied sample^{96,97}.

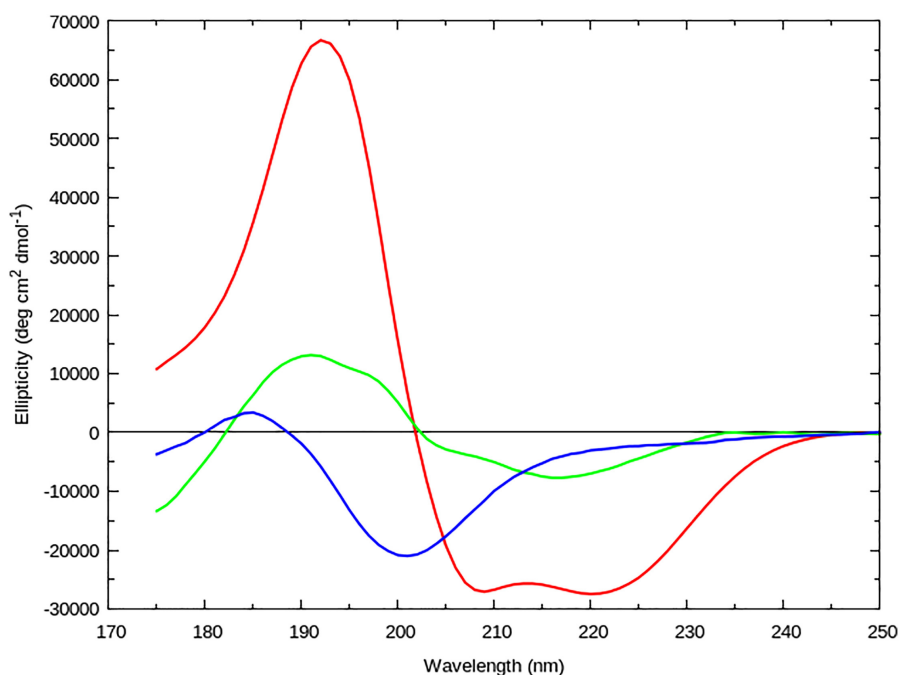


Figure 2.9: CD spectra for predominantly α -helical structure in red, for β -strand in green and for random coil in blue. Figure copied from Rogers et al.⁹⁷

As the N-terminal tail of NS4A contains a significant amount of hydrophobic residues, it is not well soluble in water. To increase its solubility, amino acids with polar properties were added at each end, resulting in a final sequence of *KKGG – STWVLVGGVLAALAAYCLTTGS – GGKK*.

The CD spectra were recorded for the peptide in water, in a phosphate buffer (pH 7.0) and in different types of crowded environment - in the presence of PEG 600, 2000 and Ficoll 400. For the peptide, a concentration of 50 μ M was used. The concentration of PEG crowders of both sizes was 50 mg/ml, however due to problems with dissolving the peptide in Ficoll solution, a Ficoll concentration of 25 mg/ml was used. To record the CD spectra, the Biokine MOS-450/AF-CD spectrometer with a Xe lamp was used. Measurements were conducted in 0.1 cm cell and for the far-UV wavelength range (190-260 nm), with a resolution of 1 nm, and acquisition time of 2 s. The high tension (HT) values were monitored during measurement, and voltage values did not exceed 600 V.

CD spectra presented in this work are the averages over two measurements, with three scans performed during each measurement. To smooth the presented graphs, the Savitzky-Golay method⁹⁸ was used.

Chapter 3

Crowders

3.1 Three models of crowder molecules

In this thesis, three types of crowders are described and used: all-atom PEG and polysucrose molecules, and the coarse-grained crowders, intended to represent the size and hydrophilic properties of 28-mer PEG molecules, but minimise the interactions with the NS3/4A protease (Figure 3.1). All crowders have a similar volume, so that the space they exclude is comparable, but the main difference between them is the degree of interaction with the protein. Thus, the observation of the protease surrounded by three types of crowders serves to disentangle the effects of volume exclusion and those resulting from protein-crowder interactions. In the current chapter, the crowders will be described in the order in which they were created during my doctoral studies. As all three models were built from scratch, I consider them as part of the obtained results.

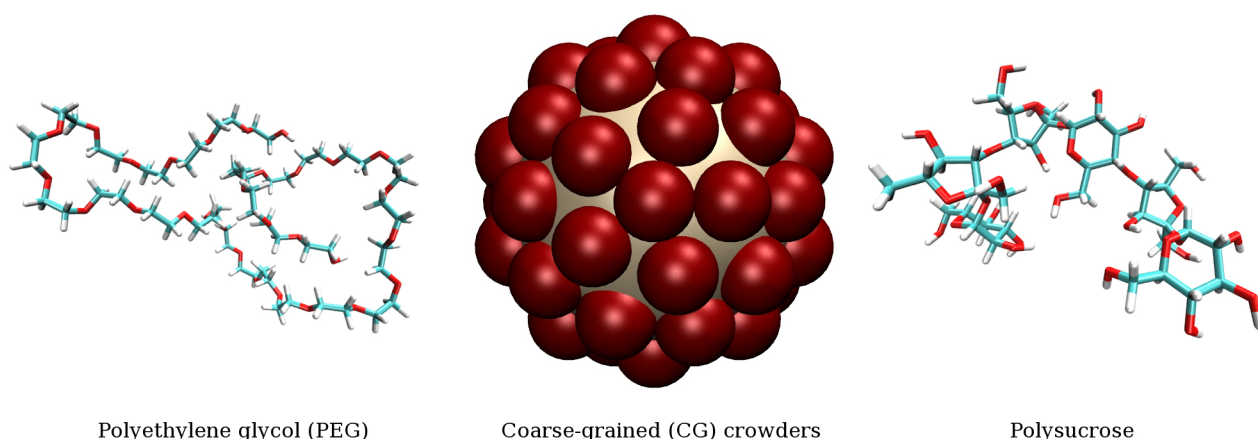


Figure 3.1: Three models of crowders used in the simulations of the NS3/4A protease in crowded environment: Polyethylene glycol (PEG) and Ficoll molecules in all-atom representation and the coarse-grained crowder built with 42 carbon-sized pseudo-atoms distributed evenly on the surface of the sphere, and an additional large pseudo-atom in the center.

The model of PEG crowders was used in an interdisciplinary study of the effect of crowding on the activity of the NS3/4A protease, which was published in *Biochimie* in 2020. Both PEG

and coarse-grained crowdres were used in simulations aiming to understand the effects of different aspects of crowding on the dynamic properties of the protease. This study was published in *Biophysical Journal* in 2021, however the properties of CG model were reported only briefly. Here I describe the properties of both PEG and CG crowdres in detail, I also include a description of the process of the CG model design that finally led to the choice of the bead-shell model instead of simple one-bead spheres.

The model of polysucrose and part of the results obtained with this model were the subject of an undergraduate thesis written by Anna Skrzyszowska, for whom I was co-supervisor, together with prof. Joanna Trylska. The thesis has been successfully defended at the Faculty of Physics, in September 2021, however the results have not been included in any journal publication yet. In the current and the following chapters I present the properties of Ficoll and a thorough comparison of its interactions with interactions displayed by PEG and CG crowdres.

3.2 All-atom PEG crowdres

Motivation for incorporating PEG in a molecular dynamics simulation is to view a single crowding experiment in two different scales - this is achieved by transferring the activity assay conditions to the MD simulation. In the laboratory assay, the outcome of collective action of many NS3/4A proteases is observed, whereas in the molecular dynamics simulation only one enzyme is viewed, but with atomistic detail. This zoom-in allows observation of changes in the diffusion rate and in the conformational dynamics of the enzyme, which is not possible during the activity assay.

Poly(ethylene glycol) (PEG) polymers of various sizes are among the most commonly used crowder molecules⁴⁶. Their advantage is that they are well soluble even at high concentrations⁹⁹. PEG interactions with the proteins are limited, although not reduced to zero - experimental data show, that it can not be considered an inert crowder. Protein-PEG interactions have been confirmed by isothermal titration calorimetry (ITC)¹⁰⁰, NMR spectroscopy^{101,102} and X-ray crystallography¹⁰³, showing that PEG preferentially binds to hydrophobic residues¹⁰². Polar interactions between bovine serum albumin and PEG, including formation of hydrogen bonds, were observed both in experimental studies and in molecular docking experiments¹⁰⁴.

The belief that PEG does not interact with proteins comes largely from the fact that it shows anti-fouling properties - a layer of PEG covering a surface would act as repellent of proteins and micro-organisms¹⁰⁵, which is useful both for biomedical materials like implants and biosensors, and for example in protecting ship hulls from marine biofouling.

A very unusual and useful property of PEG polymers is that they are considered amphiphilic molecules - they can dissolve in both organic and aqueous solutions, even in solutes such as toluene, ethanol and acetone¹⁰⁶. PEG molecules are composed of a number of interior $-CH_2 - O - CH_2-$ groups and two $-CH_2 - OH$ end groups that possess a hydrogen bond donor in their hydroxyl group. The interior groups, on the other hand, show much more favourable group interaction potential with aliphatic C¹⁰⁷, contributing to hydrophobic character of the molecule. The ratio of end to

interior groups is connected with hydrophilic properties of smaller PEG molecules and hydrophobic properties of larger ones. PEGs with low molar weight of up to 8000 MW are mainly hydrophilic, whereas long PEGs, with the mass over 20 000 MW show prevalently hydrophobic interactions with the proteins^{104,108}.

3.2.1 Molecular models of PEG - topology and parameterisation

In the simulations presented in this work, 28-mer of PEG was used - it contains 26 interior $-CH_2 - O - CH_2-$ groups and a $-CH_2 - O - CH_2 - CH_2 - OH$ group at each end, giving a total of 28 $-CH_2 - O - CH_2-$ groups, hence the '28-mer' term. The mass of such molecule equals 1295.5 Da, which places it between the PEG molecules used in laboratory experiments - PEG 600 and 6000 in the activity assay and PEG 600 and 1200 in CD experiments. The main motivation for choosing precisely this polymer length was due to the fact that for 28-mer the hydrodynamic radius of the molecule is known - this radius has been used to design spherical coarse-grained crowders of a size corresponding to all-atom PEG polymer.

For model building and parameterisation of PEG molecules, I collaborated with a research group from Bethesda (USA), lead by prof. Richard Pastor. The group conducted all-atom molecular dynamics simulations of a single 28-mer of PEG in water⁷⁵, from which they determined, among other parameters, the mean hydrodynamic radius of the polymer. Conformations sampled from those trajectories, as well as optimised charges and force field parameters were used to prepare models of PEG crowders included in my simulations. Five conformations used as starting structures in the simulations of the NS3/4A protease surrounded with PEG crowders are shown in Figure 3.2.

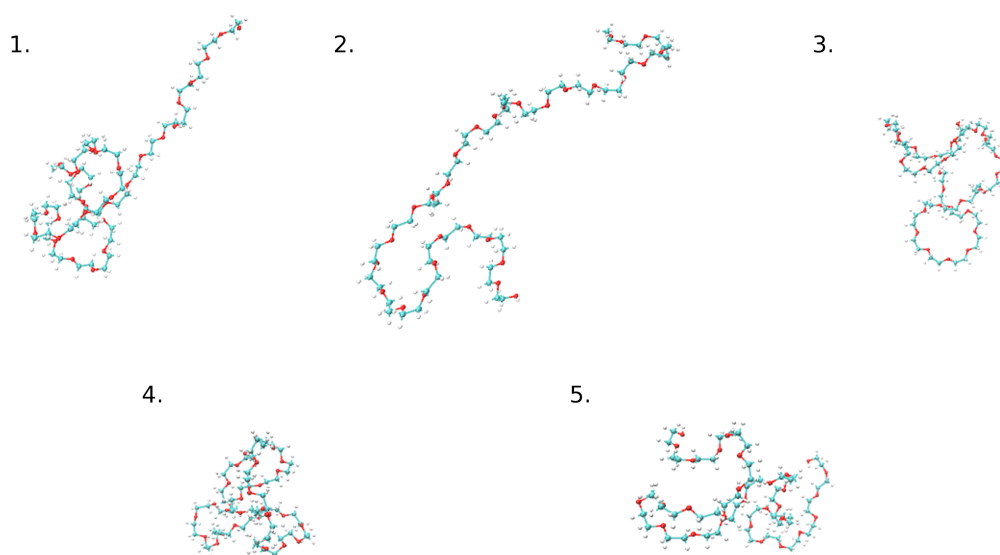


Figure 3.2: Conformations of the 28-mer of PEG sampled from a molecular dynamics simulation conducted by H. Lee et al.⁷⁵. The conformation set comprises of both extended and folded polymer chains - in the process of surrounding the protease with multiple PEG crowders, one of those conformations has been chosen randomly upon placing each crowder around the NS3/4A protease.

Topology and force field parameters of PEG polymers were constructed based on CHARMM36 force field, with optimised partial charges and torsion angles. Specifically, the PEGM residue from the *topall36_genff.rtf* topology file was used for interior CH_2-O-CH_2 groups, and the topology for $OH-CH_2-CH_2-O-CH_2$ terminal groups was written from scratch. Partial charges on the atoms were modified to match the charges used in the single PEG simulations⁷⁵. To parameterise the model, standard CHARMM36 force field parameters were used, with modifications of the torsion angles based on the work of A. Leonard et al.¹⁰⁹. Finally, the topology in the following form was used:

```
RESI PEGX      0.0000
```

! RESI PEGX is the topology of a single internal monomer of PEG

```
GROUP
```

```
ATOM C1  CG321  0.40
```

```
ATOM H1B  HGA2  -0.0575!      H11      H21
```

```
ATOM H1A  HGA2  -0.0575!      |      |
```

```
ATOM O1  OG301  -0.57  ! (-C2) - C1 - O1 - C2 -
```

```
ATOM C2  CG321  0.40  !      |      |
```

```
ATOM H2A  HGA2  -0.0575!      H12      H22
```

```
ATOM H2B  HGA2  -0.0575
```

```
BOND C1 H1A  C1 H1B  C1 O1
```

```
BOND O1 C2  C2 H2A  C2 H2B  C1 -C2
```

```
PATCHING FIRST PEG1 LAST PEG2
```

```
PRES PEG1      -0.00
```

! PEG1 and PEG2 are patches for the terminal groups of PEG. Patches comprise of entries for the atoms, whose type or charge changes in the terminal group, relative to interior groups, as well as entries for new atoms, that do not occur in the interior group. Total charge of the terminal $OH-CH_2-CH_2-O-CH_2$ parts equals zero after considering all atoms present in those groups.

```
ATOM C0  CG321  0.40
```

```
ATOM H0A  HGA2  -0.01
```

```
ATOM H0B  HGA2  -0.01
```

```
ATOM OG0  OG311 -0.65
```

```
ATOM HG0  HGP1  0.08
```

```
ATOM H1B  HGA2  -0.0100
```

```
ATOM H1A  HGA2  -0.0100
```

```
ATOM H2A  HGA2  -0.0100
```

```

ATOM H2B HGA2 -0.0100
BOND C0 C1 C0 H0A C0 H0B OG0 C0 OG0 HG0

```

```

PRES PEG2 -0.00

```

```

ATOM C3 CG321 0.40
ATOM H3A HGA2 -0.01
ATOM H3B HGA2 -0.01
ATOM OG3 OG311 -0.65
ATOM HG3 HGP1 0.08
ATOM H1B HGA2 -0.0100
ATOM H1A HGA2 -0.0100
ATOM H2A HGA2 -0.0100
ATOM H2B HGA2 -0.0100
BOND C3 C2 C3 H3A C3 H3B OG3 C3 OG3 HG3

```

! IC entries were present in the topology file, although the values were set to 0.

3.2.2 Aggregation of PEG molecules

In the first set of simulations of the NS3/4A protease in the presence of PEG molecules, I encountered a very significant issue. After only about 50 ns, the PEG molecules started to form large but not dense clusters, to aggregate completely between 100 and 200 nanoseconds (Figure 3.3); clusters of crowders formed in all three copies of the simulation, at a similar location near the surface of the NS3 protease and the N-terminal tail of NS4A.

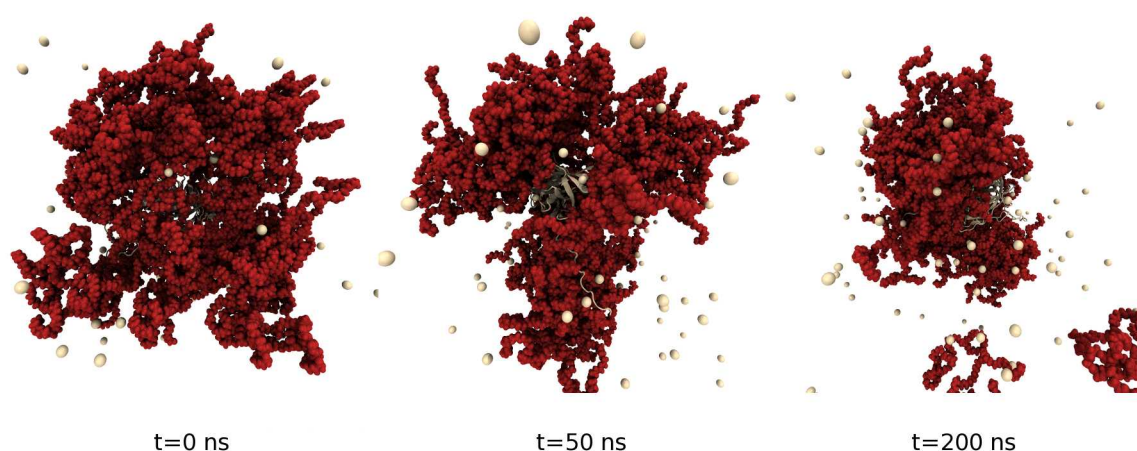


Figure 3.3: Sample screenshots from one of the trajectories of the NS3/4A protease surrounded by PEG molecules, in which standard parameters were used in the TIP3P water model. The protease and ions are shown in beige and the PEG crowders in red.

The first obvious explanation for such a problem could be an error in the parameterisation of the PEG molecules, but exactly the same parameters were already used in molecular dynamics simulations by the group of prof. Richard Pastor. However, these were simulations in which the system consisted of only one PEG polymer surrounded by water molecules⁷⁵.

The literature research established that aggregation in MD simulations of multiple molecules is a known issue, caused not by the parameterisation of aggregating molecules, but by sub-optimal definition of the interactions between the water and other molecules. It turns out that in many force fields the strength of interaction of water with other molecules is underestimated - in other words, the hydrophobicity of molecules is overestimated. This problem does not interfere with simulations of a single globular protein, but when simulating several molecules, it leads to excessive association, as well as altered conformational dynamics, especially in case of the unstructured fragments.

3.2.3 Modification of force field parameters of water restores realistic diffusion of PEG crowdres

The first work that raised the question of whether modern force fields are suitable for simulating multiple proteins was a 2014 publication by D. Petrow et al.⁷⁶; its authors noted that while there are examples of simulations of several proteins in explicit solvent, from which conclusions are drawn on their aggregation and interactions, there is no negative control, i.e., a simulation of non-aggregating peptides. They performed such simulation, using a protein called villin, and its carbonylated counterpart, that forms aggregates in vitro. Non-carbonated villin is a small protein (36 aa), which is an example of a protein that does not aggregate even at very high concentrations - that is why it later became a model system for studying aggregation in MD simulations. The first simulation of villin ensemble, conducted in the GROMOS force field, showed that the non-aggregating villins under natural conditions aggregate as much or even more than their modified counterpart. All villins present in the system formed a single aggregate in 22 out of 26 simulations in GROMOS, and 65% of the proteins did not dissociate from the aggregate until the end of the simulation. The simulation was repeated for Amber, CHARMM22 and OPLSA force fields, all with exactly the same results. In each case, the formation of the aggregate proceeded with a decrease in the potential energy of the system.

Another study published in the same year highlights a slightly different aspect of inaccuracies in the force fields - observations showing that proteins in simulations are too poorly solvated relative to experimental data⁷⁷. A factor that can affect both protein association and solvation energy is the hydrophobic effect, and the authors concluded this might be where the problem lies. The only place where we can attempt to model hydrophobic effect in traditional force fields is in the parameters of solute interactions with water molecules, specifically the epsilon parameter for water atoms, which is an input to calculating the 'strength' of the Lennard-Jones potential of interactions between the water and solute atoms. Indeed, upscaling this parameter resulted in agreement of the simulation results with experimental data on protein size, solvation energy and protein complex formation.

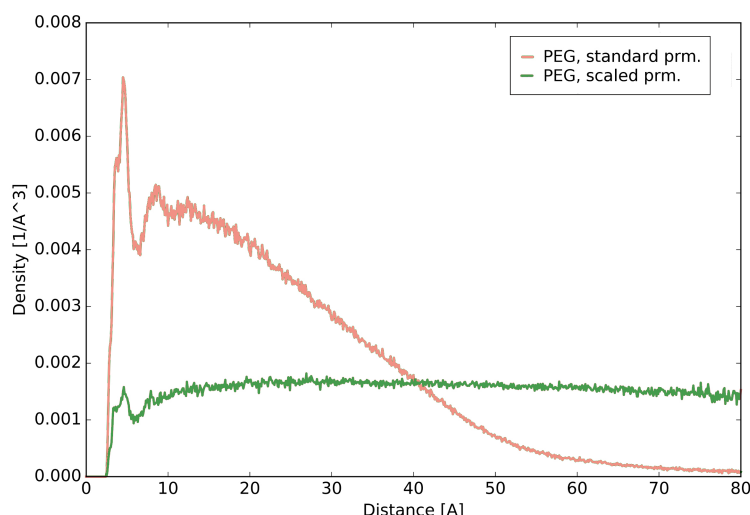


Figure 3.4: Radial distribution function (RDF) calculated for the distribution of water oxygen atoms around heavy atoms of the NS3 protease, in the simulations with standard and scaled force field parameters of water. For the simulation using the standard TIP3P water parameters, the radial distribution function gradually decreases and drops down to zero, which corresponds to a large cluster of crowdres that forms near the surface of the protein. With scaled parameters, the graph shows an even distribution of crowdres and free diffusion of PEG molecules, consistent with the experimental data.

In the simulations presented here, the VdW interactions between solute and water molecules were upscaled by the factor of 1.09, based on optimisation conducted for the CHARMM force field¹³. In the optimization process, such a scaling factor allowed the elimination of excessive aggregation of the test peptides, and was low enough not to destabilize their native structure¹³.

Technically, the LJ epsilon parameter of the water oxygen atom was increased from -0.1521 to -0.1807 *kcal/mole* and the epsilon parameter of water hydrogen atoms from -0.046 to -0.05466. This way the calculated water-solute LJ potential is upscaled by the factor of 1.09. Epsilon values for water atoms were calculated taking into account the equation for epsilon parameter for a pair interaction:

$$\epsilon_{i,j} = \sqrt{\epsilon_i \epsilon_j} \quad (3.1)$$

where ϵ_i and ϵ_j are values specific for a pair of interacting atoms. The parameters were upscaled for the interactions with all atom types present in the system, then for the water-water interactions it was reduced back to the original values using the NBFIX entries that allow modifying the nonbonded LJ parameters for selected atom pairs. Such small modification allowed to restore realistic diffusion and distribution of all-atom crowdres, observed to aggregate when simulated in the standard TIP3P water model (Figure 3.4). The comparison between the modified water parameters and original TIP3P water is shown below:

```
*** Original TIP3P parameters ***
!atom ignored epsilon  Rmin/2
!          (kcal/mol)  (Å)
```

```

HT    0.0   -0.046   0.2245
OT    0.0   -0.152   1.7682
*** Modified water parameters ***
HT    0.0   -0.05466  0.2245
OT    0.0   -0.1807   1.7682
NBFIX
!      Emin    Rmin
HT  HT   -0.046   0.449
HT  OT   -0.0836  1.993
OT  OT   -0.1521  3.5364

```

3.3 The coarse-grained Crowdiers

The purpose of introducing the coarse-grained model was to identify which elements of the effects of the presence of PEG result from volume exclusion, and which from the interactions between the protease and PEG.

Hence the CG crowdiers aim to represent folded 28-mers of PEG in every aspect but the charge-charge protein-PEG interactions: they retain the radius of folded polymer and its hydrophilicity, as well as minimal interactions with the protein in the form of LJ parameters consistent with the PEG carbons parameters. The pseudo-atoms making up the crowdiers lack any partial charges, therefore their interactions with the protease are much weaker. Thus, in the assumptions of the model, the influence of CG crowdiers on the protease dynamics is in the vast majority due to reduction of the volume available for the protein.

It is worth noting that in the case of simulations described in this thesis, the reduced CG crowder model is used in an otherwise full-atomic systems, that also include explicit solvent. It is quite an unusual multi-scale solution, used to maintain the effects of solvent dynamics and protein interactions with individual water molecules. Some technical aspects of the model finally used result directly from the necessity to adapt it for use in full atomistic simulation.

This chapter details the process of designing coarse crowdiers, with consecutive stages of design shown in Figure 3.5. It also describes the encountered issues and strategies for addressing them.

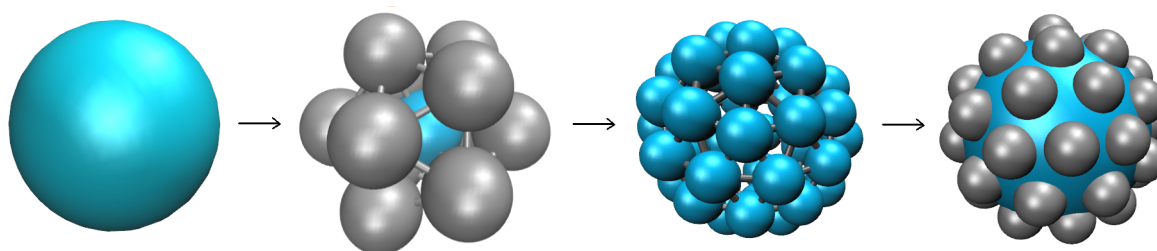


Figure 3.5: CG crowder models at subsequent stages of design.

3.3.1 Single sphere crowdors

The basic, simplest model of a coarse-grained crowder is a single sphere that represents the entire molecule that makes up a crowded environment. Such model has been previously used in many simulations, studying for example the effects of crowding on protein stability^{20,19}, protein complex formation^{23,110} and inter-domain mobility²⁷.

To implement this model in the simulations of the NS3/4A protease, I used spheres of a radius corresponding to a folded 28-mer of PEG (see Figure 3.6) which, according to a single PEG simulations, equals $7.7 \text{ \AA}$ ⁷⁵. Each crowder was modelled with a single enlarged pseudo-atom of type **PE** developed specifically for the needs of the simulations carried out as part of the doctoral project.

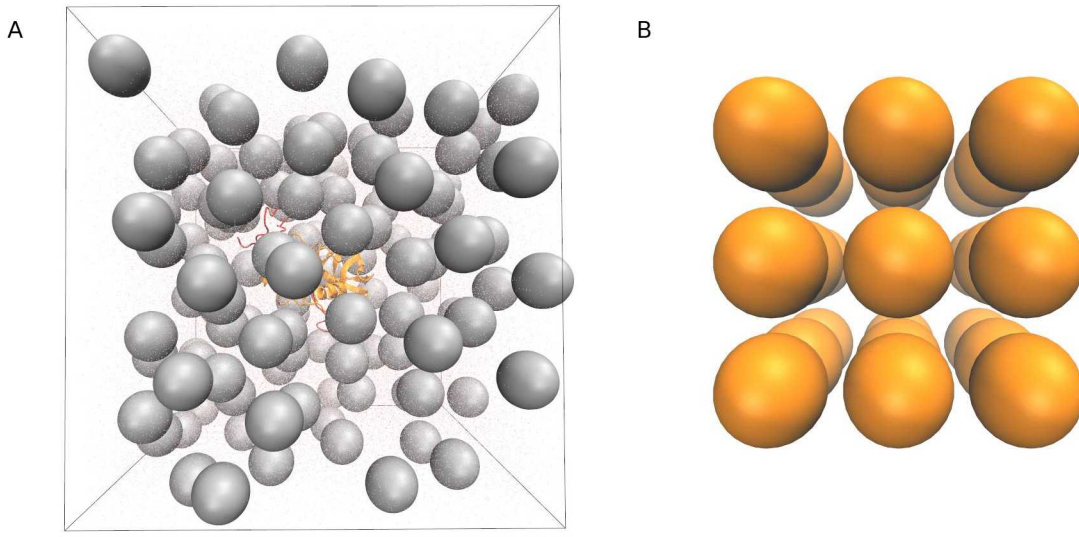


Figure 3.6: A. The all-atom model of NS3/4A surrounded with 130 single-particle crowdors; B. a test system on which the parameters of the model were optimised, containing 27 crowdors in water.

Since the pseudo-atoms have no partial charges, their interactions with other molecules are reduced to the Lennard-Jones potential U_{LJ,PE_i} in the form:

$$U_{LJ,PE_i} = \epsilon_{PE_i} \left[\left(\frac{R_{min,PE_i}}{r_{PE_i}} \right)^{12} - 2 \left(\frac{R_{min,PE_i}}{r_{PE_i}} \right)^6 \right] \quad (3.2)$$

$$\epsilon_{PE_i} = \sqrt{\epsilon_{PE} * \epsilon_i}$$

$$R_{min,PE_i} = \frac{R_{min}}{2} \epsilon_{PE} + \frac{R_{min}}{2} \epsilon_i$$

where ϵ_{PE} and $\frac{R_{min}}{2} \epsilon_{PE}$ are the parameters of the pseudo-atom PE, included in the NONBONDED section of the CHARMM parameter file read by the NAMD simulation engine. In NAMD it is possible to define different parameters for interactions with different molecules, for example, a separate set of ϵ and $\frac{R_{min}}{2}$ for the crowder-protein interaction and a separate set for the crowder-crowder interaction. However, this does not change the fact that for a reduced crowder, the entire mechanics of an interaction has to be described with a very limited set of numbers.

The function of the ϵ_{PE_i} parameter is quite simple - in practice it describes the strength of the interaction between the crowder and other atoms. The function of R_{min,PE_i} is slightly more complicated (Figure 3.7) - it is not the distance between the crowder and other atoms, at which their mutual attraction peaks. $\frac{R_{min}}{2}_{PE}$ is the contribution to R_{min,PE_i} , and R_{min,PE_i} is the distance from the centre of the pseudo-atom PE at which the (U_{PE_i}) potential crosses the X axis. In practice, $\frac{R_{min}}{2}_{PE}$ describes the "softness" of the U_{LJ,PE_i} potential - the shape of the LJ potential function.

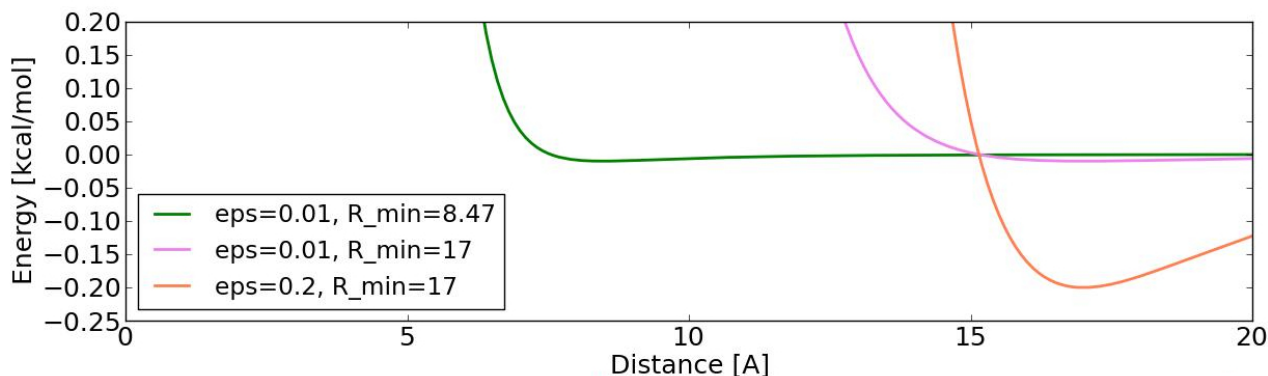


Figure 3.7: Examples of the LJ potential for the crowder pseudo-atoms of the radius of 7.7 Å, illustrating how the changes to ϵ_{PE} and $\frac{R_{min}}{2}_{PE}$ parameters affect its shape. As for the initial parameters the R_{min} for the crowder-crowder interaction would be calculated as $8.47 + 8.47$, the LJ potential for this interaction would resemble the violet graph. A comparison between the green and violet graphs - graphs with identical ϵ parameters - shows very clearly, that the ϵ parameter does not only define the strength of interaction, but also steepness of the graph - a low ϵ would result in 'softer' potential, allowing the centers of interacting atoms to locate closer than the sum of their radii.

As spherical crowders ought to only exert an excluded volume effect on the protein, initially the ϵ_{PE} parameter of the PE pseudo-atoms was set to a relatively small value of 0.01, minimizing the attraction to other molecules. $\frac{R_{min}}{2}_{PE}$ was set to 8.47, which is 110% of the radius of 28-mer of PEG. However, with these parameters, already at the stage of system equilibration single sphere crowders were observed to aggregate and overlap with each other.

To solve the aggregation problem, I built a test system containing 27 crowder particles in water. In order to facilitate the debugging process, in the parameter file I separated the crowder interactions into interactions with the protein (main parameters in the Nonbonded section), interactions with water and crowder-crowder interactions, described both with NBFIXes - parameter file entries defining modified LJ parameters only for specific pairs of atoms. Using such a test system (Figure 3.6), a series of modifications were tested, including:

- **Increasing the cutoff** for calculating nonbonded interactions from 12 to 19 Å (Figure 3.8). Due to the large size of the crowders, the distance between their centres (>15.4 Å) was greater than the cutoff distance of 12 Å. For this reason, crowder A only "noticed" crowder B when the crowders overlapped. The cutoff has been set to 19 Å, which is larger than the sum of the crowder radii plus additional 2-3 Å distance between them. Following this change, the pairlistdist parameter in NAMD configuration file was increased from 14 to 21 Å.



Figure 3.8: An illustration of the cutoff problem; with the standard LJ cutoff of 12 Å two spherical crowders with the radius of 7.7 Å start to "feel" the VdW repulsion between each other when they already overlap (A). With the cutoff increased to, say, 19 Å (B) crowders see each other and 'feel' the VdW interactions before they start to overlap, and even before the LJ potential starts to be repulsive. Another way of solving the cutoff problem is to reduce the radius of interacting pseudo-atoms - this method was used in the later stages of crowder design, by placing a set of smaller pseudo-atoms on the surface of the sphere.

- **Optimization of the crowder-crowder NBFIX parameters** by testing various combinations of R_{min} and ϵ values. Following the hypothesis that soft LJ potential might allow overlap of the PE pseudo-atoms, I also tested a substantially increasing ϵ value, making the potential steeper and the crowder particles 'harder', as well as increasing R_{min} to values larger than the sum of PE radii, to enforce distancing between interacting crowders,
- **Replacing the Lennard-Jones potential** for the PE - PE interaction for only the repulsive part of the U_{LJ} , introduced using the NB-Table functionality,
- **Exchanging the water model** from TIP3P to TIP4P¹¹¹,
- **Modifying the crowder-water interactions** to decrease hydrophobicity of uncharged PE pseudo-atoms.

Increasing hydrophilicity combined with enforcing larger distance between crowders proved successful in avoiding aggregation of single-sphere crowders; however, forcing physical distancing posed a risk of disrupting crowder diffusion. Another significant problem of this crowder model are large distances between the centres of individual crowders. To ensure that the crowders "feel" each other before they overlap, the nonbonded cutoff has been increased to 19 Å and the pairlistdist to 21 Å (Figure 3.8) - this parameter sets the distance up to which the interacting atom-pairs are listed. Nonbonded interactions involve interactions of atom i with every other atom present within a cutoff radius. As in the all-atom system, the 12 Å sphere around a given atom contains roughly 700 atoms, and the 19 Å sphere contains about 3000 atoms, with increasing the cutoff the number of interacting atom-pairs would quadruple. Therefore increasing the cutoff would dramatically increase the number of calculations performed in each simulation step, significantly slowing down the simulation.

3.3.2 12+1 crowder model

A way to avoid increasing the cutoff distance for calculating nonbonded interactions is to reduce the radius of interacting pseudo-atoms, so that the sum of their radii is less than the standard cutoff value, for example, by replacing a single large sphere representing a crowder with a set of smaller spheres. As a preliminary test I built a model combined of twelve pseudo-atoms surrounding a central atom, all with a radius of 3.5 Å. In this model, each pair of atoms was connected via bond, with the length of about 8 Å. The initial version of this model was tested on a system containing two crowders surrounded with water, as in Figure 3.9.

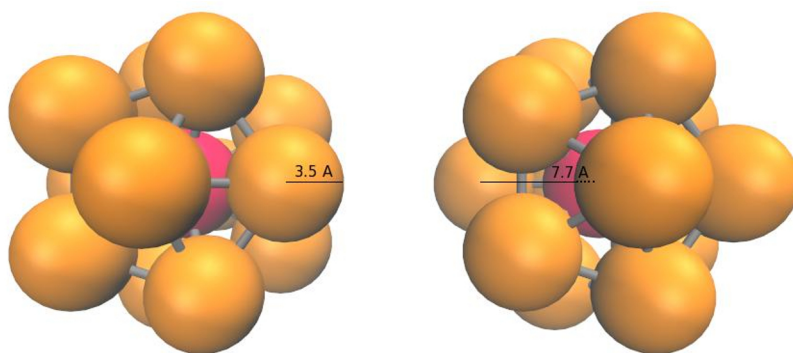


Figure 3.9: A test model of a crowder consisting of twelve pseudo-atoms evenly distributed around the central atom. In this model, all pseudo-atoms had a radius of 3.5 Å and were connected via standard harmonic bonds.

Unfortunately, the simulation with this particular model turned out impossible to be carried out successfully - all the test simulations ended with "Bad global exclusion count" errors at random times. As NAMD troubleshooting¹¹² documentation explains, such errors can be caused by either instability of the starting system, or the presence of excluded atoms, that are located on non-neighboring patches. Excluded atoms are atoms distanced by up to 4 bonds, for which nonbonded interactions are not calculated. In the 12+1 crowders, pairs of excluded atoms are separated by a maximum of the crowder's diameter, so it is possible that they will be localised in non-neighboring patches - in such case the simulation would terminate every time this happens. The solution could be to **switch off the atom exclusion**, also for the protein, or **decrease the distance between atoms on the surface of the crowder**. The other solution, which was not taken into consideration, could be running the simulation on a single processor, so the system would not have to be divided into patches.

3.3.3 42+1 bead-shell crowders

To decrease the distances between pseudo-atoms on the surface of the crowder, the sphere built with 12 atoms was replaced with 42 atoms distributed on the surface of the sphere (Figure 3.10 A), with the radius of the sphere set to 6 Å. New atoms were significantly smaller, with the VdW radius similar to carbon atoms (2.01 Å), making the overall radius of the crowder similar to the first

single-sphere model. In this way, the sizes of the pseudo-atoms and the distances between them have become comparable to those found in other molecules such as proteins.

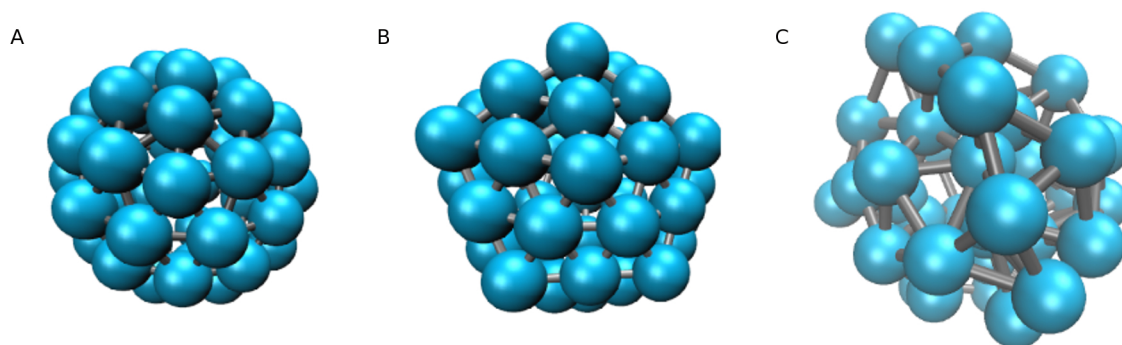


Figure 3.10: Conformations of the crowders built with 42 carbon-sized pseudo-atoms distributed on the surface of the sphere; A. the starting structure, B. the conformation adopted after minimization and C. structure collapsed inwards during the thermalization process.

The first version of the 42 beads crowder that was tested, was a shell built with carbon-sized atoms (of the artificial type BS) with a void inside (Figure 3.10 A). The atoms were connected with neighbors via "covalent" bonds. However, during the minimization procedure, the crowders adopted an icosahedral shape (Figure 3.10 B). This structural change was unexpected, but proved to be easily explainable from a mathematical point of view. Twelve is the maximum number of points that can be placed evenly on a sphere, so the distances between 42 pseudo-atoms of the crowder could not be exactly equal. The crowder had to adopt the conformation that would be energetically optimal, and the resulting structure appeared to be similar to naturally occurring structures such as, for example, viral capsids.

During solvent thermalization, the crowders that were empty inside collapsed inwards (Figure 3.10 C), likely due to water pressure and/or the VdW interactions between atoms distanced by more than 4 bonds. Therefore, to prevent the structure from collapsing, an additional atom was placed in the center of a crowder. The central atom's radius has been set to 6 Å - this way it would ensure the spherical structure of the crowder (Figure 3.11).

The additional atom stabilized the structure, but the problem that occurred with 12+1 crowders has not been eliminated - the bonds between the atoms with R_{min} of 2.01 Å (carbon-sized) turned out to be still too long. As I did not want to further reduce the radius of the pseudo-atoms, I decided to remove the bonds from the model, therefore completely eliminating the risk of 'exclusion count' error. For this purpose, I tested three strategies:

- **Extra bonds** are an additional harmonic potential that one can introduce between two atoms to, for example, make the protein conformation more stable. In my model I replaced all bonds with extra bonds. However, at the beginning of thermalization global bond count error has been reported by NAMD, which according to NAMD documentation indicates, that the extra bonds were too long.

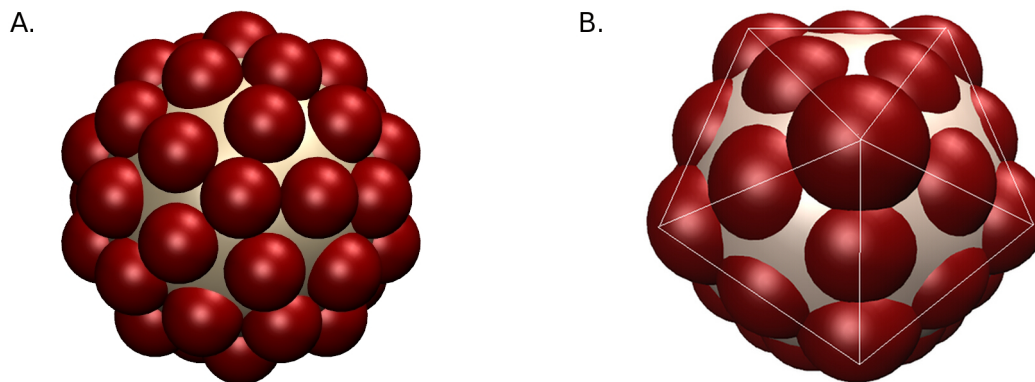


Figure 3.11: Additional large pseudo-atom in the center of the crowder ensures stability of the spherical structure of the model. During the simulation in explicit water the 42+1 model adopts the shape of an icosahedron, similar to the shape of the previous version adopted during thermalisation process.

- **NB-Table**, with the Van der Waals interactions between surface atoms replaced with tabulated harmonic potential in the form:

$$V_{BS-BS}(r) = k(r - r_{ref})^2 \quad (3.3)$$

NAMD NB-Tables functionality requires providing an energy file, with a table consisting of entries with the values of potential and force to be applied at specific distance from each atom. Energy was calculated using the formula for V_{BS-BS} harmonic potential, with the harmonic constant k set to 100 and the minimum distance r_{ref} to $3.7 \text{ \AA}$ - an average distance between the surface atoms. The force was calculated as a derivative of the V_{BS-BS} potential. Harmonic potential should work as a pseudo-bond, preventing the surface atoms from moving away from their neighbors. Although this solution seems simple, during simulations using pseudo-bonds of this type, changes in the behaviour of water around the crowder were noticed - a vacuum layer was formed near the surface of the crowder and the water pressure in the remaining regions increased. Unfortunately, it was not explained what caused such effects.

- **NBFXes as pseudo bonds**. The third and most successful approach was using NBFX entries instead of bonds. Strong VdW type interactions appear to be enough to keep the crowder's structure intact. Furthermore, the length of such pseudo-bonds is limited only by the cutoff value.

During the simulations of crowdres with pseudo-bonds I noticed that when all crowder atoms are described with a single atom type (BS), crowder A atoms can "form a pseudo-bond" with crowder B atoms. In other words, upon contact between crowdres A and B, the atoms of crowder A "feel" the LJ potential representing the pseudo-bond in crowder B and an attraction between the atoms of both crowdres occurs. This can result in unwanted clustering of crowder molecules. To address this issue, I decided to separate the atom types in the crowdres and assign crowder 1 a BS1 type for all

its surface atoms and a PE1 type for its central atom. In this way, crowder N will have surface atoms of type BSN and a central atom of type PEN.

This solution complicates the process of creating a parameter file for a system containing coarse-grained crowders because separate atom types are required for each crowder particle and the structure of such a file depends on the number of crowders in the simulated system. The solution I used was to write a program that generates a parameter file for a given number of crowder molecules. This program, *par_generator.py*, is attached in APPENDIX. A structure of the parameter file for the system containing N bead-shell crowders is presented and described in detail in the 'Final model of the CG crowders' section.

3.3.4 Artificial hydrophilicity of the CG crowders

An important element in preventing the aggregation of single-sphere crowders, identified during test simulations, was modification of crowder interactions with water molecules. It turned out that to a large extent it is the hydrophobicity of the crowders that is responsible for their aggregation, while hydrophobicity is a direct result of the lack of charges or partial charges on the crowder pseudo-atoms. The most fundamental property of water molecules is their polar nature, resulting from a shift in electron density from the two hydrogen atoms to the electronegative oxygen atom. The resulting dipoles allow for the very characteristic structure of water, in which molecules form a series of hydrogen bonds with each other. The OH dipoles also define the way of interactions between water and other molecules - through charge-charge interactions, polar groups and molecules enter into a "net of interactions" formed by water molecules, while the non-polar molecules are repelled, leading to their clustering.

In order to obtain crowders that do not aggregate and retain realistic diffusivity, it was necessary to understand a little more deeply what hydrophobicity and hydrophilicity are - so that it could be encoded in the very limited number of parameters that the NAMD program takes as an input.

The first task was to find a way to compare the hydrophilicity of two molecules - in this case PEG and the CG crowder that represents them. For this purpose, it is required to find a measure which describes hydrophilicity and which can be determined from a fragment of the molecular dynamics trajectory. Such a measure may be the distribution of water molecules around the molecule under study, described by the Radial Distribution Function (RDF). It describes the density / number of water molecules as a function of the distance from the molecule. At the surface of a hydrophilic molecule, such as PEG, a layer of water molecules directly interacting with it will form a peak on the RDF graph (the first hydration layer), then the density of water molecules will decrease and increase at the level of the next layer, and further on at a larger distance from the molecule the RDF will level off around a certain constant value.

To quantify the level of interaction of the PEG molecule with water, I decided to determine the average height of the first peak of the RDF function from the trajectory, and then to choose the LJ parameters for the crowder atoms in such a way that in the test simulation the height of the first peak

Atom selection	ϵ	RDF peak
PEG		
type C1		0.7
type C2		0.65
Crowders		
resname BSP	0.1521	0.55
	0.2	0.65
	0.3	0.675
	0.6	0.95
	1.2	1.25
resname BSP and not type PE	0.1521	0.55
	0.2	0.65
	0.3	0.675
	0.6	0.975
	1.2	1.3

Table 3.1: The height of the first RDF peak for the PEG carbons and for the crowdres with different values of the $\epsilon[kcal/mole]$ parameter. Atom selection is the VMD selection for which the water's oxygen RDF was calculated, i. e., the distribution of water oxygen atoms around carbon atoms of PEG and around the BS surface atoms of the coarse-grained crowdres (the crowder residue is called BSP and PE is the large central atom of the crowder).

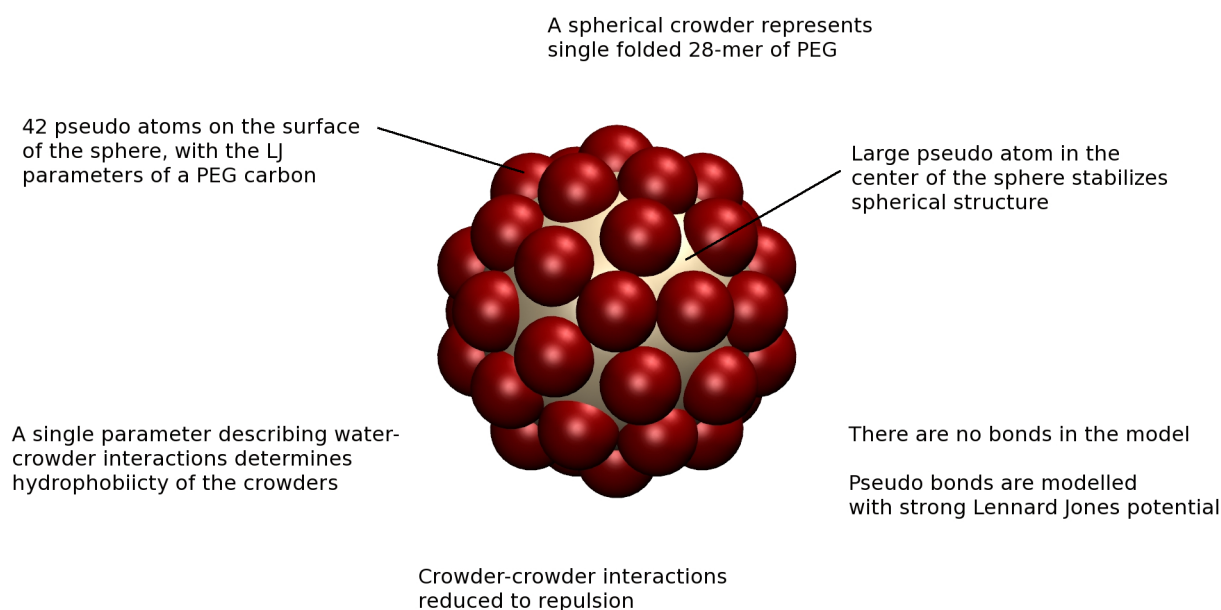
was as close as possible to the value for PEG. I also took into account the comparison of the shape of the function for both molecules.

The RDF for the distribution of water around the crowder molecules was calculated for short simulations of two separated crowdres in water with a set of different ϵ values. The RDF peak height values for the PEG and test simulations of the CG crowdres with different ϵ values for the crowder-water interactions are presented in Table 3.1. RDF was calculated using VMD Radial Pair Distribution Function extension. To obtain comparable functions, RDF was calculated for the carbon atoms of the inner monomers of PEG (type C1 and C2) and the surface pseudo-atoms of the coarse-grained crowdres.

Based on this procedure I decided to set the ϵ value for the water-crowder interaction to 0.3. Thus, in summary, the artificial hydrophilicity of coarse-grained crowdres was implemented by assigning a separate NBFIX entry in the CHARMM parameter file to the crowder-water interaction, modifying the interaction between the atoms on the crowder surface (type BS_i) with the oxygen atoms of the water molecules (type OT). As the ϵ parameter indirectly describes the interaction strength of a given pair of atoms, the modification of this parameter increased the interaction strength of the crowder with water in such a way that the distribution of water molecules around the crowder was as close as possible to the distribution of water around the PEG molecule.

3.3.5 Final model of the CG crowder

The model finally used for simulations of the NS3/4A protease in crowded environment was based on the 42+1 model - a model in which 42 carbon-sized atoms are distributed on the surface of a sphere, with an additional large atom in the centre of the crowder, however parameterisation of this model benefits from observations and optimisations obtained at all stages of the CG model design. In order to refer to the practical implementation of the model, the properties of the final version are included as comments in the individual sections of the template CHARMM parameter file for such a crowder.



* CG crowder parameters *

ATOMS

* Mass of a crowder equals $246 + 42 \times 25$ [Da], which corresponds to the mass of a 28-mer of PEG (1295.5 Da). The mass of the central atom is larger to slow down its movements and keep the crowder molecule as stable as possible.

MASS -1 PE1 246.000

MASS -1 PE2 246.000

...

MASS -1 PEN 246.000

MASS -1 BS1 25.0000

MASS -1 BS2 25.0000

...

MASS -1 BSN 25.0000

ANGLES

* The model has no bonds, so it also has no angles between bonds. However, for technical reasons, angles set to zero are included in the parameter file as they have to be included in the psf file

BS1 BS1 BS1 0.0 0.0

BS2 BS2 BS2 0.0 0.0

...

BSN BSN BSN 0.0 0.0

NONBONDED

* The standard parameters in the NONBONDED section apply only to interactions with protein atoms - for the other interactions separate parameters have been set in the NBFIX sections. The surface pseudo-atoms BS1, BS2... BSN are parameterised as PEG carbons. The same parameters are set for the PE1, PE2... PEN atoms, however, due to the large distances between the central atom and the closest atoms of the protein, they do not affect the crowder protein interactions.

PE1 0.000000 -0.0500 2.0100

PE2 0.000000 -0.0500 2.0100

...

PEN 0.000000 -0.0500 2.0100

BS1 0.000000 -0.0560 2.0100

BS2 0.000000 -0.0560 2.0100

...

BSN 0.000000 -0.0560 2.0100

NBFIX

* in the NBFIX entries, the fourth column defines the equilibrium distance between the specified pair of atoms, and the third column defines the ϵ value for their interaction - unlike the previous section, the LJ parameters are not calculated here based on values for two individual atoms.

* the NBFIX entries for the central and surface atoms and for the pairs of surface atoms replace bonds between pseudo-atoms. A LJ potential much stronger than that between the pairs of atoms in real molecules holds the crowder parts in the desired configuration. Pseudo-bonds between the surface pseudo-atoms and central pseudo-atom of each crowder maintain nearly-spherical shape of the crowders and ensure their stability.

BS1 PE1 -100.00000 6.000

BS2 PE2 -100.00000 6.000

...

BSN PEN -100.00000 6.000

* pseudo-bonds between the pseudo-atoms on the surface of the crowder

BS1 BS1 -25.0000 3.700

BS2 BS2 -25.0000 3.700

...

BSN BSN -25.0000 3.700

* artificial hydrophilicity - water-crowder interactions are modified in such a way as to obtain a water distribution around the crowder similar to that of the PEG molecules. This procedure prevents aggregation and ensures realistic diffusion of the crowders.

BS1 OT -0.300 3.570

BS2 OT -0.300 3.570

...

BSN OT -0.300 3.570

* soft LJ potential for crowder-crowder interactions, with parameters optimised to eliminate crowder-crowder attraction. As each crowder has its own type of surface atoms, it is necessary to include all possible pairs of crowdors in this section.

BS1 BS2 -0.00001 12.00

BS1 BS3 -0.00001 12.00

BS1 BS4 -0.00001 12.00

...

BS(N-1) BSN -0.00001 12.00

3.3.6 Adding crowdors can actually speed up MD simulation

The natural environment in which proteins function is highly crowded - about 30% of the space in the cytoplasm is occupied by macromolecules such as proteins, lipids and RNA. Thus, simulating the dynamics of a protein in a crowded environment can add a lot of realism to the resulting trajectories, especially due to the fact that crowding influences both conformational dynamics and activity of certain proteins.

It is widely believed that simulating macromolecules in a crowded environment is very time consuming and computationally intensive. However, this may not always be true; while adding full-atom crowdors to the simulation actually increases the simulation time, adding coarse-grained crowdors may have the opposite effect.

For example, a spherical crowder, even in the 42 + 1 representation, contains less atoms than the same volume filled with explicit water molecules. For this reason, a simulation of the protein surrounded with CG crowdors and explicit solvent can be faster than a simulation without crowdors. According to my benchmark simulations, adding 42+1 bead shell crowdors in a volume fraction $\phi = 0.18$, speeds up the simulation by 20%, compared to the simulation of the same protein surrounded only with explicit water.

3.4 Ficoll crowdors

Simulations of the NS3/4A protease in the crowded environment modelled with Ficoll molecules represented with polysucrose were performed to explain the experimental results obtained in the activity assay, in which the activity of the protease was measured in water and in the presence of PEG and Ficoll⁹¹. This experiment showed that while PEG as expected decreases protein activity, the presence of Ficoll can increase it. This result was unexpected and not obvious to explain that we took the decision to run MD simulations to study the effect of Ficoll on the protein at the single molecule scale.

Ficoll is a polysaccharide consisting of polymerised sucrose molecules, formed by the reaction of sucrose with epichlorhydrin. The structure of resulting molecules is very diverse, with long, highly

branched chains classified by weight - hence terms such as Ficoll 400. The molecules are neutral, and due to the presence of many hydroxyl groups highly hydrophilic and well soluble in water. Ficoll is typically described as inert molecule, which means it does not interact with proteins and in a crowded solution acts as a good approximation of a hard sphere. However in the publications describing experiments with Ficoll crowders, the statement about Ficoll's inertia usually appears as a dogma, that does not require citation of the supporting work.

For the reason of the alleged inertia, Ficoll molecules have been widely used as crowders, which influence the protein only through volume exclusion. Although this is one of the most popular models in experimental studies^{113,6}, so far it has not been applied in molecular dynamics simulations.

3.4.1 Parameters of the model

The model of polysucrose molecule created for the purpose of the NS3/4A simulations was designed to match the volume of other crowders used in this study. A polymer consisting of three sucrose molecules turned out to be the closest to the size of a sphere of a radius of 7.7 Å. To obtain the initial version of the model, the sucrose molecules were aligned using Chimera, so that the points to be bonded were as close together as possible; for this purpose, a structure deposited in the PubChem database was used, with an ID of 5988.

The topology and parameter files for this structure were written based on CHARMM force field for carbohydrates, which includes parameters for building blocks of sucrose, glucose and fructose^{114,115}, and the linkers for building polysaccharide molecules⁷⁴. In the Ficoll molecules used as crowders in laboratory experiments, the sucrose monomers are connected with glycerol linkers, which form during the reaction of sucrose with epichlorohydrin¹¹⁶. Since the sucrose-glycerol linkages are not parameterised in the CHARMM force field, I decided to use a simplified model in which sucrose molecules are connected with glycosidic bond (Figure 3.12). For a further justification of this choice, and plans for future improvements of the model, see section 3.8 in the discussion after this chapter.

The polysucrose model was built by breaking down the sucrose structure into glucose and fructose components, and then combining the three glucose and three fructose molecules with available forcefield patches. While connecting two glucose or fructose residues, a -OH group of one monomer and the -H of the hydroxyl group of the other monomer had to be removed, in order to create a pair of monomers linked with the ester bond. So in each contact place, 2 hydrogen atoms and 1 oxygen atom were removed, i.e. the water molecules produced during the formation of glycosidic bonds.

Since the chosen polymer is short, its topology was described within a single RESIDUE. Parts of the topology file were prepared by Anna Skrzyszowska (under my supervision) during the undergraduate student practices. Topology fragments corresponding to glucose (ID:AGLC) and fructose (ID:BFRU) were arranged successively, according to their order in the initial model. The linkers between the fragments were added based on patches for building polysaccharide molecules (ID:SUCR and IMAL for the linkers inside sucrose molecules; ID: MELZ, 6KES for the linkers connecting

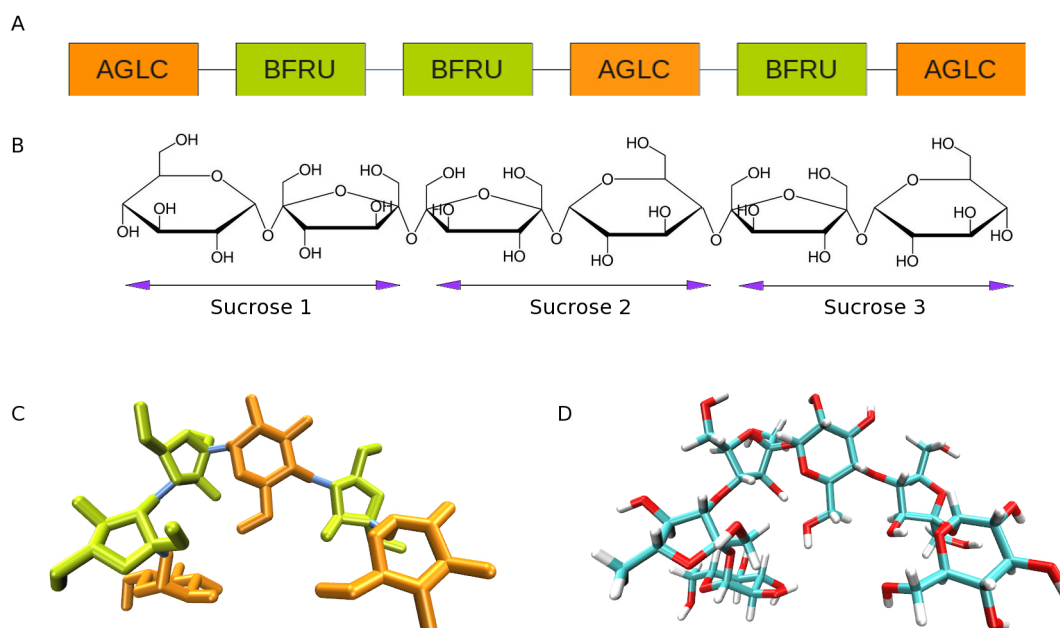


Figure 3.12: The structures of polysucrose (used as a representation of Ficoll) crowder, illustrating the order of glucose and fructose monomers (A and B), and the optimised conformation selected based on the MD simulation of a single polymer molecule in water (C and D).

neighbouring sucrose molecules). Partial charges and atom types, and thus also the force field parameters, were assigned entirely based on the fragments described in the CHARMM36 force field, i. e., no parameter modifications were made. The only interference in the parameterization of the molecule was the addition of three missing dihedral angles, with the parameters assigned by analogy to existing ones.

The last step in building the model was energy minimization and a test simulation of a single polysucrose molecule in the environment of explicit water with ions. Obtained trajectory was used to select an energetically optimal polysucrose conformation, which was used as a starting structure in the simulations of NS3/4A in a crowded environment. This method is similar to the way in which the starting conformations of PEG were obtained; however, due to the much lower mobility of polysucrose (Figure 3.13), only one conformation was used.

3.4.2 The concentration problem

Considering crowded molecular systems, two ways of understanding the crowder concentration can be distinguished: the traditional mass per volume or 'volume per volume' concentration, i.e. the volume of the crowder in relation to the total volume of the system or the volume of solvent. The latter is actually more intuitive in the case of such systems, as much of the impact of crowding is due to the excluded volume effect. For this reason, when constructing systems containing protein surrounded by crowdres, I took into account the concentration as a measure of the volume by which the crowdres limit the space available to the protease.

The size of the polysucrose molecules was chosen to correspond to the volume occupied by a single molecule of PEG in the semi-globular conformation. However, this resulted in a 'mass per volume' concentration mismatch between the systems crowded with PEG and polysucrose molecules. PEG molecule, with 206 atoms and a mass of 1295.5 Da is heavier than polysucrose with 129 atoms and a mass of 990.84 Da - similar volume at different masses is due to the difference in the structure of the two molecules - while PEG is a chain built with $-CH_2-CH_2-O-$ monomers, polysucrose contains six aromatic rings, significantly adding to its volume.

In the continuation of the project, new model of polysucrose was designed, with the mass corresponding to the mass of 28-mer of PEG, however this part was not included in the thesis as the research and model design are ongoing.

3.5 Dynamic properties of the crowdors

A first insight into the behaviour of crowder molecules was obtained by observing their internal dynamics and mobility. While coarse-grained crowdors maintain their semi-icosahedral shape with the RMSF of 0.1 Å, all-atom crowdors show much higher internal flexibility (Figure 3.13, Table 3.2). The greatest conformational variability is exhibited by PEGs, which are able to dynamically fold and unfold, adopting a wide range of conformations from fully stretched strands to compact globular structures. PEG conformations shift rapidly, with the transition from unfolded to globular states on the order of a few nanoseconds. Polysaccharide crowdors on the other hand are less mobile, with an RMSF of 2.1 Å - this can be partially explained by their structure rich in aromatic rings, which is not as flexible as the much simpler linear PEG chain.

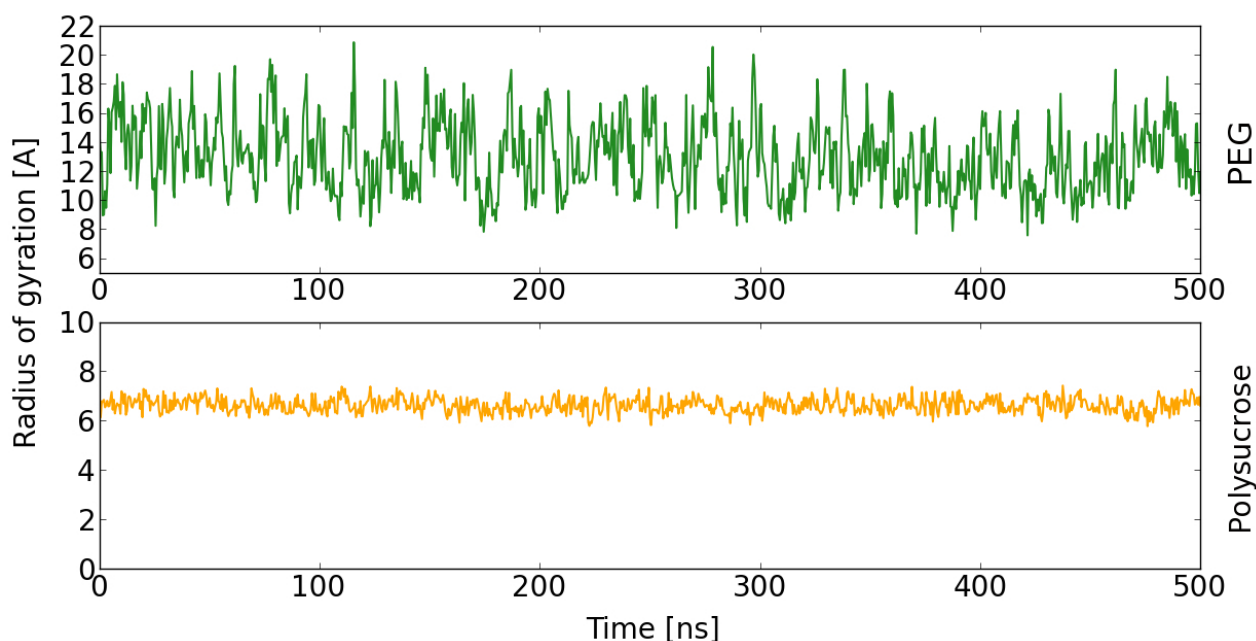


Figure 3.13: Internal dynamics of the crowdors visualised as changes in the radius of gyration of sample PEG and polysucrose molecules.

Environment	CG crowders	Polysucrose	PEG
RMSD [$\text{\AA}$]	0.42 ± 0.01	3.25 ± 0.82	9.30 ± 1.57
RMSF [$\text{\AA}$]	0.09 ± 0.01	2.12 ± 0.86	7.08 ± 1.36

Table 3.2: RMSD (eq. 2.13) and RMSF (eq. 2.14) calculated for heavy atoms of sample crowder molecules of each type. RMSD and RMSF values are averaged over three trajectories of the NS3/4A protease surrounded with given crowders, and presented with SEM and SD errors respectively.

One can notice a discrepancy between the actual radius of gyration measured based on trajectories of PEG molecules (in Figure 3.13) and the average hydrodynamic radius calculated by H. Lee et al.⁷⁵ ($7.7 \text{ \AA}$). However, when the values of the radius of gyration are considered, both results are in agreement; mean value of the radius of gyration of individual PEG crowder calculated from my trajectories equals $12 \pm 0.2 \text{ \AA}$, while the R_g value calculated for the same PEG molecule simulated by H. Lee et al. equals $11.8 \pm 0.2 \text{ \AA}$.

Another observed aspect of crowder dynamics was their translational diffusion, which was measured based on MSD values (eq. 2.18), taking into account the size of the simulation box and the effect of crowded environment on solvent viscosity (see Simulation Methods: Trajectory Analysis). Since conformational variability makes it impossible to accurately calculate rotational diffusion, this property of the crowders was not included. Below is a summary of both corrected translational diffusion coefficients ($D_{corrected}$, eq. 2.22) and raw values (D_0 , eq. 2.19) as calculated from the trajectories (Table 3.3).

Environment	Raw diffusion from MD ($\text{\AA}^2/\text{ns}$)	Corrected diffusion coefficient ($\text{\AA}^2/\text{ns}$)
PEG standard FF	3.89 ± 0.03	5.55 ± 0.01
PEG scaled FF	13.09 ± 0.19	8.74 ± 0.06
CG crowders	19.43 ± 0.44	11.17 ± 0.15
Polysucrose	25.41 ± 0.39	13.49 ± 0.13

Table 3.3: Translational diffusion coefficients of the crowder molecules of different types. The table contains both raw values of diffusion coefficients calculated from the MD trajectories (D_0 , eq. 2.19), and the coefficients corrected for PBC and TIP3P viscosity artifacts ($D_{corrected}$, eq. 2.22).

The D_t analysis offers a numerical proof that upscaling water-crowder interactions prevents aggregation of PEG and restores their realistic diffusion. An interesting feature of the corrections can be seen - in most cases they decrease the value of the coefficient, except for the case where crowder aggregation occurred, i.e., for the PEG crowdres simulated with standard force field parameters (Table 3.3).

As for the corrected values themselves, it can be seen that PEGs diffuse the slowest of all crowdres. This is certainly influenced by their conformational variability, and the resulting ability to restrict the movement of crowdres; however, it may also be indicative of interactions with both other crowdres and the protease. Coarse-grained crowdres diffuse faster than PEG simply because their interaction is reduced to repulsion and their spherical shape makes it easier for them to bypass each other. Both of these results are in line with what would be expected. In contrast, a somewhat surprising result is the diffusion coefficient of polysucrose, which moves even faster than the non-interacting CG particles.

Many studies suggest that polysucrose molecules are highly hydrophilic, so that they form a strong hydration layer around them¹¹⁷, preferentially binding with water rather than with any other partner. This property might serve as good explanation for the fast diffusion of polysucrose, since a layer of water around each crowder should facilitate diffusion. And indeed, analysis of the radial distribution function confirms increased density of water molecules directly around polysucrose, compared to the water surrounding PEG (Figure 3.14).

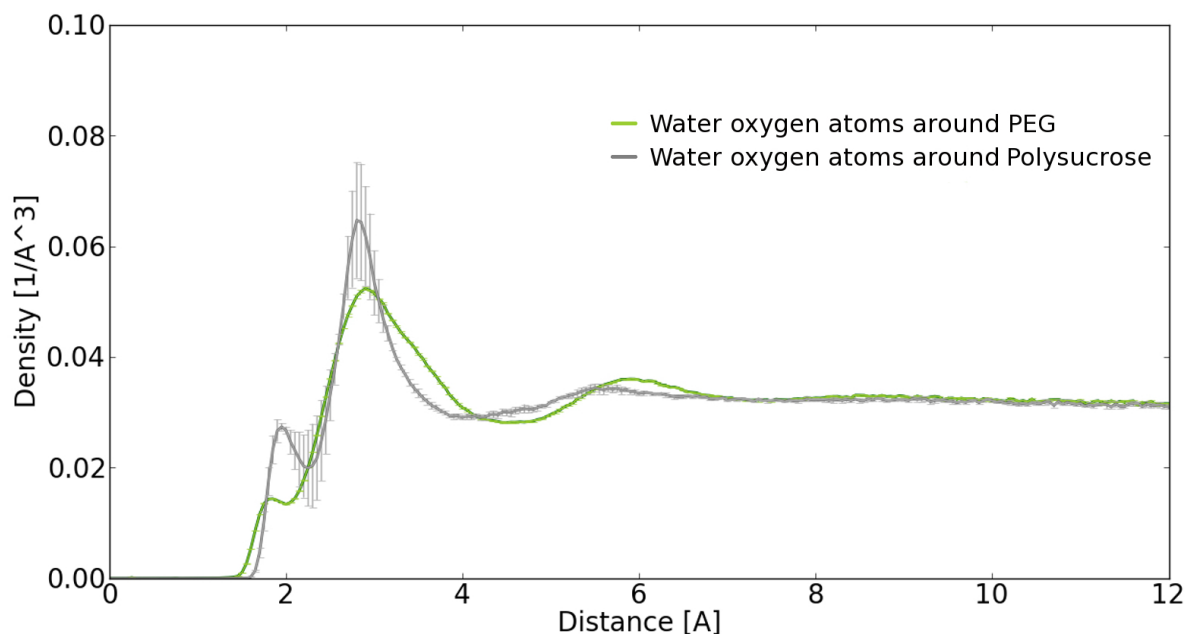


Figure 3.14: Distribution of water oxygen atoms around PEG and polysucrose molecules, calculated for water molecules surrounding all crowder atoms, including the hydrogens.

Typically, the RDF peak representing the first hydration shell - i.e. the water molecules directly interacting with the solute - appears at $\sim 2\text{-}3$ Å. The RDF graph presented here shows distances between atoms of the crowdres and the oxygen atoms of water molecules. For this reason, we see

the peak representing the first hydration shell divided into two distinct peaks at 2 Å and ~ 2.8 Å. The 'subpeaks' represent two positions of water molecules – positioned with oxygen towards the crowder (first, lower peak) and positioned with hydrogen closer to the crowder surface (second, higher peak).

Numerous hydroxyl groups of polysucrose are excellent donors of hydrogen bonds, and so it is understandable that its interactions with water molecules are particularly strong. However, even theoretically, it is hard to imagine that a molecule with such polar properties does not interact with proteins at all and acts as a completely inert crowder, as is commonly believed.

To find out if and how the polysucrose molecules interact with the protease and with each other, I calculated a set of radial distribution functions describing surroundings of the protein and crowdres, measured the contact times between the molecules and performed the analysis of hydrogen bonds forming in the trajectories. The results were compared with the interactions of PEG molecules and those of the coarse-grained crowdres.

Crowder-crowder contacts			
Crowder	$\tau_c^s[ns]$	$\tau_c^l[ns]$	S_c
PEG	0.45 ± 0.01	10.99 ± 0.46	0.74
CG	0.03 ± 0.01	3.88 ± 0.04	0.99
Polysucrose	0.36 ± 0.04	3.40 ± 0.51	0.71
Crowder-protein contacts			
Crowder	$\tau_c^s[ns]$	$\tau_c^l[ns]$	S_c
PEG	1.20 ± 0.26	29.78 ± 7.64	0.79
CG	0.43 ± 0.02	21.40 ± 0.41	0.93
Polysucrose	2.16 ± 0.10	41.13 ± 4.04	0.85

Table 3.4: Estimated times of crowder-crowder and protein-crowder contacts, measured on shorter and longer time scales, along with the weight associated with the shorter contact time (S_c).

Estimated contact times were calculated for both crowder-crowder and crowder-protein contacts, and averaged over three trajectories of each type of crowded systems containing the NS3/4A protease surrounded by 130 copies of crowder particles (Table 3.4). Contacts were calculated for two time scales, shorter and longer, by fitting correlation functions to a double exponential function (eq. 2.28), obtaining two contact times and the weights associated to the shorter contacts (see Chapter 2, section 2.2.7, equations 2.27 and 2.28).

As for the crowder-crowder contacts, a first observation is that the time of contact between pairs of mostly repulsive CG crowders goes to zero, with a weight of 0.99. Their longer contact, although having a weight of 0.01, is of similar length to the longer scale contact time (τ_c^l) of the polysucrose molecules. On this basis it can be assumed that a time on the longer scale of the order of 3-4 ns can be called a minimum contact time. At the other end of the scale are the PEG molecules, whose longer contacts reach lengths of more than 10 ns; thus PEG-PEG interactions are much more stable than polysucrose-polysucrose. The estimated contact times on the shorter scale (τ_c^s) are of a similar order for both all-atom crowders, although, as with longer scale contacts, they indicate greater stability in the contacts between the PEGs.

Environment	Simulation copy	Hydrogen bonds per frame	Mean value
PEG	sys1	3.46	4.35 ± 1.75
	sys2	2.76	
	sys3	6.76	
Polysucrose	sys1	10.56	10.67 ± 1.98
	sys2	8.30	
	sys3	13.14	

Table 3.5: Number of hydrogen bonds in each copy of the simulations of NS3/4A protease surrounded with PEG and polysucrose crowdres. Mean numbers of bonds per simulation frame are given with SEM errors.

Analysis of the protein-crowder contacts suggest that PEGs and polysucroses interact with the protein surface to a comparable extent, and even with an indication of longer, more stable contacts with polysucrose molecules. This result is not consistent with the common knowledge on Ficoll, according to which the polysaccharide displays only minimal interactions with the proteins, whereas it is the PEG that is considered to be more interacting⁴⁶.

Because of this inconsistency, I looked carefully at the parameters of the polysucrose molecules, thinking that it might be due to an error made during the parameterisation process. However, there is no indication that such an error could have occurred - all the LJ parameters and partial charge values were taken from a standard well-tested force field. The only place where new elements were introduced were the parameters of the three missing torsion angles, which were also copied from analogous linkers between saccharides from the CHARMM force field - so even if they were introduced incorrectly, they would only affect the geometry of the particles, not the mode or strength of interactions.

Residue	Hbonds fraction [% of all hydrogen bonds]		Contact fraction [% of the simulation time]	
	PEG	Polysu-crose	PEG	Polysu-crose
ARG	14.9	12.7	96	92
LYS	4.0	4.7	59	74
ASN	0.9	1.8	22	50
GLN	5.4	6.5	61	75
CYS	2.8	3.5	61	63
SER	8.1	6.5	90	83
THR	9.4	9.2	88	92
TYR	9.3	6.8	84	85
TRP	2.7	1.5	51	28
HIS	0.0	0.0	5	0
GLY	6.1	4.8	79	88
ALA	8.2	4.8	78	81
VAL	4.9	3.3	83	80
ILE	2.8	1.9	78	67
LEU	12.0	7.7	96	90
MET	1.2	1.8	47	71
PHE	1.5	1.9	57	52
ASP	1.0	7.4	44	70
GLU	1.9	9.0	46	66
PRO	2.8	4.7	82	87

Table 3.6: Hydrogen bonds and contacts between the NS3/4A amino acids and crowder molecules, calculated for single 500 ns trajectories of the systems containing the NS3/4A protease surrounded with PEG and polysucrose crowders. Residue types are ordered by their number of hydrogen bond donors. The Hbonds fraction columns show the fraction of unique hydrogen bonds between a given type of amino acid and PEG or polysucrose crowders, as a percent of all hydrogen bonds formed between NS3/4A and crowders. The maximum donor-acceptor distance was set to 3.5 Å and the donor-hydrogen-acceptor angle criteria were extended to no less than 120 degrees. The Contact fraction columns show the fraction of the simulation time when any crowder atom was present within 3.5 Å from the atoms of the amino acids of a given type.

The next step in the attempt to understand the obtained results was to search for hydrogen bonds between PEG or polysucrose molecules and amino acids on the protein surface (Table 3.6). The analysis confirmed the presence of such hydrogen bonds in the case of both crowdres; moreover, there are significantly more bonds with polysucrose molecules than with PEGs. Considering the maximal distance of 3.5 Å and maximal angle of 140 deg, there are over four protein-crowder hydrogen bonds per frame in the trajectories with PEGs, and about eleven with polysucrose crowdres (Table 3.5).

PEG and polysucrose are characterised by different hydrogen bond forming capacities. While PEG has a large number of acceptors and only two donors at each end, polysucrose has significantly more hydrogen bond donors (Figure 3.15). From this may come the main difference in what type of residues they interact with most frequently; it appears that polysucrose forms hydrogen bonds with acidic amino acids, namely ASP and GLU, far more frequently. The implications of this difference will be discussed further in the following chapter.

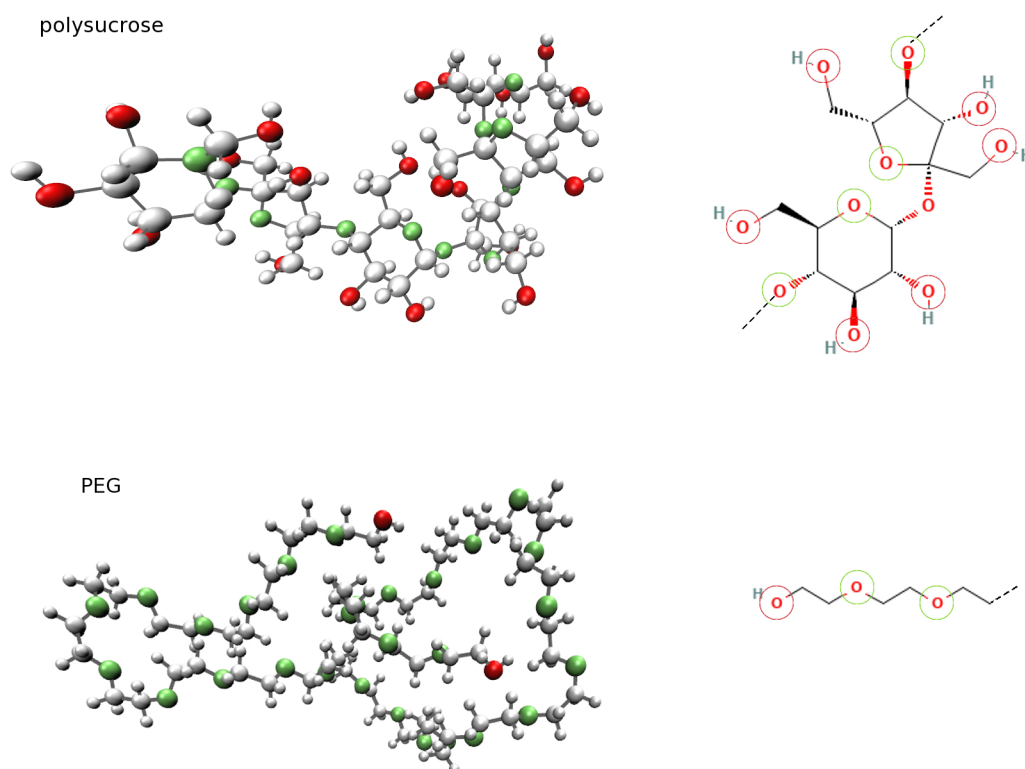


Figure 3.15: The structures of PEG and polysucrose crowdres with hydrogen bond donors marked in red and acceptors in green.

Other than that, it is hard to see other significant differences in the amino acid type preferred by PEG and polysucrose. However, looking from a different perspective makes it possible to notice the difference in their interactions elsewhere. Protein-crowder contacts mapped to the protein surface show that while PEG molecules accumulate around certain preferred regions of the protein, polysucrose distributes more evenly across the protein surface (Figure 3.16).

Differences in the distribution of crowdres on the surface of the protein may have downstream

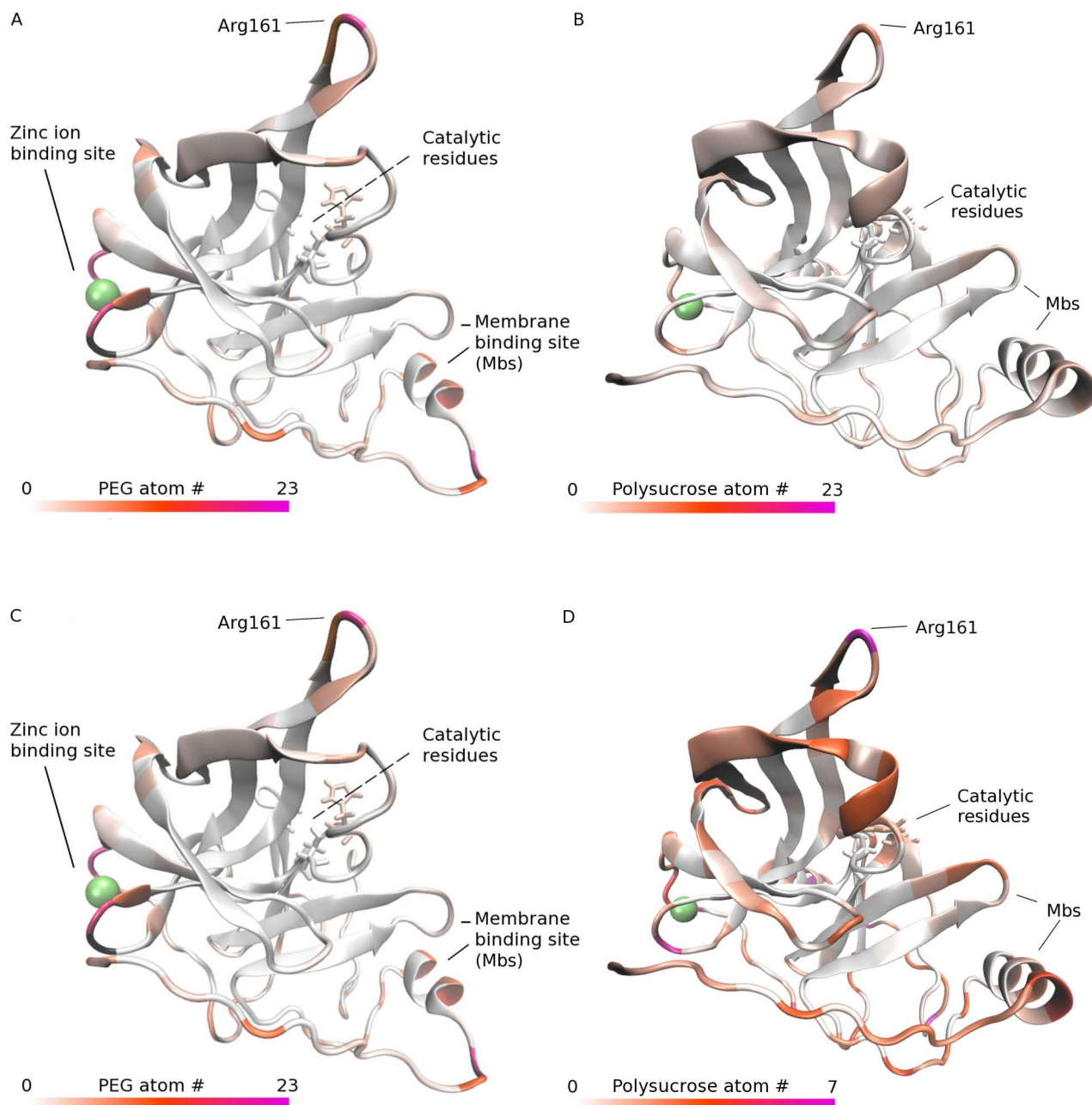


Figure 3.16: Distribution of PEG and polysucrose crowders around the NS3 protein, visualised as the ribbon model structures colored by the average number of crowder atoms around each amino acid per trajectory frame. Two pairs of structures, A-B and C-D, compare the distributions in a slightly different way: at the top, the structures are colored according to the same scale, to compare the amount of crowder atoms around the amino acids of each structure, while at the bottom, each structure is colored according to its own scale, to show different nature of the two distributions - PEGs gathered at specific regions of the protein and polysucrose molecules evenly distributed over the entire surface of the protease.

consequences for the rotational diffusion of the protease (see next chapter for exact diffusion analysis) - a transient cluster of protein and evenly distributed polysucrose molecules will have a more globular shape than a protein with PEGs accumulated at several locations, and thus in an environment crowded with polysucrose, rotational diffusion of the protein will be easier.

A final and concluding analysis of crowder interactions is the observation of the distribution of crowder atoms as a function of distance from the protein surface (Figure 3.17). It is easy to notice a clear peak in the graph for the PEG oxygen distribution, indicating the presence of a layer of crowder atoms directly at the protein surface. The irregular shape of the peak most probably corresponds to a wide range of conformations of PEG molecules directly contacting the protease.

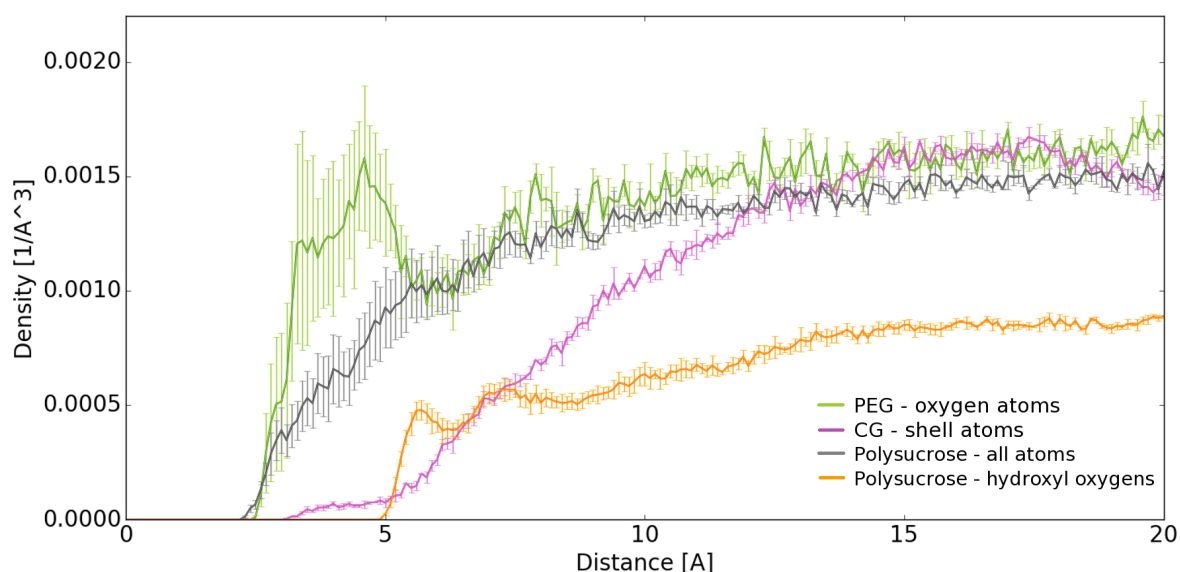


Figure 3.17: Distribution of PEG, CG and polysucrose crowdres around heavy atoms of the NS3 protease. The radial distribution functions have been normalised with respect to the number of atoms in distinct types of crowdres and to their volume fraction. On the graph, two lines are shown to characterise polysucrose crowdres distribution, to highlight the difference between distribution of all crowder atoms, and the distribution of hydroxyl groups interacting with the protein. In case of PEG crowdres, oxygen atoms for which the RDF was calculated are distributed evenly along the polymer chain, therefore only one line is shown, as oxygen locations are a good approximation of PEG shape.

Another observation that fits with previous results is the shape of the graph for the distribution of coarse-grained crowdres - these crowdres distribute around the protein in a random manner and do not form a layer of molecules interacting with its surface. This result is a clear confirmation that the CG crowdres do not interact with the protein, and they behave exactly as designed.

Once again, trajectories with polysucrose proved to be the most complicated to analyse. Calculating the RDF for the distribution of all polysucrose atoms around the protein yields a plot that is a textbook example of a graph for non-interacting molecules - the number of polysucrose atoms increases continuously, and no defined layer of polysucrose atoms is visible. Such a result would, of course, be inconsistent with previous analyses that indicate multiple interactions with protein

residues. It is due to the specific structure of the polysaccharide; each semi-globular molecule has twenty-five hydroxyl groups evenly distributed on its surface. Measurement of the RDF for the distribution of oxygen atoms of these groups clearly indicates that they form layers around the protein surface, partially resulting from the presence of numerous hydrogen bonds with the protease residues (Figure 3.17).

The peaks observed for PEG and polysucrose crowders both refer to the distribution of oxygen atoms, but the location of these atoms in the molecules is different, hence the difference in peak location. The oxygens of PEG molecules are composed into linear polymer chain, and located directly at the surface of the protein, acting as hydrogen bond acceptor; whereas analysed oxygens of polysucrose are components of the hydroxyl groups being hydrogen bond donors, and therefore they are separated from the protein surface by a bond and a hydrogen atom.

3.6 Discussion

3.6.1 Non-standard crowder choice

In the laboratory experiments investigating crowding, technical aspects make non-protein polymers, such as PEG, Ficoll or dextran the most common choice of crowded environment representation^{46,6}. At the same time, in the simulations studying proteins in crowded environment, there are two main directions for choosing the crowders. The first one is using reduced models, providing only the volume exclusion. The second one is deciding on crowders representing molecules actually present in living cells, providing interactions with the studied protein as well¹¹⁸. This is understandable, as if like in the case of molecular modelling and simulations, the choice of crowders is virtually unlimited, the most desired form of crowded environment model would be a realistic depiction of a fragment of cytoplasm. In this way, a kind of division is created, into 'laboratory crowders' and 'simulation crowders'. The results obtained in my simulations suggest, that there might be a gap in understanding of how exactly 'laboratory crowders' affect the proteins, which is partly related to said division. A clear example of this gap is the case of polysucrose that represent Ficoll molecules. Ficoll is generally considered to be inert and not interacting crowder^{113,6}, while in the simulations it forms more contacts and hydrogen bonds with the protease surface than PEG, which is known to form such interactions^{100–104}.

The idea of introducing 'laboratory crowders' into molecular dynamics simulations is very uncommon. There are only a few very recent computational studies employing PEG as crowder molecules - simulations of Lysozyme surrounded with all-atom PEG polymers^{119,120}, and a study on intrinsically disordered proteins surrounded with PEG crowders in a coarse-grained representation¹²¹. With Ficoll and polysucrose crowders the case is simpler - to my knowledge there are no molecular dynamics simulations published to date, that would study the impact of polysucrose crowders on a protein. There is one publication that describes a MD simulation system containing both multiple molecules of PEG and polysucrose - a study on the effects of crowding on the hydrogen

bond network of water¹²², however no details are given on the design of the models themselves.

In my work, PEG and polysucrose were chosen as crowders to replicate as closely as possible the system used in laboratory experiment investigating the effect of crowding on the activity of NS3/4A⁹¹. The assumption at the beginning of the project was that being able to look at the same solution from two perspectives was likely to create new possibilities for interpreting the results - hence the simulations with PEG polymers. Later, non-obvious experimental results confirmed that this was the right direction, and further simulations were carried out with another 'laboratory crowder' (polysucrose), this time specifically to explain the obtained results.

The choice of non-standard crowders, however, had its consequences in that there was no opportunity to pick up solutions to technical problems in the literature - at the time the simulation systems were developed, no publications described such crowders, only the polymers in other settings.

There are many simulations of PEG in the context of PEGylation, i.e. attaching PEG chains to proteins, for example in order to facilitate transport of drug molecules into the target cells. An extensive review on computational studies of PEGylation was published by S. Sousa et al.¹²³. The work covers simulation studying properties of single PEG molecules, including the work I used while designing and generating my simulation systems⁷⁵, followed with many simulations of PEG conjugates with small molecules, peptides and nanocarriers like micelles and carbon nanotubes. Ficoll on the other hand was simulated by Y. Mao et al.¹²⁴ in a work studying the use of Ficoll as cryoprotective agent - a single polysucrose molecule representing Ficoll 70 was modelled as linear polymer of sucrose molecules.

3.6.2 Ficoll or polysucrose?

In the assay performed to compare activity of the NS3/4A protease in water and in crowded environment, three types of crowder molecules were used: PEG 600, PEG 6000 and Ficoll 400. The numbers in product names indicate the mean molecular mass of each crowder - 600, 6000 and 400 Da. However, the masses of crowders used in the molecular dynamics simulations were different - I used PEG and polysucrose of the mass of 1295.5 Da and 990.84 Da respectively. The choice of the PEG molecule was motivated by the availability of hydrodynamic radius data for the specific polymer, while the polysucrose was created to match the volume of the coiled PEG molecule and the CG crowder.

However, mass is not the only element that differs between the polysucrose used in the simulations and Ficoll. Under the name Ficoll, which is a commercial product protected by trademark, hides a polymer formed by the reaction of sucrose with epichlorhydrin; this produces polymers of various masses and compositions, which are known to be highly branched (Figure 3.18) and form nearly spherical structures in water^{125,126,116}. In such polymers, sucrose molecules are connected by glycerol linkers (Figure 3.18), however there is no available data on the probability of attaching glycerol to given sucrose carbons, or on the degree of branching, that is on the average number of glycerol linkers per sucrose monomer.

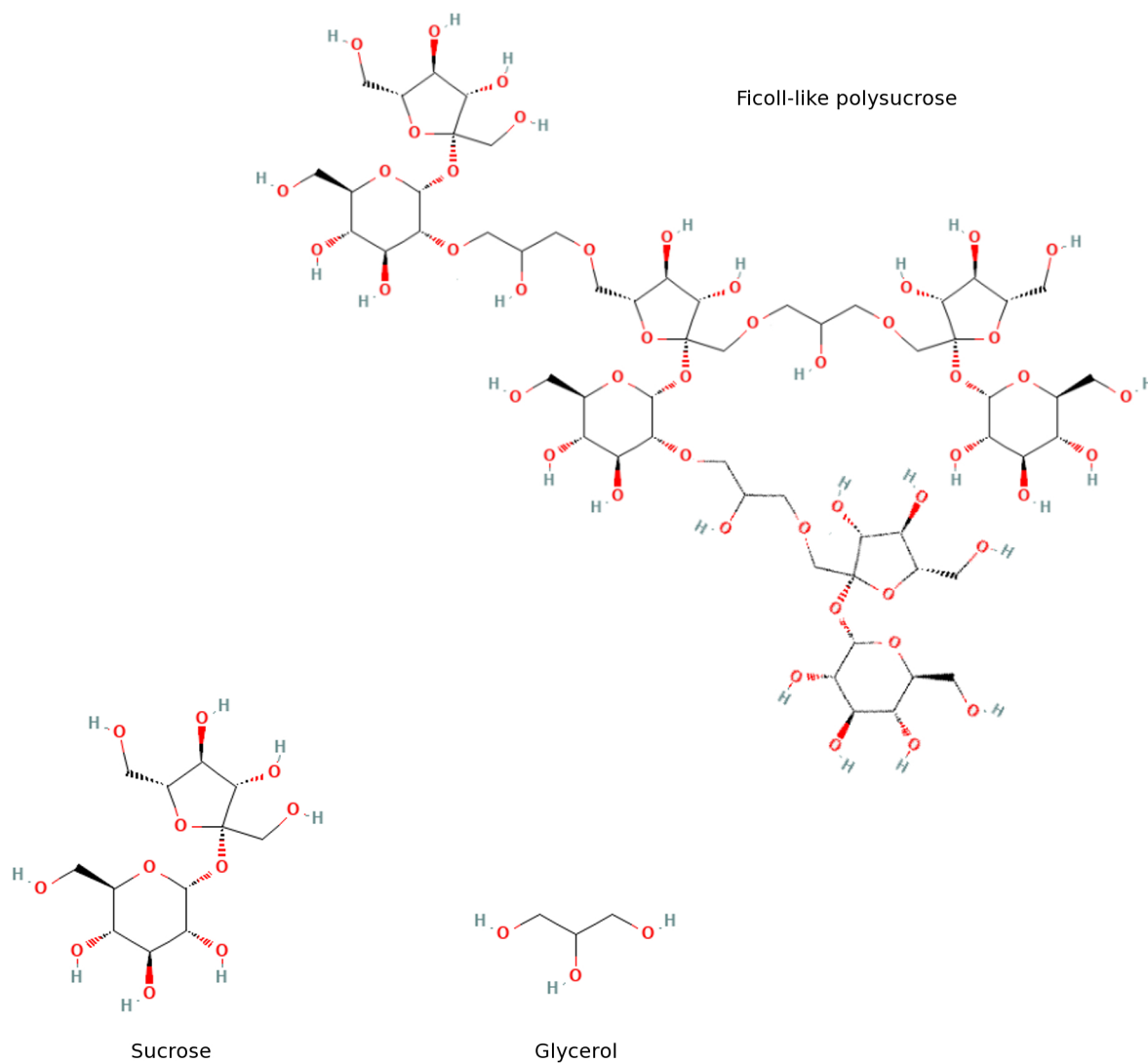


Figure 3.18: A model of Ficoll polymer composed of four sucrose molecules with three glycerol linkers. It is worth noting, that cutting off any sucrose molecule (and thus reducing the molecule to a trimer) would make the polymer linear. Figures prepared using structures from the PubChem database¹²⁷ (sucrose ID:5988, glycerol ID:753).

In my simulations, the priority was to capture how the presence of Ficoll affects the dynamics of the NS3/4A protease. Thus, the most critical element in designing crowders representing Ficoll was to reproduce how polysucrose molecules interact with the protein surface, and to achieve a similar volume fraction as in the other simulations, with which the results were compared.

Given the availability of molecules and linkers already parameterised in the CHARMM force field^{114,73,74}, I decided that it would be sufficient to choose a linear polymer consisting of three sucrose molecules to represent protein-crowder interactions. It is worth noting, that the glycerol linkers are built with very similar groups to those found in sucrose, i.e. carbons with hydroxyl groups. Therefore, there is no reason to suppose that their presence would bring new types of interactions with protein residues. Due to the fact that the polymer used has only three monomers, it is also difficult to introduce any degree of branching. As my crowders are much shorter than Ficoll 70 or 400, they are unbranched and lack glycerol, I resigned from referring to them as Ficoll, and used the name polysucrose instead.

The model used in the thesis is not a final attempt to investigate the mechanism of Ficoll's effect on NS3/4A protease dynamics. In a new project on which work is currently underway, the model of this crowder is being improved in several ways, to more acutely analyse the dynamics of Ficoll molecules and their effects on the protein.

3.6.3 The coarse-grained crowders

The main motivation for choosing the bead-shell crowder model was the desire to represent both protein and water with atomistic detail. All stages of the model design were driven by technical considerations and the need to adapt the model elements for an otherwise all-atom simulation in NAMD:

- I resigned from the most commonly used representation, i.e. a single pseudo-atom, because it would require increasing the cutoff for nonbonded interactions from the standard 12 to 19 Å or more (at a cutoff of 12 Å, crowders with a radius of about 8 Å would start overlapping before they felt any interaction from a neighbouring crowder),
- I have chosen the radius of the pseudo atoms corresponding to the radius of the carbon atom, in order to keep a 'realistic' shape of the LJ potential for interactions with water and protein, corresponding to the potential of other atoms in the simulation,
- I have replaced bonds between crowder atoms with a strong interaction described by LJ potential, because in the case of a model with bonds NAMD returned errors suggesting that pairs of atoms within bond exclusion list (separated with less than four bonds) located in non neighbouring patches. Patches in NAMD are groups of atoms for which the interactions are calculated on a given processor, and the construction of program code requires atom within bond exclusion to be located within neighbouring patches. This suggests that I could have

continued using the model with bonds in two cases: either in a simulation performed on a single processor, or with the bond exclusion disabled and thus nonbonded interactions calculated for every pair of atoms within the cutoff, including the pairs directly connected with bonds. In such case, bond exclusion would be disabled also within the protein and water molecules, substantially increasing the simulation time. In this sense, resignation from using bonds is a simulation performance optimization.

Resulting crowder model is perfectly fitted to be included in all-atom systems, and I consider this its biggest advantage. Using this model in the simulation allows to retain the dynamics and interactions with individual water molecules, and their potential influence on studied macromolecule. At the same time, it reduces the simulation time compared to a simulation with all-atom crowdors, and even compared to the system containing only the protein in explicit water. Since the bead-shell crowder contains less atoms than the same volume of explicit water, a simulation with them is less computationally expensive and according to this benchmark allows a 20% reduction in simulation time compared to a simulation without crowdors (Figure 3.19).

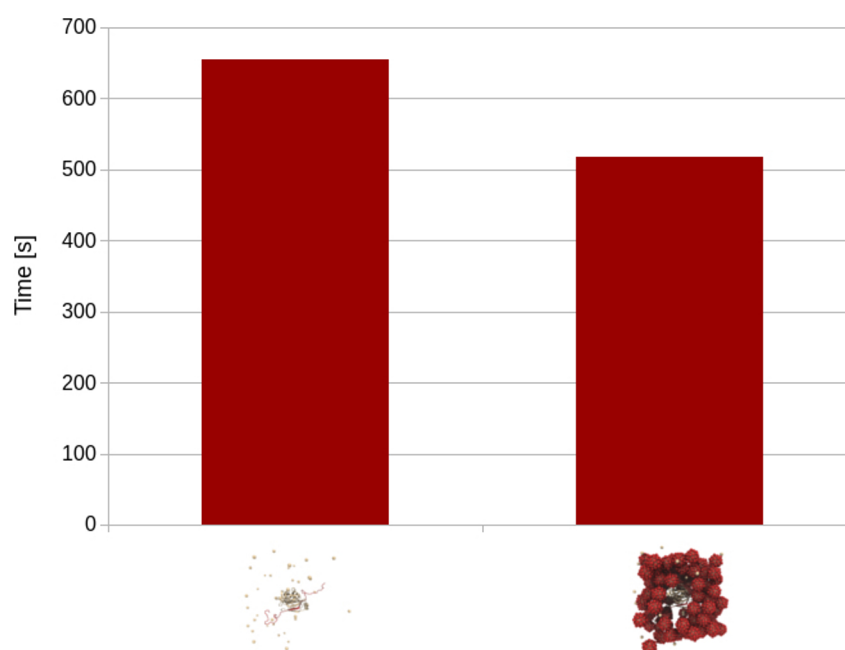


Figure 3.19: Benchmark times for 10 000 steps of molecular dynamics simulation of the NS3/4A protease in explicit water with ions (left) and in explicit water with ions and 130 copies of the bead-shell crowdors. Test simulations were performed for the system with exactly the same volume of the simulation box.

Of course, such crowdors can only be used for simulations in which only volume exclusion is needed. There are many examples of simulations where researchers restrict any interaction of crowdors with the protein, even excluding the attraction part of the Lennard Jones potential^{27,22}. Spherical crowdors themselves, in the most often used form of single spheres representing whole crowder molecules, which interact only via LJ potential, have been used in simulations studying the

effect of crowding on aspects of protein dynamics such as diffusion, domain mobility, stability of proteins and the formation of protein complexes^{27,19,21,128,22,129,110,130}.

Simulations with reduced crowders typically also use some level of coarse-graining in representing the protein of interest, which leads to the loss of some information about the internal dynamics of the protein and its interaction with binding partners. This is where the bead-shell model described in this work can apply; by allowing multi-scale simulations, with an all-atom representation of the system and a reduced crowder representation, it allows for the acquisition of full information about the dynamics of the protein under study.

Possible modification of the model include changing the size of the pseudo-molecule, the strength of LJ potential of interactions with the protein, water or other crowder molecules and adding partial charges on the pseudo-atoms of the crowder. By adding charges, the way the crowders interact with other molecules can be altered further, introducing more complex effects on the protein. Charges can either be added directly on the surface pseudo-atoms, or on the central pseudo-atom(s). The central atom can be for example replaced with four smaller pseudo-atoms, with two of them charged positively and two negatively. This operation would create patches of low-magnitude positive and negative charges on the surface of the crowder, making the crowder more similar to realistic molecules.

3.6.4 Comparison of dynamic properties of the crowders

The effect of a crowded environment on protein activity and dynamics is the result of two main phenomena: nonspecific interactions of crowder molecules with protein residues and volume exclusion, i.e. the limitation of the space available to the protein. In the simulations described in this thesis, the crowder models used to investigate the effect of crowding on NS4A protease dynamics were selected to disentangle the outcomes of these two effects. Thus, the assumption was that the CG crowders would only provide excluded volume effect, and the all-atom PEG molecules would represent the crowders interacting with the protein. Since, according to the available literature, Ficoll is an inert, non-interacting crowder^{131,132,133,134}, the polysucrose molecules representing Ficoll were initially assumed to have intermediate properties - be an all-atomic crowder, but showing minimal interactions with protease.

The first notable observation from the simulations of NS3/4A protease in the crowder surroundings was that the results obtained from the trajectory analysis were consistent with most, but not all of the assumptions:

- PEG crowders were found to form numerous contacts with both hydrophobic and polar residues of the NS3/4A protease, and with some of the contacts resulting in hydrogen bond formation (see Tables 3.6 and 3.5)
- As intended, the coarse-grained crowders did not interact with the surface of the protease. This was confirmed by analysis of the crowder distribution as a function of distance from the

protease (RDF), according to which the CG crowders did not form layers around the protein (see Figure 3.17). Additional confirmation is the lack of observed long-scale contacts with the NS3/4A residues (Table 3.4)

- Polysucrose molecules, contrary to literature data, evidently showed interactions with the protease. They engage in many short- and long-scale contacts with the protease surface, with average contact times in both cases longer than for the protein-PEG contacts. They also form much more hydrogen bonds with NS3/4A amino acids - on average 11 hydrogen bonds per simulation frame, compared with 4 bonds formed with the PEG crowders. (see Tables 3.4, 3.6 and 3.5)

Trajectory analyses confirmed that although different than expected, there is nevertheless a fundamental distinction between crowders interacting and not interacting with the protein, with both all-atom crowders in the first group (Figure 3.20). The assumed separation of the observed effects on protease dynamics into those due to volume exclusion and protein-crowder interactions should therefore still be feasible.

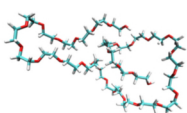
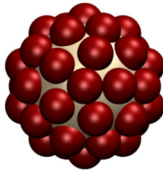
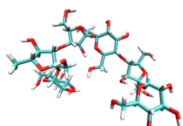
Crowder			
Protein - crowder interactions	✓	✗	✓
Crowder - crowder contacts (long scale)	✓	✗	✗
Corrected diffusion coefficient [$\text{\AA}^2/\text{ns}$]	8.74	11.17	13.49
RMSF [$\text{\AA}$]	7.08	0.09	2.12

Figure 3.20: A summary of dynamic properties of crowders - all-atom PEG (left), polysucrose (right) and the coarse-grained crowders (center).

In addition to the degree of interaction with the protease, two major differences were observed between the crowder models: the level of flexibility and differences in diffusion rates (Figure 3.20).

As expected, the structure of the CG crowders remained almost rigid, while PEG, a polymer with a chain structure ($-\text{CH}_2 - \text{CH}_2 - \text{O}-$) showed very high mobility. PEG crowders adopted a wide range of conformations, from fully extended, to compact, folded structures. In terms of internal dynamics, the polysucrose molecules were closer to CG crowders than to PEGs; with an RMSF of about 2 $\text{\AA}$, it adopted a constant small range of conformations, with the hydrocarbon rings arranged in a streamlined shape (see Table 3.2 and Figure 3.13).

For the translational diffusion of the crowders, some of the most unexpected results were obtained - it turned out that polysucrose crowders are characterised by very fast diffusion, with a translational diffusion coefficient of about $13.5 \text{ \AA}^2/ns$. This value is higher not only than the diffusion coefficient of PEG polymers, but also that of spherical bead-shell crowders (Table 3.3). While PEG crowders are slowed down by interactions with proteins and the formation of transient clusters between each other, as well as by the formation of networks of variously arranged chains, saccharide molecules and CG crowders have a semi-globular shape and lack long crowder-crowder contacts (Table 3.4).

The question of why then polysucrose diffuses at a higher rate than CG crowders is not trivial to answer. With the data collected so far, I can only suppose, that the reason for this phenomenon might be related to repulsive charge-charge interactions occurring upon crowder-crowder collisions. The surface of polysucrose, as well as 'real-life' Ficoll molecules is dotted with polarized hydroxyl groups (Figure 3.15), forming charge-charge interactions with water molecules. As a result, polysucrose crowders that are not in contact with the surface of NS3/4A, diffuse as spherical complexes of a saccharide and its hydration shell. Depending on the orientation of the water molecules and their mobility, there might be for example charge patches transiently forming on the surface of such crowder-water complexes. The properties such as distribution, orientation and mobility of molecules in the hydrodynamic layer of polysucrose can be measured, and it will be analysed in the trajectories obtained with the improved model of polysucrose crowders, better reflecting the properties of Ficoll.

Chapter 4

Effects of crowders on the NS3/4A protease

4.1 Impact of crowding on the activity of the NS3/4A protease

Central to the entire project, including the selection of the crowder model, was the idea of transferring experimental conditions to molecular dynamics simulations so as to combine observations of the system at two scales - the activity of multiple enzymes in a test tube and the dynamics of a single protein molecule. This approach is likely to allow a deeper understanding of the mechanism of enzymatic activity of the protease under study and how it is affected by the crowded environment.

In both laboratory experiments and simulations, the NS3/4A protease was immersed in water, and surrounded by the non-protein polymers (PEG and polysucrose), whose main task is to limit the space available to the protein, while minimising protein-crowder interactions. The activity assay measured the effect of crowding on the rate of the enzymatic reaction catalysed by the protease, while simulations allowed observation of the processes that directly determine protein activity - its internal dynamics, and the diffusion of both the enzyme and its substrates.

4.1.1 Experimental measurement of NS3/4A protease activity - the general idea

Determining the activity of a protein requires measuring how fast the enzymatic reaction catalysed by it progresses. This can be achieved by directly or indirectly measuring the decreasing concentration of substrates or increasing concentration of products of the reaction. Doing so we measure the reaction rate - by definition, it is defined as the rate of product formation: $v = \frac{d[P]}{dt}$, where [P] is the concentration of the reaction product.

In the NS3/4A activity assay, this concentration was measured indirectly by measuring the fluorescence intensity of the product (Figure 4.1). The peptide used as a substrate in the assay has two fluorescent tags - EDANS and DABCYL on either side of the cleavage site. When the uncleaved peptide absorbs light, the excitation energy of EDANS, instead of being emitted, is transferred via FRET to a second tag that acts as a quencher. This mode of energy transfer is closely related to the distance between tags - when the peptide is cleaved by the protease, the distance between the

tags increases, and with it less and less energy is transferred to the quencher and more and more is emitted as fluorescence. Thus, while measuring changes in EDANS fluorescence, the progress of the enzymatic reaction is also quantified.

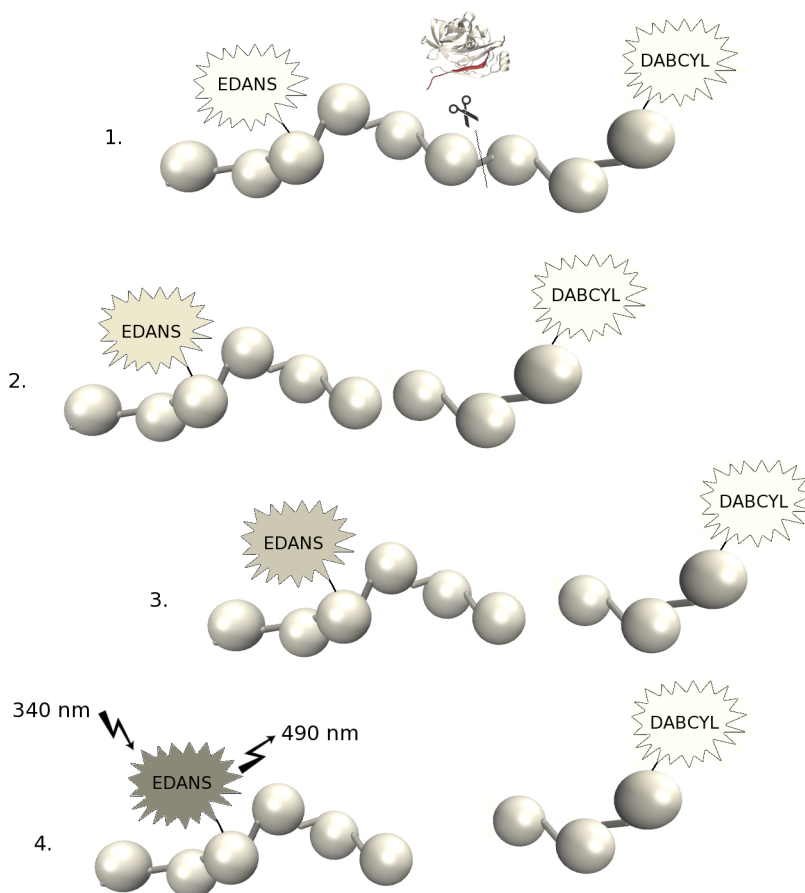


Figure 4.1: The schematic illustration of the assay used to measure the activity of the NS3/4A protease. Residues of the substrate are represented as spheres located at the C_α atom positions, the FRET donor and quencher are represented with starburst callouts. After enzymatic cleavage performed by NS3/4A (1.), the intensity of EDANS fluorescence increases with the distance between EDANS and DABCYL tags (2. – 4.).

To quantitatively compare the activity of the enzyme in different environments, initial reaction velocities (in relative fluorescence units per second) versus substrate concentrations were fitted to Michaelis-Menten model, in which the progress of the reaction is described in terms of the kinetic constants:

$$v = \frac{d[P]}{dt} = V_{max} \frac{[S]}{K_M + [S]} = k_{cat}[E]_0 \frac{[S]}{K_M + [S]} \quad (4.1)$$

where $[S]$ is the concentration of the substrate, V_{max} is the maximum rate achieved by the reaction, K_M is the Michaelis constant, which equals to the substrate concentration in which the reaction rate is equal to $0.5V_{max}$, and k_{cat} is the catalytic rate constant and $[E]_0$ is the initial concentration of the enzyme.

4.1.2 Activity assay results

The experimental assay was performed by Monika Wojciechowska and Agnieszka Popielec. I helped with the synthesis of the peptide substrate. The results are briefly presented here because of their key importance for the interpretation of other results presented in the thesis.

The activity of NS3/4A was measured in the presence of three crowders - PEG 600, PEG 6000 and Ficoll 400. Three concentrations of each type of crowder molecules were considered: 50, 100 and 200 g/l. For each group of experiments, control measurements were performed in buffer without the crowders. Obtained results are presented as raw graphs of initial reaction velocities versus substrate concentration (Figure 4.2), and as calculated kinetic parameters (Table 4.1).

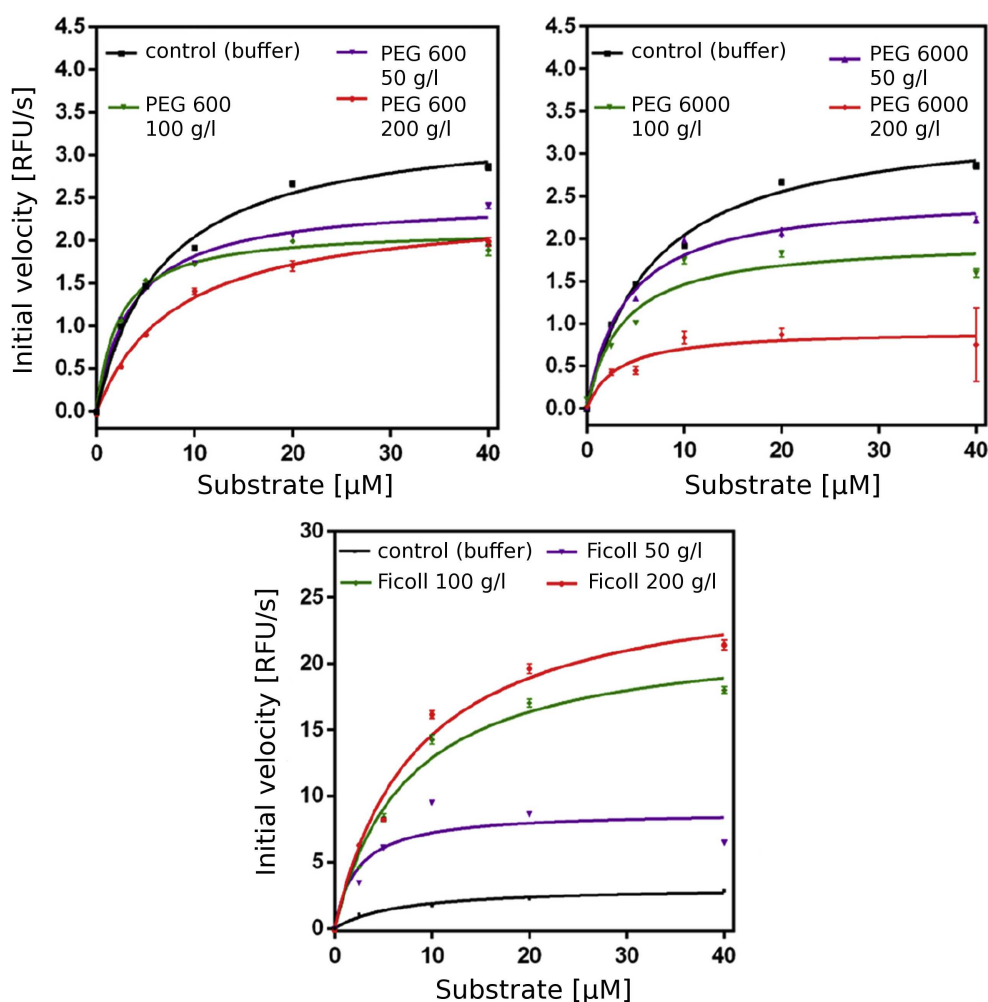


Figure 4.2: Michaelis-Menten graphs illustrating the effects of crowding on the activity of the NS3/4A protease. Initial velocities of the reaction in each crowder concentration are averaged over three measurements with standard deviations (SD) errors. V_0 parameters of the peptide substrate cleavage were measured in the buffer and in solutions crowded with PEG 600, PEG 6000 and Ficoll 400. Figure from a published work⁹¹.

media	concentration [g/l]	$K_M \pm \text{SD}$ [μM]	$V_{max} \pm \text{SD}$ [$\mu\text{M/s}$]	k_{cat} [s^{-1}]
Buffer	0	6.7 ± 0.4	3.4 ± 0.1	1700
PEG 600	50	3.7 ± 0.2	2.5 ± 0.1	1250
	100	2.3 ± 0.2	2.1 ± 0.1	1050
	200	8.2 ± 0.4	2.4 ± 0.1	1200
PEG 6000	50	4.0 ± 0.3	2.5 ± 0.1	1250
	100	3.6 ± 0.9	2.0 ± 0.1	1000
	200	3.1 ± 1.6	0.9 ± 0.1	450
Ficoll 400	50	2.3 ± 0.4	8.9 ± 0.8	4450
	100	7.4 ± 1.0	22.4 ± 0.9	11200
	200	8.4 ± 1.0	26.9 ± 1.1	13450

Table 4.1: Kinetic parameters for the hydrolysis reaction catalysed by HCV NS3/4A protease. K_M , V_{max} and k_{cat} parameters of peptide substrate cleavage were measured in the buffer and in solutions crowded with PEG 600, PEG 6000 and Ficoll 400. Table based on the published work⁹¹.

Since in a highly crowded environment the diffusion of molecules is impeded by steric effects and crowder interactions, the rate of enzymatic reactions can be expected to be lower with the higher the concentration of crowdres. Meanwhile, although the dependence of the reaction rate on crowder concentration was demonstrated for all crowdres, the direction of these dependencies was quite the opposite.

Both PEG 600 and 6000, for which previous studies have shown some degree of interaction with proteins, indeed slowed the reaction in proportion to the concentration of crowdres. However, the presence of Ficoll, which is widely regarded as the 'ideal' inert crowder, resulted in increased rate of enzymatic reaction catalysed by NS3/4A. Moreover, with increasing concentrations of Ficoll, the rate increased even more (Table 4.1).

The observed effect is definitely unexpected, and also non-trivial to understand. There are two main directions to searching for possible explanations: the first one is a possibility that the enzyme conformation changes upon crowding, in a way that would improve or impair protease-peptide binding or the cleavage reaction; the second one is quite possibly altered diffusion of the enzyme or substrates - this includes translational movements, rotations, formation of transient clusters or aggregates. Most of the resulting questions can be answered far more easily by running and analysing molecular dynamics simulations than based on laboratory experiments.

4.1.3 Molecular dynamics simulations of the NS3/4A protease in water and in crowded environment

To study the effects of crowding on the dynamics of NS3/4A protease, two main groups of molecular dynamics simulations were conducted:

- the first group studied **the NS3/4A protease complex in four different surroundings**: control simulations in explicit water with ions, and explicit water with three types of crowders: all-atom PEG molecules, all-atom polysucrose molecules representing Ficoll, and the coarse-grained crowders. These simulations aimed to determine how the crowded environment affects internal dynamics of components of the NS3/4A protein and its diffusion. Using different types of crowders allowed to disentangle the effects of volume exclusion and nonspecific protein-crowders interactions,
- the second group contained **systems including peptide substrates**, of the sequence representing the substrate used in the activity assay. The aim of this group of simulations was to investigate the mechanism of crowding effects on enzyme activity, by taking a closer look at the distribution of crowder molecules around the enzyme's active site, as well as the diffusion of peptide substrates and the protease.

First of all, to dispel any concerns about the impact of PEG and polysucrose on the active site of the protease, obtained trajectories did not show significant effects of crowders on either the conformation of this region or its flexibility. The active site is one of the most stable regions of the NS3 protease, with the RMSF of heavy atoms within 5 Å from the catalytic triad (His57, Asp81, Ser139) of 0.9 Å, both in water and in the presence of polysucrose and PEG. The only flexible part of this region is His57, which often flips out of the 'active' conformation, orienting towards the solvent. The conformation which I called active is the one in which the distance between all three catalytic amino acids is minimised, and the one that have been captured in the crystallographic structure. The frequency of 'His flipping' is increased in crowding (61% of the observed 500 ns in water and over 70 % in the presence of all-atom crowders). It is not clear whether the flipping has a significant effect on the substrate binding process, or does the NS3's amino acids align correctly with the substrate upon binding regardless of their initial conformation; however in my thesis I have not studied the process of substrate accommodation in the active site.

The differences between the distribution of PEG and polysucrose around the protein surface have been described in the previous chapter, with a much more even distribution of saccharide molecules. However, in the vicinity of the active site in both cases the distribution has a similar, very interesting property: both PEG and polysucrose leave a crowder-free zone around the catalytic triad, with the radius of over 7 Å in case of PEG, and over 6 Å in case of polysucrose.

Given these observations, it can be said with a high degree of confidence that the crowders do not interfere with the peptide bond cleavage reaction itself - they do not interact with the catalytic amino acids and do not block substrate access directly at the active site of the protease.

4.1.4 Crowding affects rotations and translational diffusion of the protease

Crowders do not block the active site of the protease itself, but have the potential to significantly alter the diffusion of both the protease and its substrates:

- By occupying a significant amount of space around the protease, they reduce the range of possible translational movements - this effect has the potential to considerably reduce the frequency of contacts of the protease with its substrates
- Interactions with crowder molecules and formation of transient protein-crowder clusters can disrupt the rotational diffusion of the protein
- Diffusion of the crowders themselves affects the movement of water around the protein, indirectly affecting also the diffusion of other molecules, in particular smaller, such as the peptide substrates

To begin with, I present the results regarding the translational diffusion of the protein in water and in different types of crowded environment. As the diffusion has been calculated from trajectories obtained from simulations in a relatively small box with periodic boundary conditions, the obtained values have been corrected for PBC effects. Additionally, the water viscosity in the TIP3P model, differing from the experimentally measured water viscosity has also been taken into account (see Methods, equations 2.20 and 2.21). The raw and corrected values are shown in Table 4.2.

Environment	Raw diffusion coefficient from MD, D_0 [$\text{\AA}^2/ns$]	Corrected diffusion coefficient, $D_{corrected}$ [$\text{\AA}^2/ns$]
Water	3.14 ± 0.04	6.98 ± 0.04
PEG scaled	1.98 ± 0.03	4.66 ± 0.01
Polysucrose	2.55 ± 0.16	5.18 ± 0.05
CG crowders	2.47 ± 0.20	4.85 ± 0.05

Table 4.2: Translational diffusion coefficients of the NS3/4A calculated from MSD of the centre of mass of the NS3 protease in water and in crowded environment. Table compares both raw values calculated from the MD trajectories and values corrected for PBC and viscosity artifacts.

For translational diffusion of proteins, the more diverse and complex shape of PEG molecules may be relevant. They have a simpler and more flexible structure than polysucrose, with the ability to rapidly fold into compact globular structures and unfold to completely stretched conformations. Such properties form this type of crowder into a network that makes it very difficult for the protein to traverse - hence the lowest translational diffusion coefficient.

Polysucrose crowdres are characterised with denser hydration shell around them (Figure 3.14), they also almost lack long time crowder-crowder contacts (Table 3.4), as their predicted long scale contact time is comparable to this of the CG particles, whose interaction with each other is limited to repulsion. Thus, the faster diffusion of proteins around these crowdres is fully justified. By far the highest value of the diffusion coefficient in the surroundings of polysucrose (higher than for the CG model) can be related to the very fast diffusion of the crowdres themselves (Table 3.3), and the resulting dynamics of the water molecules.

The shape of both NS4A and crowder proteins may also be relevant to the effect of the crowded environment on the rotational diffusion of the protein, but here far more importance is attributed to the interactions of the rotating molecule. In this sense, rotational diffusion is a more complex process - it will be limited both by steric effects and by the formation of smaller or larger complexes with other molecules.

In this work, rotations of the protein have been determined by calculating rotational correlation times, obtained by comparing the trajectory with a set of vectors pointing random directions. The results of this analysis are presented in Table 4.3.

Environment	Rotational correlation time τ_r [ns]	Rotational diffusion coefficient D_r [ns ⁻¹]
Water	45.1 ± 2.5	7.5 ± 0.4
PEG	76.7 ± 7.5	12.8 ± 1.3
Polysucrose	61.3 ± 2.2	10.2 ± 0.4
CG crowdres	63.8 ± 2.3	10.6 ± 0.4

Table 4.3: Rotational correlation times τ_r and rotational diffusion coefficients D_r calculated for the NS3 protease in water and in crowded environment, with SEM errors. Both sets of values are corrected for the altered viscosity of TIP3P water.

In the presence of PEGs, the protein rotates as a complex of NS3, unstructured NS4A and irregularly distributed polymer chains in various shapes that additionally interact with other crowdres, forming a multi-layered structure whose shape is very far from globular. In contrast, the protein surrounded with polysucrose rotates as a complex of NS3/4A clustered with evenly distributed compact-shaped polymers that do not form stable contacts with other polysucrose molecules - those crowdres remain layered with water preventing the formation of stable crowder-crowder interactions (see Chapter 3, Figure 3.14, 3.16 and Table 3.4).

Given the different interaction networks in each environment, we are looking at rotating multi-component complexes of significantly different structures - from highly irregular in the case of a protein-PEG complex to almost globular in the case of a protein-polysaccharide complex. Looking at rotational diffusion from this perspective, the different results for these two environments are

easily understood.

What is somewhat less straightforward, the rotational diffusion of CG crowdres is also slowed down relative to the theoretical value for proteins in the vicinity of hard spheres, as if significant protein-crowder interactions were taking place. Taking into account the lack of protein-crowder contacts, and the RDF strongly indicating the absence of interactions (Table 3.4, Figure 3.17) - the results can be only explained by non-globular shape of the protease itself.

In such case, the difference between rotational diffusion of the supposedly-globular protein surrounded with hard spheres (in the theoretical model) and the diffusion of NS3/4A in the presence of almost non-interacting spherical particles can be interpreted as the magnitude of rotational slow-down that is purely shape-induced (as shown in the following section, the shape of the NS3/4A protease complex is very similar in the presence of CG crowdres and polysucroses). Going further, the smaller reduction in diffusion for proteins in the environment crowded with saccharides may be related to the fact that polysucrose molecules add globularity to the protein-crowder complex, thereby facilitating rotational diffusion despite the fact that protein-crowder interactions occur.

A final conclusion that can be drawn from the analysis of the diffusion coefficients obtained from the trajectories is further information on protein-crowder interactions. Since the formation of complexes with crowdres affects the rotation of the protein to a much greater extent than translational movement, analysis of (de)coupling of those two effects may provide quantitative information on the degree of the interactions of the protein with particular types of crowdres.

Environment	Translational diffusion coefficient [$\text{\AA}^2/ns$]	Rotational correlation time [ns]
Water	6.98 ± 0.04	45.1 ± 2.5
PEG scaled	$4.66 \pm 0.01, (-33 \%)$	$76.7 \pm 7.5, (+70 \%)$
Polysucrose	$5.18 \pm 0.05, (-26 \%)$	$61.3 \pm 2.2, (+36 \%)$
CG crowdres	$4.85 \pm 0.05, (-30 \%)$	$63.8 \pm 2.3, (+41 \%)$
HS model	4.61	52.3

Table 4.4: Translational diffusion coefficients D_t and rotational correlation times τ_r calculated for the NS3 protease in water and in crowded environment, compiled to show the (de)coupling of rotational and translational diffusion. In brackets I have put the percentage differences between the D_t and τ_r coefficients in the presence of a given type of crowdres and the value of the coefficients in water.

For all three types of crowdres the retardation of rotational diffusion is increased over the values predicted for HS model. This effect is caused partly by the non-spherical shape of the protein and partly by interactions with crowder molecules. In distinguishing between these two effects, it is helpful to compare how much the values of D_t and the rotational correlation times τ_r change in different environments, relative to the diffusive properties in water (Table 4.4). The relations of changes in translational to rotational diffusion for polysucrose and CG crowdres are in fact highly

comparable, whereas for the protease surrounded with PEG the effect of crowding on both types of diffusion is considerably more decoupled, with the limitation of protein rotation far greater than for other crowdiers indicating the presence of protein-crowders interactions. This picture confirms the results obtained earlier indicating the presence of protein-PEG interactions, but at the same time it shows that interactions with polysucrose, although present, do not interfere with the diffusion properties of the protein. Moreover, in the environment crowded with polysucrose the translational diffusion of the protein is slowed down less than with other crowdiers, which may contribute to the increased activity of NS3 in this environment.

4.1.5 Simulations with the peptide substrates

To measure what effect the presence of a crowded environment might have on NS3/4A protease substrates, sets of four types of simulations were compared - simulations of the protein in water and in the presence of PEG were complemented by simulations of the protein with substrates, and the protein surrounded by PEG crowdiers and substrates (Figures 4.4, 4.5). The peptide substrate was chosen to best represent the peptide used in the activity assay, removing elements that were only useful in the experiment and whose parameterisation was a risky undertaking - the parameters obtained for the EDANS and DABCYL fluorescent tags using the CgenFF service were found to have large errors and introducing them into the simulation would risk producing inaccurate results. Therefore, I decided to simplify the structure (Figure 4.3), using a peptide of a sequence *Ac – DEDEEAASK – NH₂*, reducing the elements which are important to the laboratory assay, but not necessarily needed in the simulation.

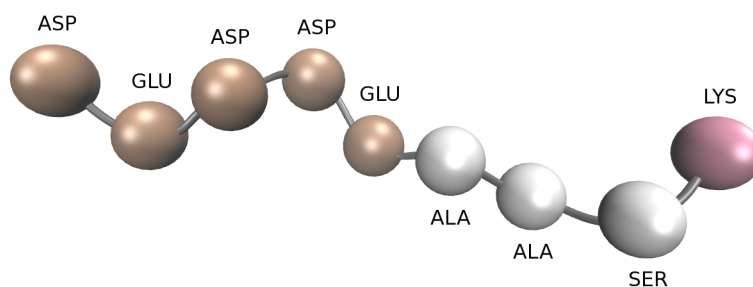


Figure 4.3: Peptide substrate of the NS3/4A protease in schematic representation, colored by amino acid type, with acidic, negatively charged residues in brown and positively charged residue in pink.

Both fluorescent tags have been removed, along with a minor modification to one of the alanine residues and the ester bond at the cleavage site. Introducing the ester bond enhances effectiveness of the enzymatic reaction⁹⁰, but is not expected to affect diffusion of the substrate. During the simulation time I was not expecting any protein-peptide contacts or reaction events, for this purpose entirely different simulation profile would be needed, with significantly increased time scale and probably different representation of the macromolecules, and perhaps with some level of coarse-graining included. For the purpose of observing changes in diffusive properties, all-atom representation and explicit solvent appeared most suitable.

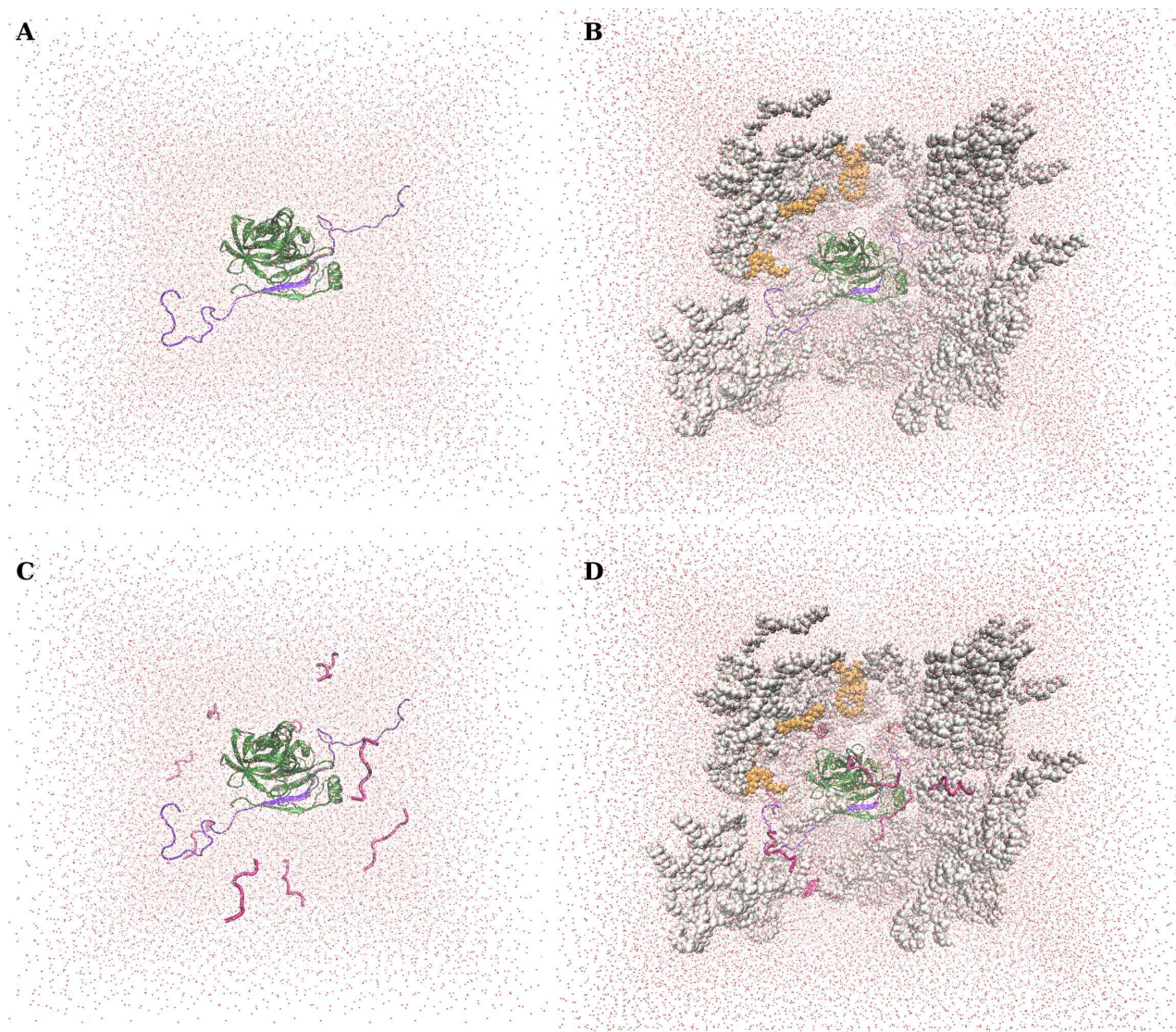


Figure 4.4: Four types of systems used for all-atom MD simulations of the NS3/4A protease in different surrounding. All simulations contained water molecules in an explicit solvent model, of which only the oxygen atoms are shown in a dot representation. A. The control simulation of the protease in water, B. the protease surrounded by PEG crowders in 130 copies, C. the protease with 10 copies of peptide substrates, and D. the protease surrounded by crowders and substrates. The NS3 protease is shown in green, the NS4A in violet, the peptide substrate in magenta and the crowders in grey. Two example conformations of PEG molecules are highlighted in brown. Figure from a published work⁹¹.

Molecule	Decrease in diffusion coefficient upon adding substrates	
NS3/4A protease in water	from 6.89 to 6.82 [$\text{\AA}^2/ns$], (-1 %)	
NS3/4A protease with PEG	from 4.73 to 4.65 [$\text{\AA}^2/ns$], (-1.7 %)	
Diffusion coefficient of the substrate D_t [$\text{\AA}^2/ns$]		
	In water	In water with PEG crowdres
Peptide substrate	13.61 ± 0.07	8.85 ± 0.32

Table 4.5: Translational diffusion coefficients of the peptide substrates in water and in the presence of PEG crowdres, and the effect of adding peptide substrates on the diffusion of the NS3 protease.

Addition of peptide substrates to the system has a negligible effect on the diffusion of NS3/4A protease, both in water and in the presence of crowdres. Substrates were added in low concentrations, not adding much to the restriction of the space available for the protein (Figure 4.5).

In contrast, in the presence of crowdres, translational diffusion of the substrate molecules themselves was reduced by 35 % compared to their diffusion in water (Table 4.5). This difference is even higher than the effect of crowding on the translational diffusion of NS3/4A, shown earlier in Table 4.2.

Restriction of both protease and substrate diffusion provides an explanation for the significant decrease in protein activity in the PEG environment - slowed movement means also a reduced frequency of protein-substrate contacts leading to catalytic events. Another factor might be a significant slow-down of protein rotations in the presence of PEG, which can also negatively affect the probability of these contacts.

A more interesting question is how the diffusion of the protein and substrate is affected by the presence of polysaccharides. Although simulations with the protein, Ficoll-like crowd and substrates have not been carried out yet, a few things can be said about this topic based on the data already collected. It is known from laboratory experiments that the presence of Ficoll, in contrast to the presence of PEG crowdres, increases the efficiency of NS3/4A catalysed peptide cleavage⁹¹. However, simulations of the protease in the setting of polysucrose molecules suggest that it is highly unlikely that the increase in activity is related to conformational changes within the protein or to crowd interactions with the active site of the enzyme. Therefore, it remains to look for the reasons for this effect in changes in the diffusion of macromolecules and substrates.

Indeed, one unexpected effect on diffusion has already been observed in this system - relative to other crowdres, polysucrose is by far the least restrictive of the translational diffusion of the protein. The decrease in the diffusion rate is even smaller than for the non-interacting coarse-grained

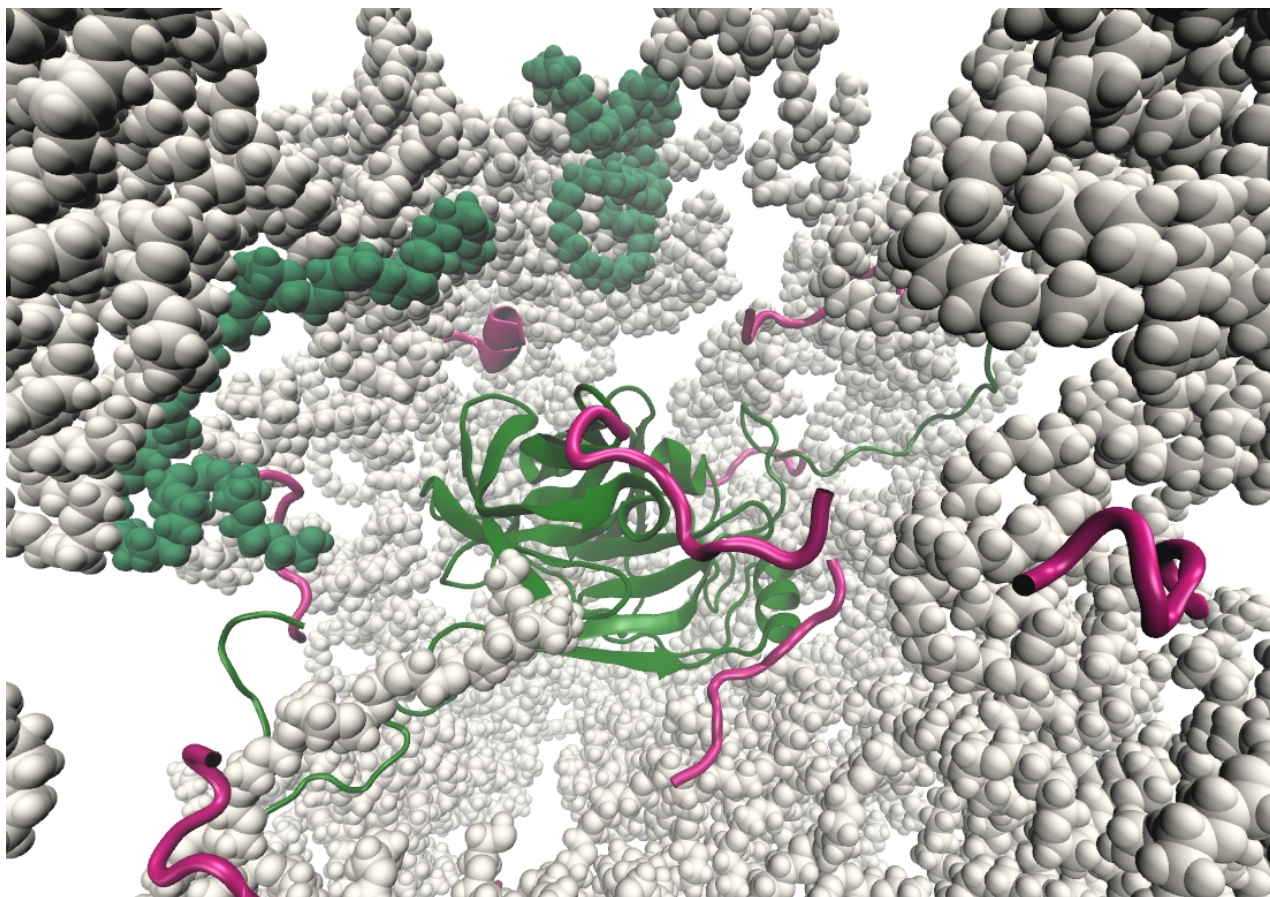


Figure 4.5: A simulation system with the NS3/4A protease surrounded with PEG and the peptide substrates, seen from another perspective. The protease is shown in green, the substrates in magenta. The PEG crowders are shown in all-atom representation, with two crowder molecules coloured in sage-green, to highlight the variety of conformations of these polymers. Figure from a published work⁹¹.

crowders. Polysucrose molecules themselves also move unexpectedly fast. It is possible that the accelerated diffusion of crowdres also entails increased diffusion of peptide substrates. In the absence of trajectories, what can be done is to re-analyse the amino acid sequence of the substrates and, by comparing it to data on the interaction of polysucrose crowdres with the protein surface, estimate the substrate-crowders interactions.

The peptide substrate used in the activity assay has a high acidic residue content, and the resulting charge of $-4e$ (Figure 4.3). Interestingly, for the interactions with acidic amino acids, I observed a difference in the number of contacts and hydrogen bonds formed with PEG and polysucrose crowdres (Table 4.6).

Residue	Hbonds fraction [% of all hydrogen bonds]		Contact fraction [% of the simulation time]	
	PEG	Polysucrose	PEG	Polysucrose
LYS	4.0	4.7	59	74
SER	8.1	6.5	90	83
ALA	8.2	4.8	78	81
ASP	1.0	7.4	44	70
GLU	1.9	9.0	46	66

Table 4.6: Hydrogen bonds and contacts between the NS3/4A amino acids and crowder molecules, calculated for single 500 ns trajectories of the systems containing the NS3/4A protease surrounded with either PEG or polysucrose crowdres. Residue types are ordered by their number of hydrogen bond donors. The Hbonds fraction columns show the fraction of unique hydrogen bonds between a given type of amino acid and PEG or polysucrose crowdres, as a percent of all hydrogen bonds formed between NS3/4A and crowdres. The maximum donor-acceptor distance was set to $3.5\text{\AA}$ and the donor-hydrogen-acceptor angle criteria were extended to no less than 120 degrees. The Contact fraction columns show the fraction of the simulation time when any crowder atom was present within $3.5\text{\AA}$ from the atoms of the amino acids of a given type.

Table 4.6 shows that, contrary to PEGs, polysucrose molecules prefer to hydrogen bond with ASP and GLU which constitute more than half of the substrate sequence. Given that the amino acids in the substrate have a high potential to interact with polysucrose molecules, it is also very likely that Ficolls will interact more readily and for a longer time with the substrate than PEGs. Adding to this the increased diffusion rate of polysucrose, one can venture to conclude that the presence of Ficolls has a high chance of accelerating substrate diffusion.

4.2 Structural dynamics of the NS3/4A protease in crowded environment

The NS3/4A protein consists of the NS3 protease domain and the NS4A cofactor, which remains largely unstructured in water. As might be expected, it is the unstructured tails (residues 1-22 and 35-54) that show flexibility and are affected by the changing environment, while the globular protease remains stable (Figure 4.6). This is also true for the NS4A fragment which forms a number of stable bonds with NS3, being a part of the NS3 β -barrel. In the trajectories this central part of NS4A behaves as if it was a part of the globular protease.

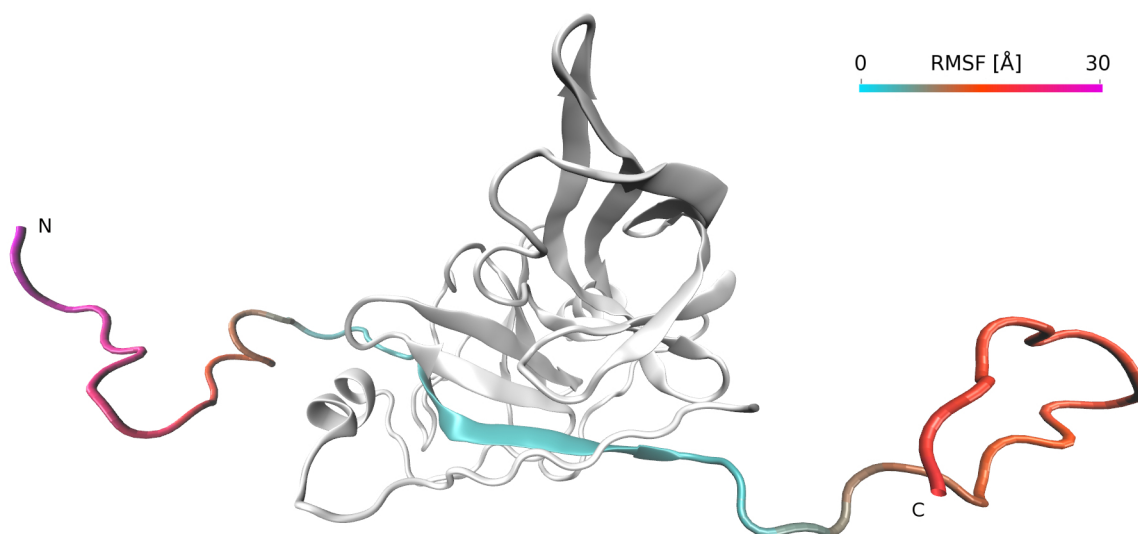


Figure 4.6: The structure of NS3/4A complex with the NS3 protease in white and the NS4A protein cofactor colored by RMSF values for the $C\alpha$ atoms, averaged over three copies of the trajectories of the protease surrounded with PEG crowders. Conformation has been chosen based on the average structure adopted during a simulation with PEG crowders. Figure from a published work⁴⁴.

High mobility of the NS4A cofactor, with RMSF reaching 30 Å and the variability of the structural dynamics affected by the crowded environment makes the N- and C-terminal NS4A the main object of observation in the obtained trajectories. In contrast, the RMSFs of the NS3 protein and the central part of NS4A remain below 5 Å and maintain stable secondary structures in all trajectories.

4.2.1 Structural flexibility of the NS3 protease

Although the protease is generally stable, loops as in any protein show increased flexibility and higher RMSF values (Figure 4.7). This can be seen in the RMSF plot, which shows numerous peaks corresponding to loop regions. One would expect that the presence of crowders, especially full-atom crowders that show some interaction with the protein surface would in general increase the mobility of any loop. However, a closer look at the graphs of RMSF in a crowded environment compared to RMSF in water shows that there is no simple correlation - the mobility of some of the regions

increases in a crowded environment, others decrease, with different effects of different crowdors (Figure 4.8).

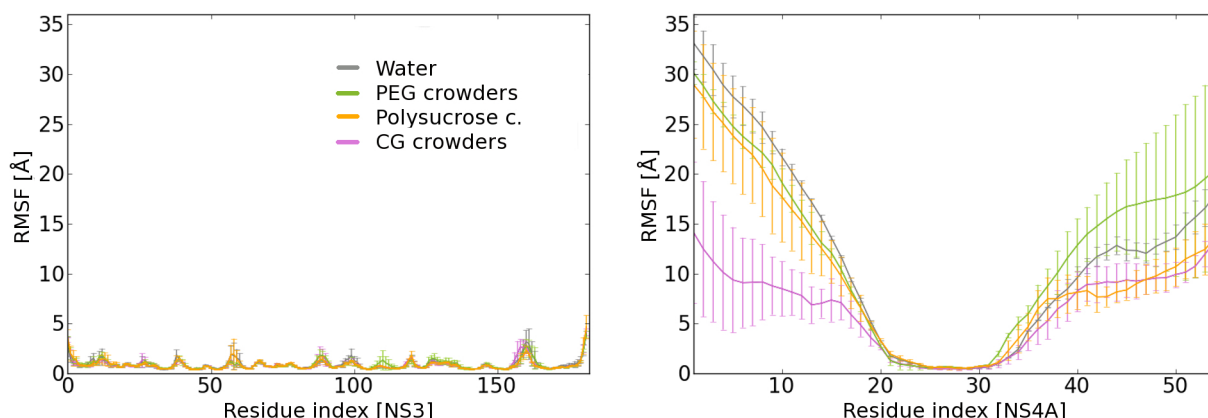


Figure 4.7: RMSF of the NS3 protease and the NS4A protein cofactor in water and in the presence of three types of crowdors. RMSF values were calculated for CA atoms, with the average structure used as reference positions of the atoms. Errors are calculated as the standard errors of the mean, based on three copies of each simulation.

Flexibility of certain regions is enhanced by both the presence of PEG and polysucrose - such elements include amphipathic helix 12-24 and the zinc-binding region. Both regions are critical for the structure and biological function of the protein (Figure 4.8). Other interesting elements include Arg161 - an amino acid which is located at the top of one of the β -sheet turns. It is a large residue, exposed towards the solvent, which has both one of the highest RMSF values and one of the highest numbers of contacting crowder atoms (Figure 4.9). However, the flexibility of this turn is not enhanced by crowding in any trajectory.

We already know, that the vicinity of active site of the protease is less populated with crowder molecules. To figure out which regions of the protease are most densely surrounded with crowdors I calculated average numbers of crowder atoms within 5 Å from each residue. Within the most populated regions are those most relevant to the protease functionalities - the membrane binding site and the region that binds a zinc ion - an element which is crucial for the stability of the whole protease. Additional region, which is closely surrounded by both PEG and polysucrose crowdors is the region around Arg161. It is a large residue sticking out towards the solvent at the tip of a β -turn, corresponding to a large peak on the RMSF plot. However it is not the contact with crowdors that influences its flexibility, but the position in the protein structure - RMSF values for this region are high also for the simulations in water (Arg161 position shown in Figure 4.8, RMSF peak visible in Figure 4.7. Fig. 4.8 shows that this residue is actually most flexible in water).

What catches attention are generally larger atom counts for amino acids surrounded by PEGs - despite lower count of close contacts and hydrogen bonds relative to polysucrose molecules. That results from far more versatile conformations adopted by PEG - the chains often wrap around the protease or single amino acids, as in the case of Arg161, increasing the number of crowder atoms in the immediate vicinity of particular amino acids.

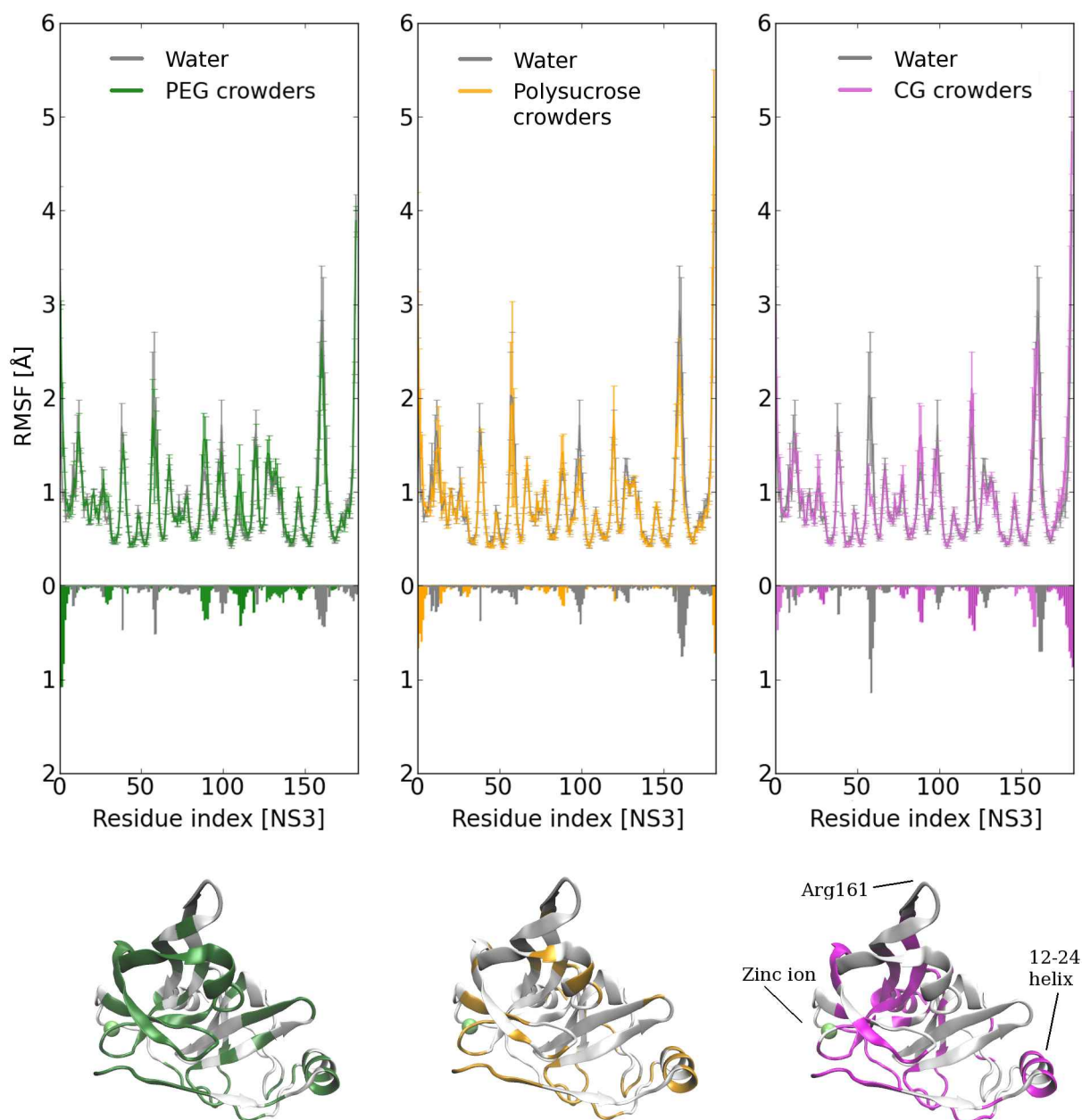


Figure 4.8: RMSF of the NS3 protease in crowded environment, calculated with respect to the average structure. RMSF in the presence of crowders is compared with RMSF obtained for the protease in water, and the differences are shown on the graph below $y=0$, with the bars in gray for the residues which were more mobile in water and in colour when the RMSF was larger in the presence of respectively PEG, polysucrose and the CG crowders. Additional way of presenting the differences between RMSF in water and in crowded environment is showing the values on NS3 structures. At the bottom figure, on each structure residues with higher RMSF in crowded environment are colored in green, orange or magenta for amino acids showing higher mobility with respectively PEG, polysucrose and CG crowders.

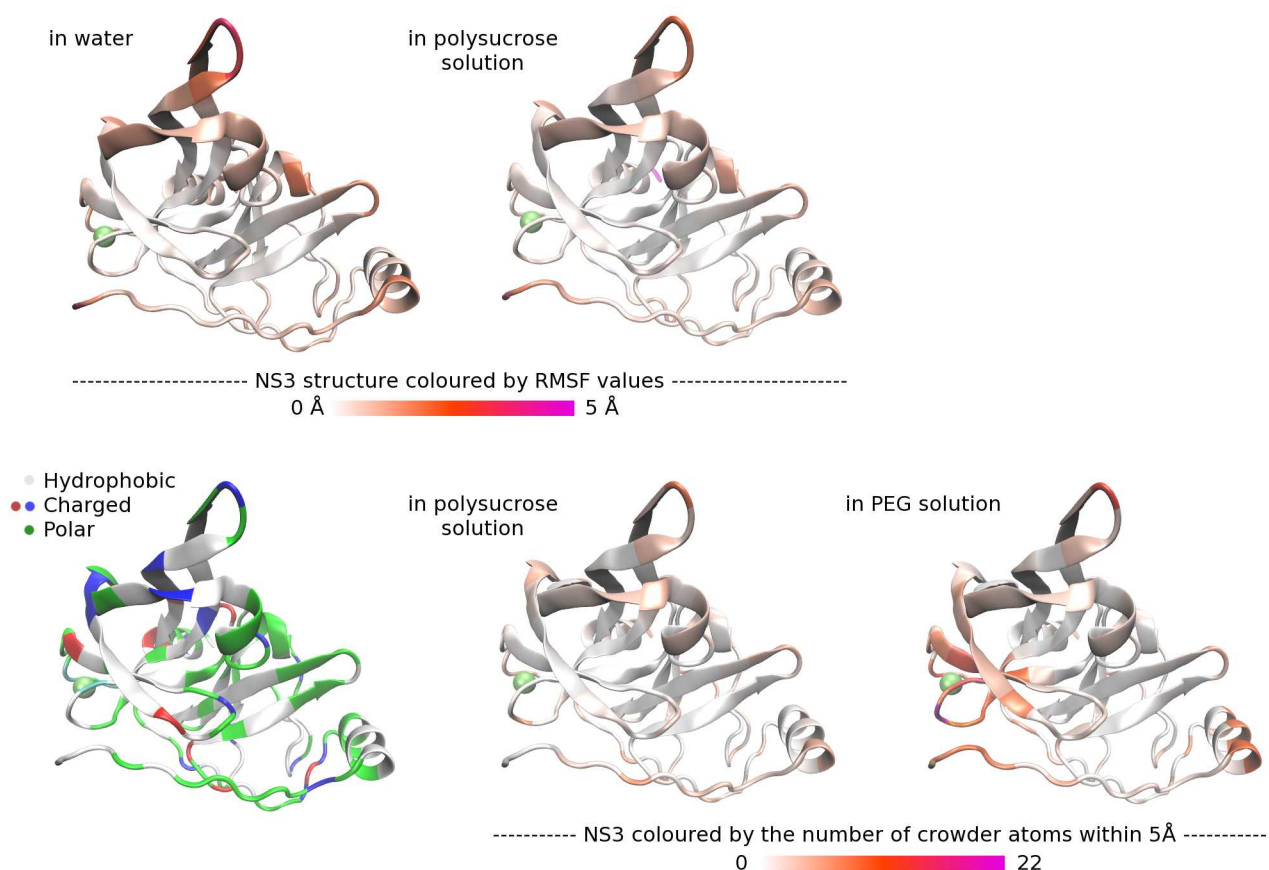


Figure 4.9: RMSF of the NS3 protease in water and in the presence of polysucrose compared with the number of crowder atoms within 5 Å from a given amino acid of NS3 simulated with polysucrose and PEG crowdres. Comparison of structures coloured according to different properties helps to assess the extent to which the increased mobility of certain NS3 fragments is related to the number of crowder molecules in its vicinity. Additionally the structures are compared with the NS3 structure colored by residue type.

4.2.2 Mobility of the N- and C-terminal tails of the NS4A protein and their interactions with the NS3 surface in two water models

The unstructured N- and C-terminal fragments of NS4A move freely in the aqueous environment, forming transient interactions with the NS3 protease surface. The tails interact with different amino acids on its surface, with no particular preferred site. How stable these interactions are largely depends on the hydrophobic properties of the residues, as was observed when performing control simulations: the protease in water in the standard TIP3P model and the protease with modified water interactions to prevent aggregation of PEG crowdors (see Methods, section 2.2.5). As a result of the need for this modification, two sets of trajectories were obtained - in the most common water model and in an improved, more realistic model that does not overestimate the hydrophobicity of molecules dissolved in water, including protein residues. Although only the second improved set of trajectories was used for further analysis, the comparison of the sets allows for an useful analysis of protein behaviour as a function of hydrophobicity.

The essence of the introduced modification of the TIP3P model is that it enhances the interactions of water molecules with the solute, and thus, it relatively weakens the interactions between amino acids. Consequently, the unstructured NS4A tails adhere more stably to the NS3 surface. This can be seen in the radius of gyration plot (Figure 4.10), which shows that in a standard force field for most of the trajectory the NS4A cofactor adopts a single conformation with a relatively small radius, whereas with upscaled parameters the radii and hence conformations are remarkably diverse. This is due to the greatly increased conformational stability in which the N- and C-terminal tails adhere tightly to the protein surface.

Scaling water parameters also affects another important property of the protein, namely its secondary structure (Figure 4.11). The increased hydrophobicity of amino acids in a standard TIP3P water force field results in increased stability of helical structures formed mainly on the N-terminal fragment of the cofactor. The helical fragments form in both versions of the water model, however, they undergo rapid unfolding in scaled water. This behaviour of the N-terminal NS4A fragment, in this case caused by a difference in the degree of protein-water interaction, also appears in further analyses, where it is caused by factors related to the protease surroundings.

4.2.3 Mobility of the N- and C-terminal tails of the NS4A protein and their interactions with the NS3 surface in different environments

The easiest way to quickly assess the conformational changes of a partially unstructured protein is to measure its radius of gyration. This method does not give an accurate characterisation of the structural dynamics, but it allows a preliminary comparison of these dynamics in several different environments: in water (with upscaled parameters) and in the crowded environment, with full-atomic PEG and polysucrose molecules and with the coarse-grained crowdors (Figure 4.12).

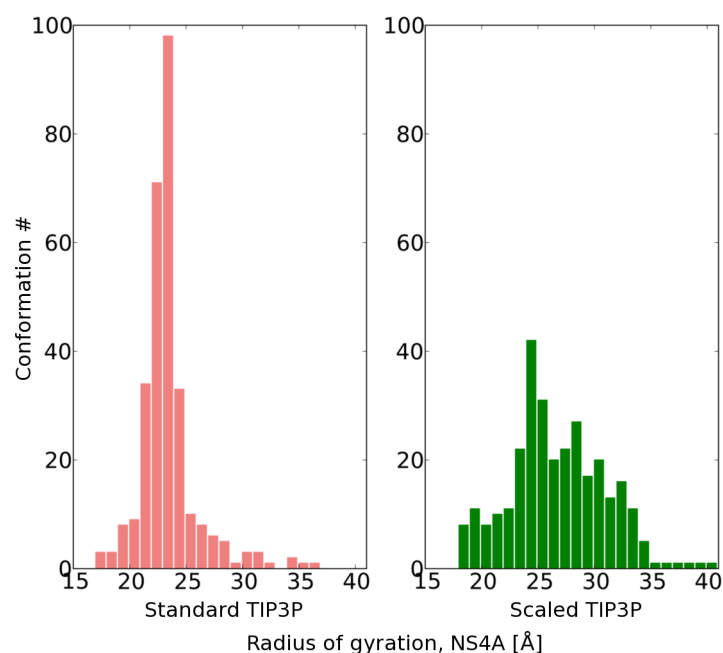


Figure 4.10: Radius of gyration of the NS4A protein cofactor measured for the trajectories in standard TIP3P water and in water model in which water-solute interactions were upscaled by the factor of 1.09. Radius of gyration is shown as a histogram of conformations with a given R_g , with conformations collected for three 100-frame trajectories of each kind. Figure from published work⁴⁴.

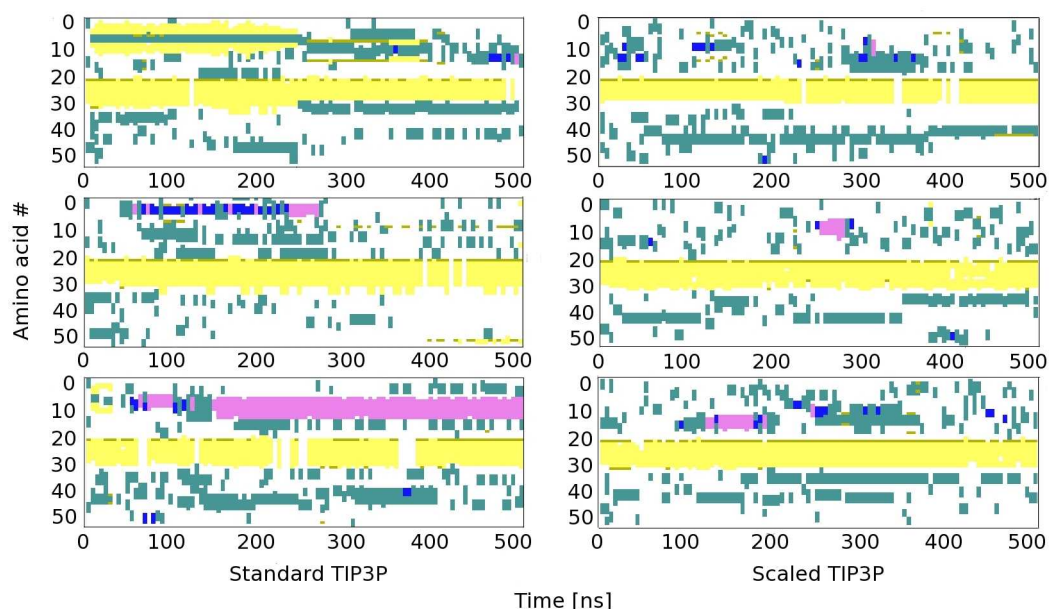


Figure 4.11: Secondary structure of the NS4A protein cofactor, observed in the simulations of NS3/4A in water in standard TIP3P model and with scaled parameters for the protein-water interactions. Figure from published work⁴⁴.

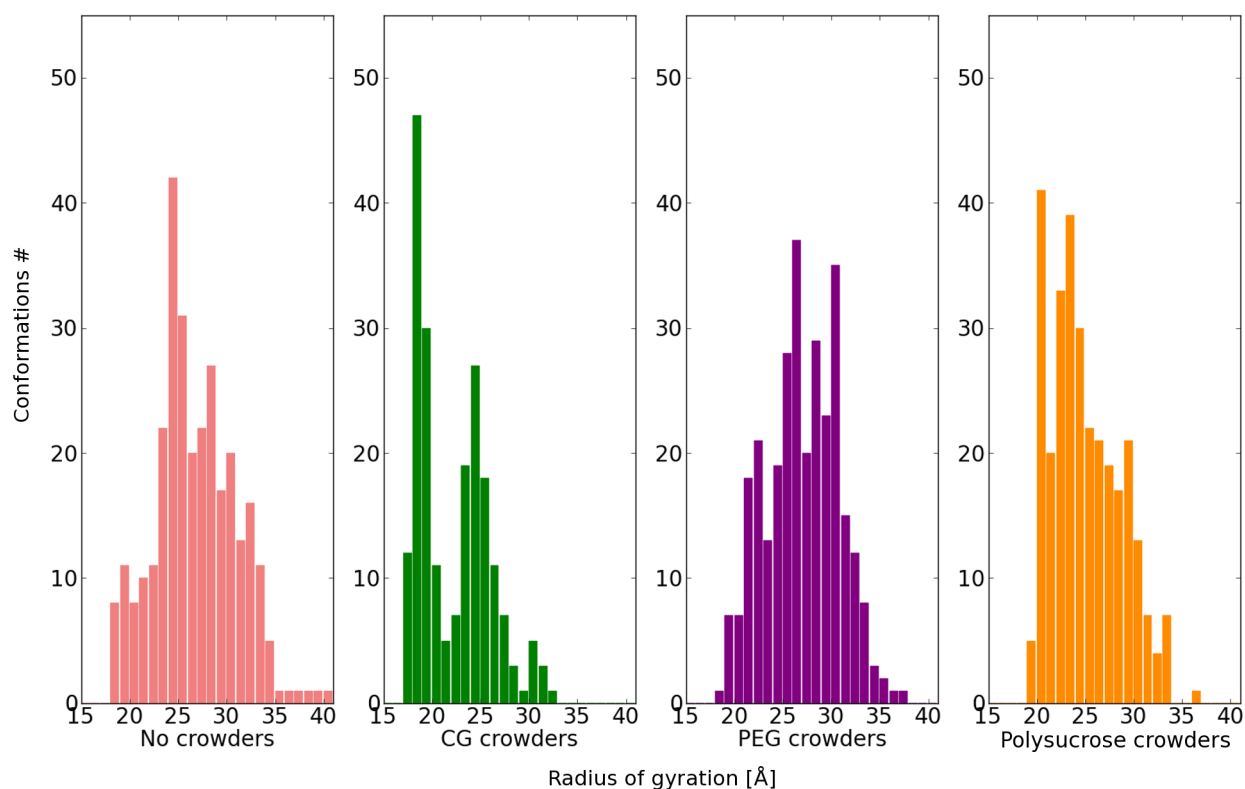


Figure 4.12: Radius of gyration of the NS4A protein measured for simulations of the NS3/4A protease in water and in the presence of crowders (PEG, polysucrose and CG particles). Radius of gyration is shown as a histogram of conformations with a given R_g , with conformations collected for three 100-frame trajectories of each type.

Only a rough picture of the NS4A dynamics can be deduced from the R_g analysis. However, already at this stage one can see a trend that will only become more apparent in further analyses - non-interacting CG crowders cause a definite shift of the observed conformations towards compact structures, while full-atom crowders, as well as aqueous environment, allow the cofactor to retain higher flexibility. In the case of the NS3/4A complex, conformations with low R_g values are those in which the tails (both or one) adhere tightly to the NS3 surface, whereas those with higher values are those in which the tails move freely without adhering to the protease. Thus, the presence of crowders that only provide volume exclusion favours 'sticking' of the tails to the NS3 surface and more stable NS3-NS4A interactions.

A wider range of R_g values in almost all systems other than crowded with CG molecules certainly corresponds to a wider range of adopted conformations, but to say more about them, more case-specific methods are needed.

To characterise the dynamics of this particular protein it would be more meaningful to obtain information on the conformation of unstructured tails of NS4A in relation to the NS3 protease, and clearly identify at which points in the trajectory the tails are moving freely and at which points they are positioned just off the surface of NS3. Such a measure would have to be calculated taking into account the distances of individual residues from the protein surface (Figure 4.13). For technical

reasons, it is much more straightforward to obtain information about the distance of residues from a certain point - for example, of the centre of the protein; and in the case of a globular protein like NS3, this value can be used to estimate the distance from its surface.

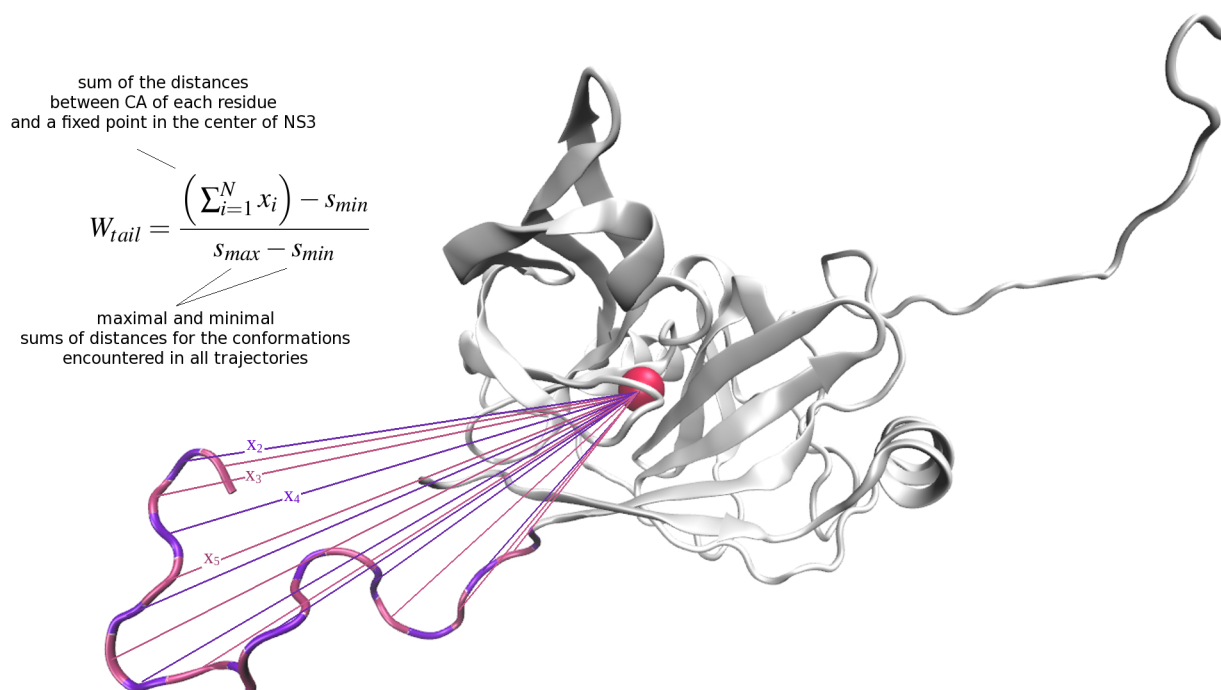


Figure 4.13: A scheme of calculating the wrapping coefficient (W), which helps classify the NS3/4A protein conformations to determine the extent to which the NS4A tails adhere to the NS3 surface. The formula for the coefficient is based on calculating the sum of the distances of the individual residues from a fixed point located at the centre of the NS3 protein (one of the atoms of Gly141, marked with sphere in magenta). In the diagram, the C-terminal amino acids of the NS4A tail (33 to 54) are coloured alternately pink and purple, with the lengths included in the W factor as x_i values.

With this in mind, I introduced a measure that is based on calculating the sum of the distances between the C_α atoms of successive amino acids of the C- or N-terminal tails and a specific fixed point in the centre of the NS3 protein. Comparing the sum calculated for a given conformation with the maximum and minimum sum for the conformations observed in the trajectory, a number between 0 and 1 is obtained, where 0 indicates a conformation in which the tail maximally adheres to the NS3 surface, and 1 indicates a conformation with the amino acids of the tail are maximally distant from the NS3 surface (Figure 4.13).

The W coefficients calculated for the conformations obtained in the NS3/4A protease simulations for both cofactor tails were used to prepare two types of analyses; in the first, the coefficients for each system are shown in the histogram form to represent the contribution of each conformation to the trajectories obtained; in the second, the coefficients are used to trace the conformational changes of NS4A over time.

4.2.4 The C-terminal tail of NS4A

Both C- and N-terminal tails of NS4A remain unstructured in water, but their conformational dynamics observed in the trajectories obtained from the molecular dynamics simulations are significantly different. The C-terminal tail mainly adopts conformations with the tail adjacent to the protein, described with the wrapping coefficients (W) close to zero. W graph shows a peak between $W=0.2$ and 0.3 and a rapid decrease in the number of conformations above these values (Figure 4.14).

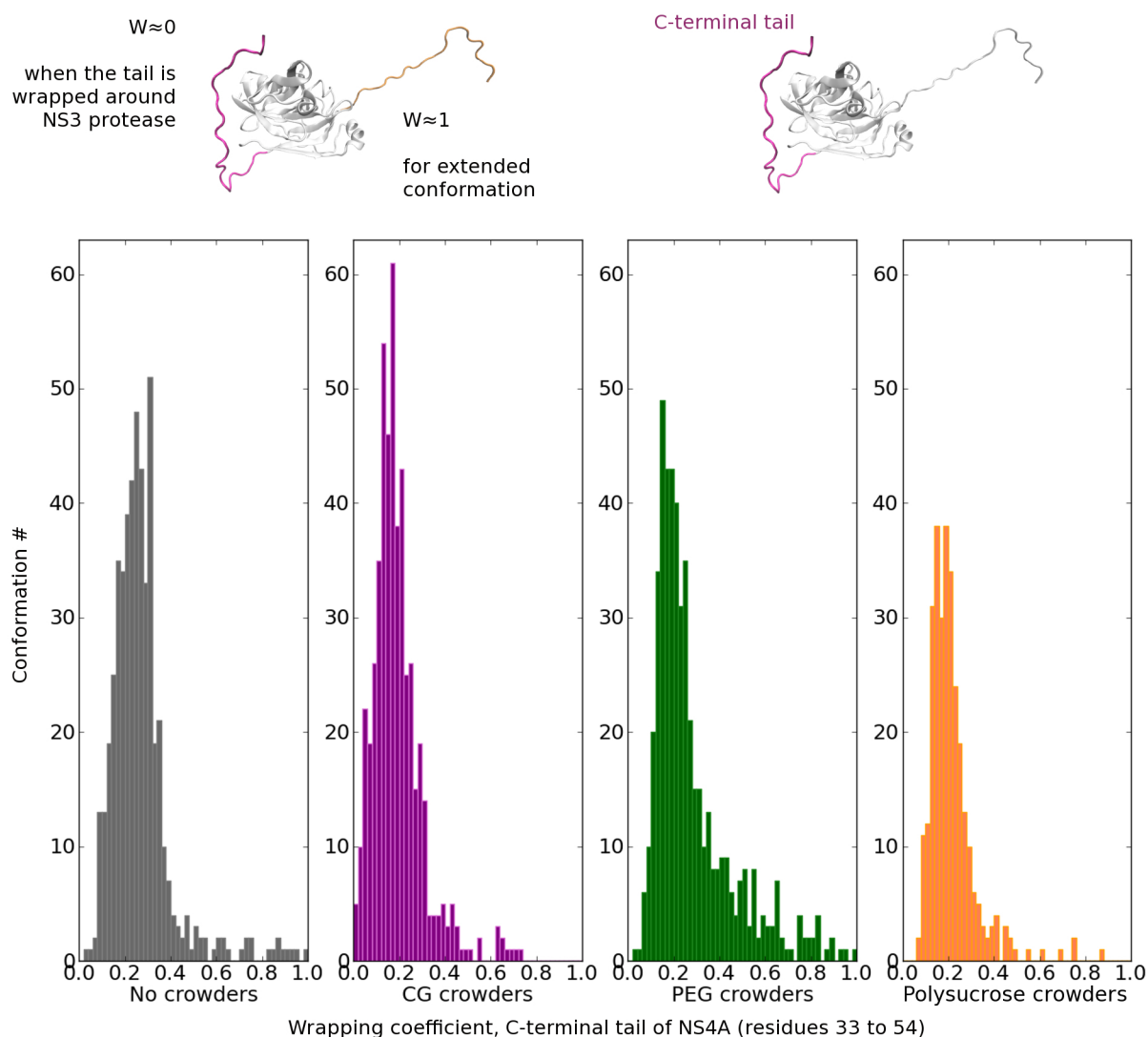


Figure 4.14: Wrapping coefficients for the C-terminal tail of NS4A (residues 33 to 54) shown as a histogram with bin size of 0.2. As shown above the histogram, $W=0$ for the conformations with minimal sum of distances between $C\alpha$ atoms of the tail and a fixed point in the center of NS3, and $W=1$ for the conformations for which the sum is maximal.

In the PEG surroundings, a slightly higher proportion of extended conformations is observed, and in the CG crowder setting the peak is sharper and shifted towards lower W - both of these small differences are in line with observations from the R_g analysis. However, the overall picture of C-terminal fragment behaviour is consistent - the tail sticks to the NS3 surface and forms a stable

protein-protein interaction. Visual observations of the trajectories confirm a strong preference for this fragment to remain in contact with the protease surface.

The sticking of the C-terminal tail of NS4A to the NS3 surface is related to the sequence characteristics of this fragment, namely the high content of acidic residues - as many as 7 out of 22 amino acids. The presence of these residues is important for the biological function of the C-terminus of NS4A. Depending on the degree of protonation of the Asp and Glu residues, the structure of this fragment alters. CD experiments showed that at pH lowered to 3.8, the C-terminal tail partially coils into an α -helix⁴⁵.

Theoretically, an analogous structural change is also possible as a result of other changes in the protein's environment, for example in the hydrophobic interior of another protein. The NS3 and NS4A proteins are not the only HCV proteins that enter into protein-protein interactions; on the contrary, they are part of a large replication complex. An additional consideration is that mutations within this region result in dysfunction of other HCV proteins - namely, decreased hyperphosphorylation of NS5A and disruption of virus assembly and release, processes in which NS3/4A is not involved. The environment in which the formed viral replication complex functions may also be important in terms of NS4A structure - HCV replication occurs inside the membranous web, a network of membrane vesicles formed during viral infection. In addition to the proximity of lipid membranes, the environment inside the web is not necessarily identical to that in the cytoplasm.

Regarding the interaction of the C-terminal tail of NS4A with the protease surface itself, one amino acid pair appears to be particularly important. Glu50 and Glu52 of NS4A form salt bridges with NS3 arginine or lysine residues in most of the obtained trajectories, both in water and in a crowded environment (Figure 4.15). The bridges with either Glu50 or Glu52 form with a wide range of amino acids on the surface of NS3, like Arg24, Lys68, Lys165 or Arg161, with some of them located on the two opposite sides of the protein structure (Figure 4.15). Apart from the salt bridges with Glu50 and Glu52, many other interactions mainly of electrostatic nature are formed, although they are highly unspecific - individual interacting pairs rarely reappear in more than one simulation copy. An important factor for the strength of those interactions in both the abundance of acidic residues on the C-terminal tail of NS4A and of negatively charged residues on the surface of NS3.

However, unspecific, dipol-dipol interactions and salt bridges appear to be strong enough to maintain the C-terminal tail at NS3 both in water and in the presence of all tested crowdiers. For this reason we do not observe any change in the behaviour of this fragment depending on the environment.

4.2.5 The N-terminal tail of NS4A

The N-terminal NS4A fragment shows much more conformational 'freedom' than the C-terminus - by not having such a high content of acidic residues, it does not form strong charge-charge interactions with the NS3 surface and does not 'stick' to it permanently. Conformational dynamics of this fragment is also strongly altered by the presence of crowded environment, and by the properties of

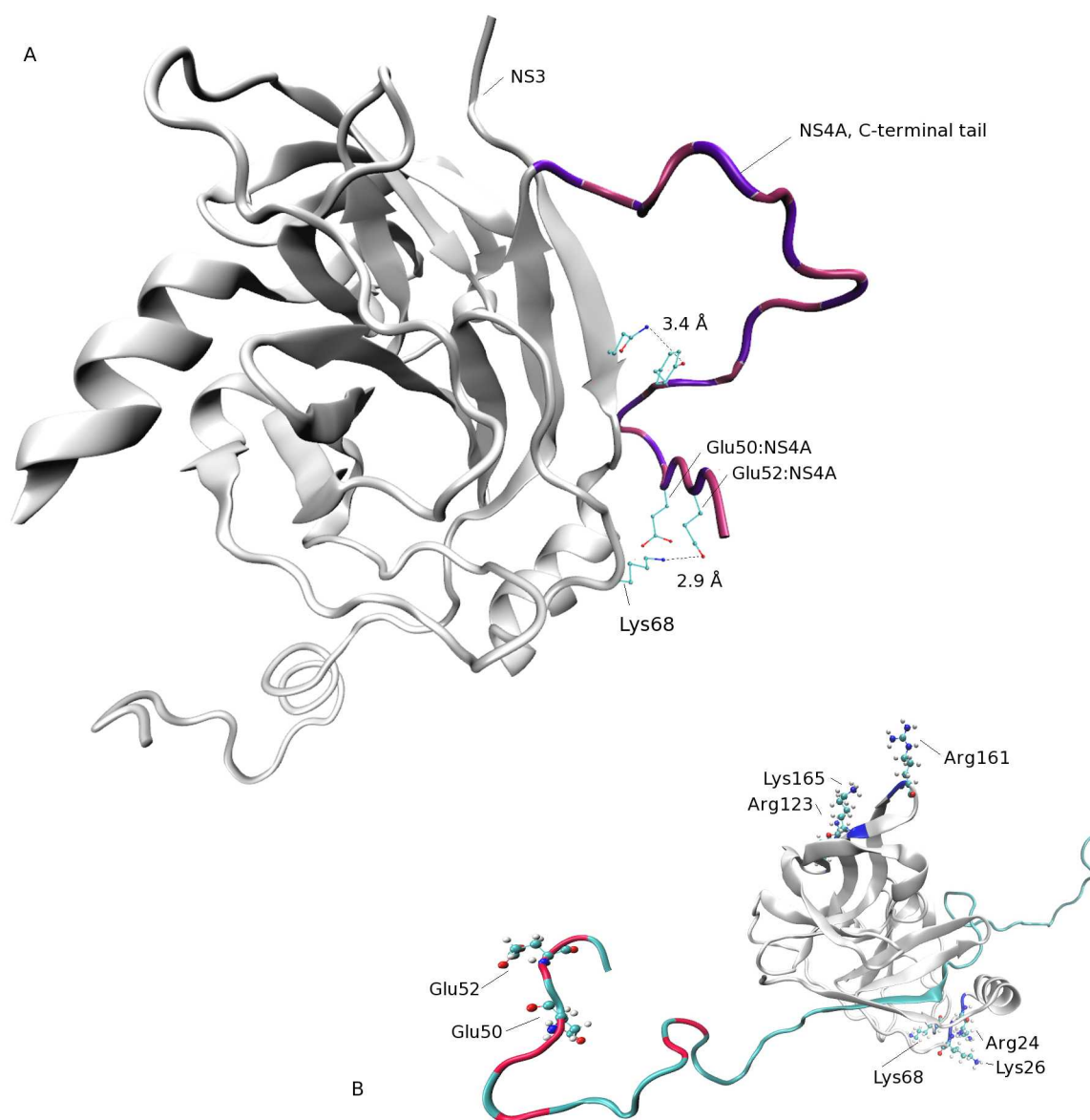


Figure 4.15: Interactions between the C-terminal tail of NS4A and the surface of the NS3 protease; A. Example of the conformation adopted by the C-terminal tail in which it partially adheres to the NS3 surface, with a characteristic pair of interacting residues shown in atomistic representation. The position of the salt bridge between Glu52 and Lys68 is also marked, together with the length of the bond. B. Structure of NS3/4A on which several elements characterising NS3 interactions with the C-terminus of the cofactor are marked, with the structure in the conformation obtained for simulations of NS3/4A in water. Acidic amino acids were marked on NS4A in magenta, with Glu50 and Glu52 in full-atom representation. Amino acids forming salt bridges with Glu50 or Glu52 in other obtained trajectories were also marked in this way.

crowder molecules (Figure 4.16).

In an aqueous environment, the wrapping coefficient for the N-tail conformations takes on a full range of values, with a distribution close to the normal distribution (Figure 4.16). This means that the tail moves fully unrestricted, adopting conformations ranging from adherent to the protease to completely extended, that is with maximum values for the distance of amino acids from the NS3 centre.

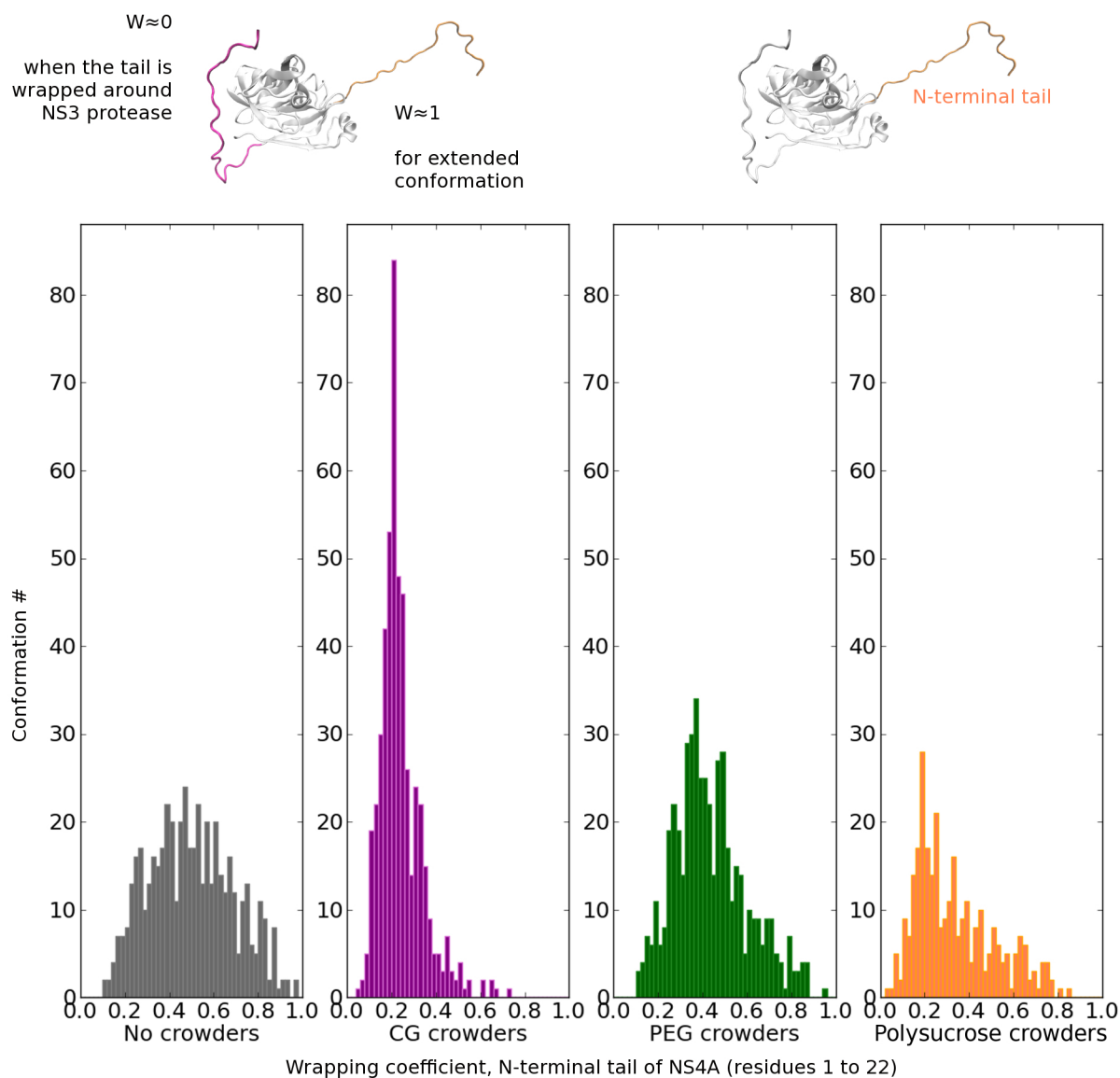


Figure 4.16: Wrapping coefficients for the N-terminal tail of NS4A (residues 1 to 22) shown as a histogram with bin size of 0.2. As shown above the histogram, $W=0$ for the conformations with the minimal sum of distances between $C\alpha$ atoms of the tail and a fixed point in the center of NS3, and $W=1$ for the conformations for which the sum is maximal.

In a crowded environment, the W range shifts more or less towards lower values, with by far the greatest effect in the case of coarse-grained crowders. In these simulations, the contribution of the extended conformation is minimal, and a sharp peak can be seen for $W=0.2$. This indicates that, in the presence of non-interacting crowders, the N-terminal NS4A fragment spent most of the

simulation time adhered to the protease surface.

Less obvious to interpret are the results obtained for simulations with full-atom crowders. Both PEG and polysucrose form transient clusters with the protein showing a range of interactions including hydrogen bonds, of which there are many more between the protein and polysucrose molecules (see Chapter 3, section 3.5). One would expect that the results of simulations with these crowders would be equivalent, however while PEGs reduce the share of extended conformations only marginally, polysucrose limits conformational diversity to a much greater extent.

Investigating the reason for the difference in how the distribution is affected in the presence of both types of crowders requires further insight. A quantitative measure of what the conformational distribution of the N-terminal tail looks like was obtained analysing the variability of wrapping coefficient over time (Figure 4.17).

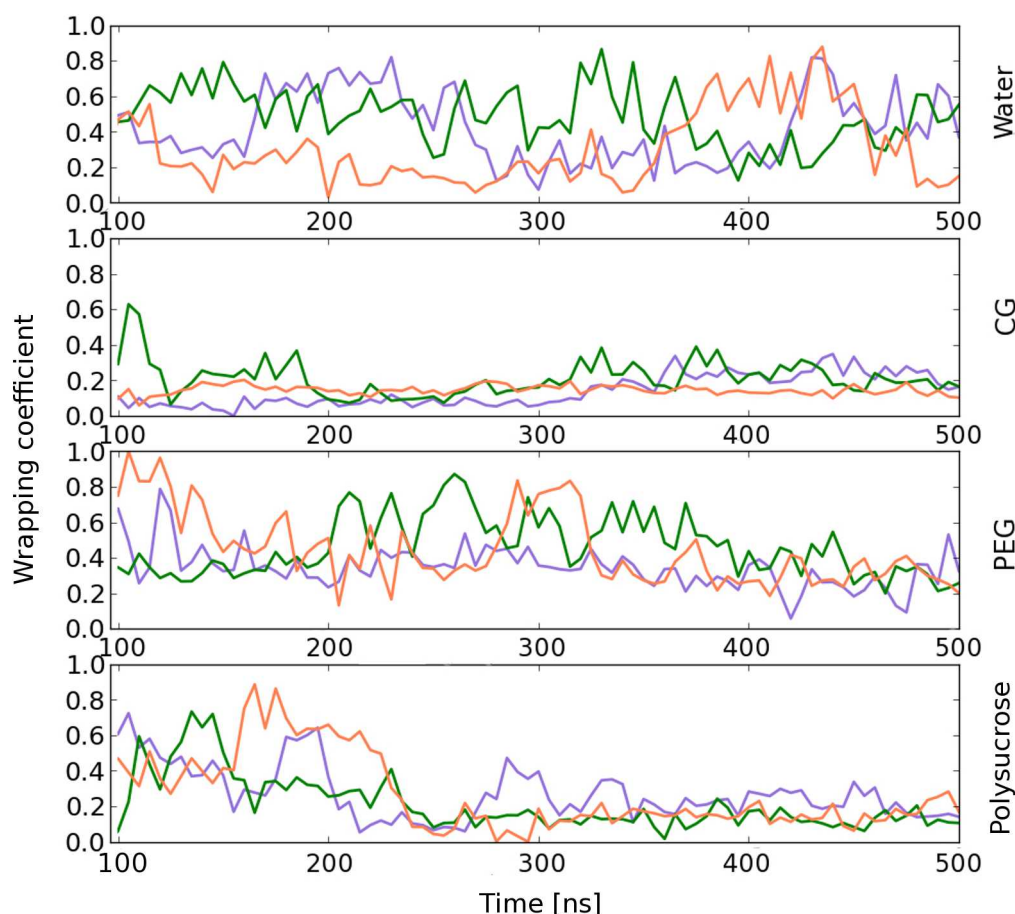


Figure 4.17: Wrapping coefficients for the N-terminal tail of NS4A (residues 1 to 22) shown as a graph of W values over simulation time. An important element of this graph is the presence of events (or lack of thereof) where the W value increases substantially from a value close to zero - those indicate situations where the tail completely detached from the surface of NS3. First 100 ns are not shown as in case of conformational dynamics it was considered the equilibration stage - during first 100 ns in the simulation with CG crowders W rapidly fell to minimal values.

The basic information that can be drawn from such a representation of wrapping coefficients is to identify events when the tail attaches to the NS3 surface and when and if it detaches. The tail is

adhered to the protease surface when W takes on low values, and the moment of detachment occurs when a significant increase in W is observed after a period of adhesion to the surface.

Regarding the dynamics of conformational changes of the N-terminus of NS4A, a clear division of trajectories into two groups can be observed. In water and in simulations in a PEG crowder environment, 'detachment events' can be easily distinguished. However, when the protein is surrounded with coarse-grained crowders or polysucrose molecules, after a single event of contact between NS4A N-terminus and NS3 the wrapping coefficient remains under 0.5 or lower. This means that the tail doesn't fully detach from the protease surface. This situation repeats in all five trajectories with CG crowders and three trajectories with polysucrose.

If the tail could detach in the presence of PEG and polysucrose, but remained adhered in the presence of the CG crowders, the mechanism could be explained easily. This would indicate a clear link between the presence of crowder interactions with the protein and the possibility of detaching the N-terminus from the NS3 surface.

However, here the situation is different: in the presence of CG crowders the tail does remain adhered to the surface of the protein, and the same is true for polysucrose. On the other hand, the tail has the ability to dissociate from NS3 in the presence of PEG crowders, even though PEGs have far fewer interactions with the NS3 surface than polysucrose - not only are the polysucrose-protein contacts themselves more numerous, but the resulting charge-charge interactions are also stronger due to the presence of highly polarised hydroxyl groups.

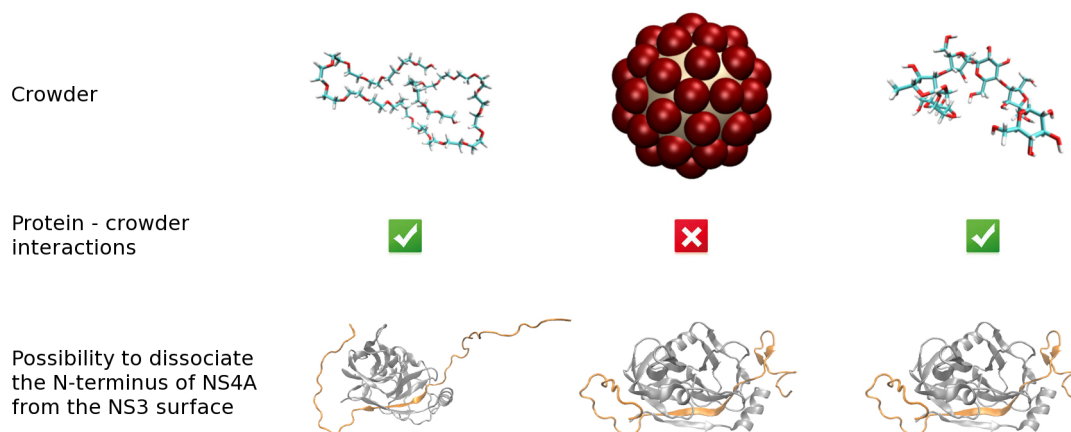


Figure 4.18: Comparison of trajectory observations with different types of crowdres - PEG, CG crowdres and polysucrose, regarding the relationship of the presence of protein-crowder interactions and the ability of the N-terminus of NS4A to disconnect from the NS3 surface after adhesion has occurred.

The crowding effect consists of two somewhat counteracting forces - the excluded volume effect and non-specific interactions with crowder molecules. The effect caused by CG crowdres is a classic example of volume exclusion effect - the protein adopts a compact conformation and the whole complex is structurally more stable. PEG, on the other hand, induces effects characteristic for interacting crowdres - proteins surrounded by these polymers adopt more diverse conformations, in particular interactions are strong enough for the N-terminal tail to detach from the NS3 surface (Figures 4.18,

4.17).

While both forces are present in the case of proteins surrounded by polysucrose molecules, volume exclusion ultimately wins - the effect of these crowdors on the NS3/4A structure observed in the trajectories is comparable to the effect induced by the presence of non-interacting CG molecules.

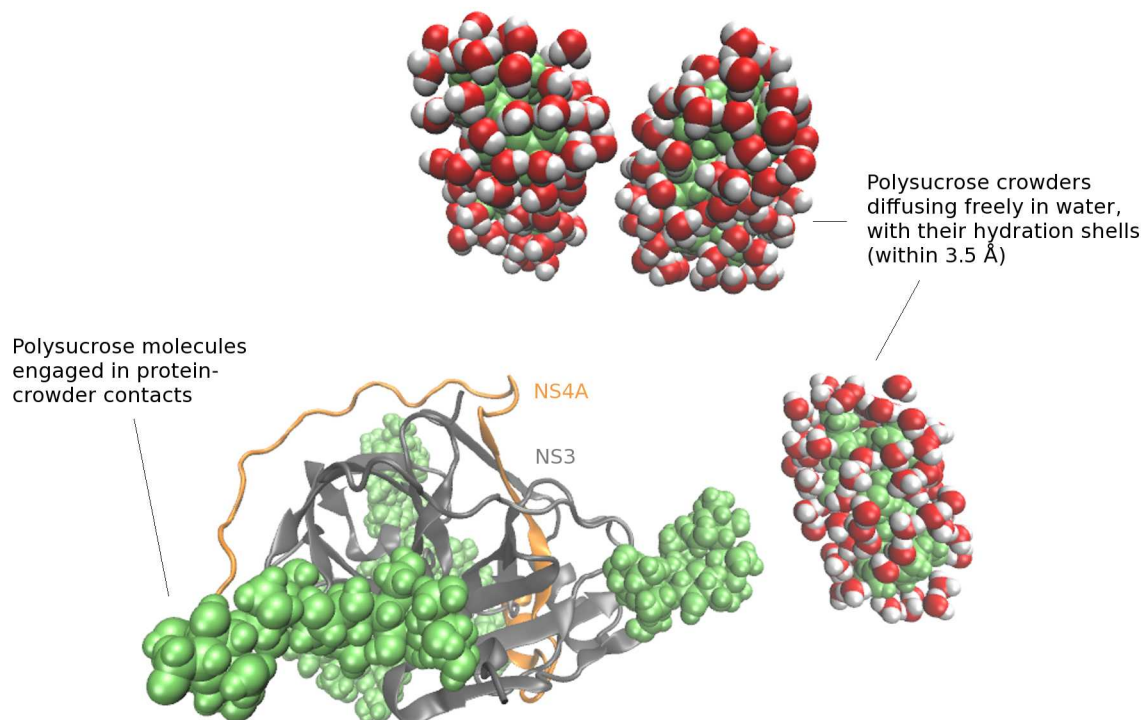


Figure 4.19: Illustration explaining the hypothesis that some polysucrose molecules behave in a similar way to non-interacting CG crowdors, despite the fact that protein-polysucrose interactions are possible and frequent.

Given the properties of polysucrose molecules, many of them may indeed behave similarly to CG crowdors:

- Polysucrose and CG crowdors diffuse much faster than PEG, with even faster diffusion of polysucrose (Table 3.3)
- Polysucrose molecules don't form crowder-crowder clusters, they almost lack long scale contacts (Table 3.4)
- Polysucrose crowdors are surrounded with 'dense' hydration layer (Figures 3.14, 4.19), which screens crowder-crowder attraction, and theoretically could screen interactions with the protein or existing protein-polysucrose complex.

In other words, some fraction of polysucrose crowdors could indeed act as inert crowdors, as they are referred to in literature^{113,6} - by limiting the volume available for the protein but having their charge-charge interactions screened by hydrodynamic layer.

This could explain the increased contribution of the excluded volume effect in the way polysucrose affects NS4A dynamics.

4.3 Crowding stabilizes a helix formed on the N-terminal tail of NS4A

The last aspect of conformational dynamics that was analysed in the trajectories obtained for the protease in water and in a crowded environment is the effect of crowding on the secondary structure of the protein, especially the NS4A cofactor. Variability in the secondary structure, particularly in the N-terminal tail of NS4A, plays an important role for the biological functions of the entire NS3/4A complex. In water this fragment remains unstructured, but upon the contact with host endoplasmic reticulum (ER) membrane, it is predicted to fold into a long transmembrane helix⁴⁰. Together with the 38-40 loop and the 12-24 helix of NS3 it forms a stable three-point contact interface anchoring the viral protease to the host ER, helping to establish a replication complex of the hepatitis C virus. The predicted helix is flanked on each side with typical membrane interface residues interacting with polar headgroups of the lipid molecules: Thr2 and Trp3 on one side, and Thr19, Thr20 on the other, with 16 residues interacting with the hydrophobic lipid tails inside the membrane. The presence of a transmembrane fragment within this region has been confirmed experimentally by addition of a glycosylation acceptor fragment to the N-terminus of NS4A⁴⁰.

In the trajectories obtained for the NS3/4A protease in explicit water, the N-terminal tail of NS4A remains unstructured, which is in line with the expected structure. Whereas in a crowded environment, the formation of a helical fragment was observed in one of the simulation copies with each type of the studied crowdors (Figure 4.20). The helical fragment formed within the region predicted to fold into the transmembrane helix, with a length between 4 and 10 residues; the helix with by far the greatest length formed in an environment crowded with polysucrose molecules. It is worth noting that control simulations also show helical structures forming within the N-terminus of NS4A, but these are significantly less stable and rapidly unfold, whereas in a crowded environment the helices are much more stable.

Importantly, helix formation was observed in each of the crowded environments tested, both in the presence of protein-interacting PEG and polysucrose polymers and in the presence of coarse-grained crowdors whose interactions with the protein were reduced to a minimum. This means that the helix at the N-terminus of NS4A is stabilised irrespective of the extent to which a given crowder interacts with the protein, and suggests that the main factor stabilising this structure is the limitation of the space available to the protein, i.e., the excluded volume effect plays a major role.

Simulations of the protease in various environments were performed in a small number of replicates, and only one instance of stable helix formation was obtained in each type of the crowded environment. It is therefore questionable whether the same results would be obtained with a larger number of trajectories. Therefore, to ascertain the reliability and statistical significance of the results obtained, a set of additional simulations were carried out. The peptide of the sequence representing the N-terminal tail of NS4A (amino acids 1 to 22) was simulated in water and in the presence of CG crowdors, starting from two conformations: entirely helical and extended (Figure 4.21). Based on statistical analysis, helix formation was found to be more likely in the presence of crowdors than

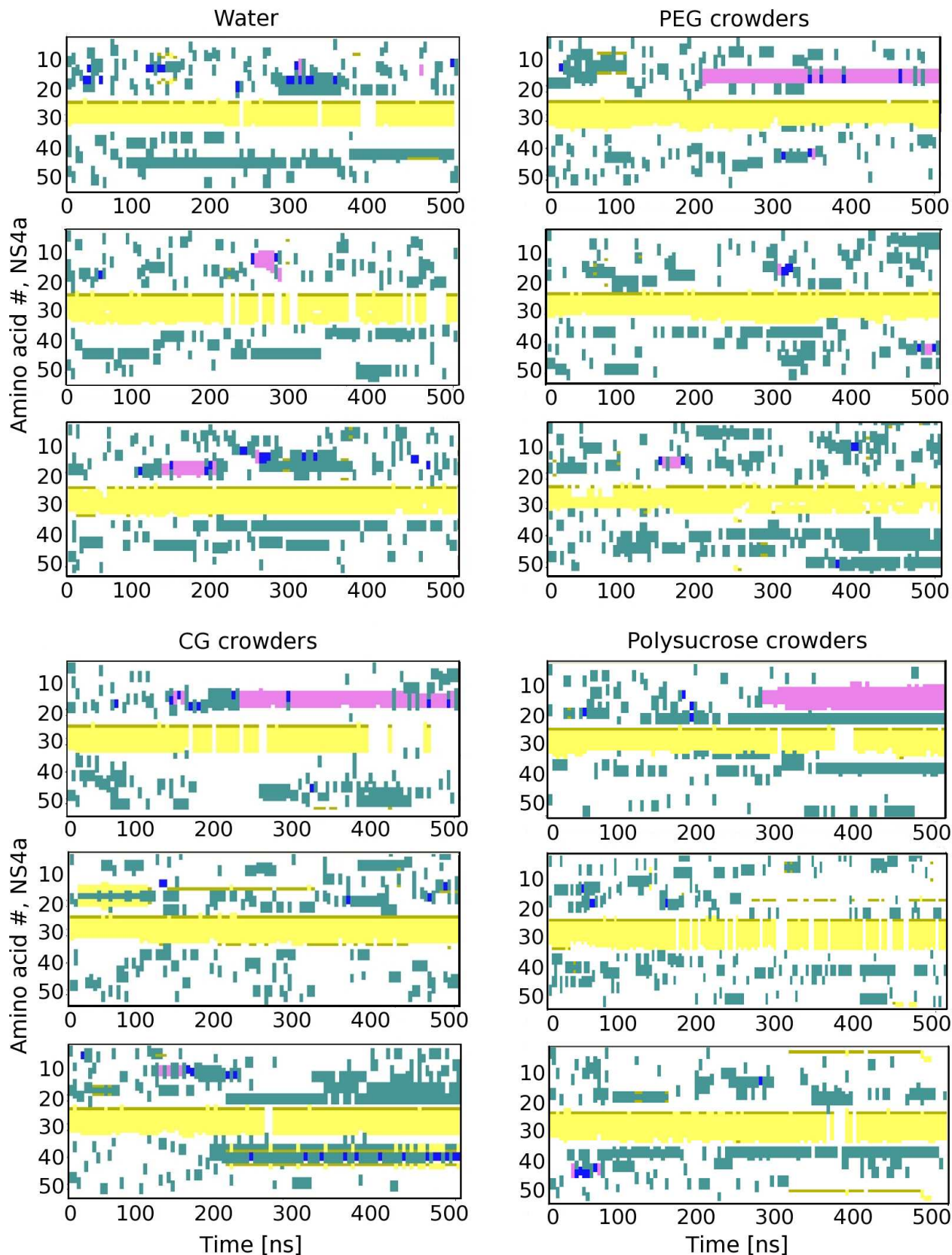


Figure 4.20: Secondary structure of the NS4A protein cofactor in water and in crowded environment. The helical fragments are marked with pink (α - *helix*) or blue (3-10 helix), β -strands in yellow and turns in green.

in water, with moderate statistical significance ($p=0.2$). However, completely in line with expected results, this simulation experiment showed that

- in water extended peptide does not fold into any stable secondary structure, and helical peptide unfolds in all simulation copies,
- in the crowded environment the helical peptide retains its structure, and the extended peptide folds into a helix in two out of three simulation copies.

The final step performed to test whether the crowded environment actually promotes the folding of the N-terminus of NS4A into a helical structure were experiments based on circular dichroism (CD). CD spectroscopy allows characterisation of the secondary structure of a protein or peptide by measuring to what extent each direction of circularly polarised light is absorbed by a sample. As proteins are chiral molecules, among other things, information about the content of helical structures can be read from this type of data.

CD experiments were performed on a peptide with a sequence corresponding to only the N-terminal tail of NS4A (residues 1 to 22), with *GGKK* solubility tags at each end (see Chapter 2, section 2.3.4). Spectra of a peptide representing this fragment of NS4A were previously recorded by Brass et al.⁴⁰ - to mimic the environment in the vicinity of ER membranes, they measured CD in the presence of sodium dodecyl sulfate (SDS) and dodecylphosphocholine (DPC) micelles. Their experiments showed increased content of helical structures of the peptide, regardless of the charge of the micelles - both with negatively charged SDS and neutral DPC.

The experiments I performed were carried out in water and in solutions containing crowdors corresponding to those included in the MD simulations: with Ficoll 400 and with PEG in two different sizes, including PEG 1200, whose mass is close to that of the PEG model used (Figure 4.22). The CD spectra confirmed the simulation results. For the peptide in water, a spectrum was obtained suggesting a mixture of different structures or the absence of secondary structure elements. However, in the presence of both Ficoll and PEG, the spectra contained characteristic minima at 208 and 222 nm, and a maximum around 192 nm. This shape of the spectrum corresponds to a peptide with a helical structure.

The obtained CD spectra confirm that crowding stabilises the helices formed on the N-terminal fragment of NS4A. This is the region of the protein that is responsible for anchoring the NS3/4A complex to the ER membrane through trans-membrane helix formation. Thus, the presence of a crowded environment may in some way support the function of viral proteins by facilitating trans-membrane helix (TMH) folding.

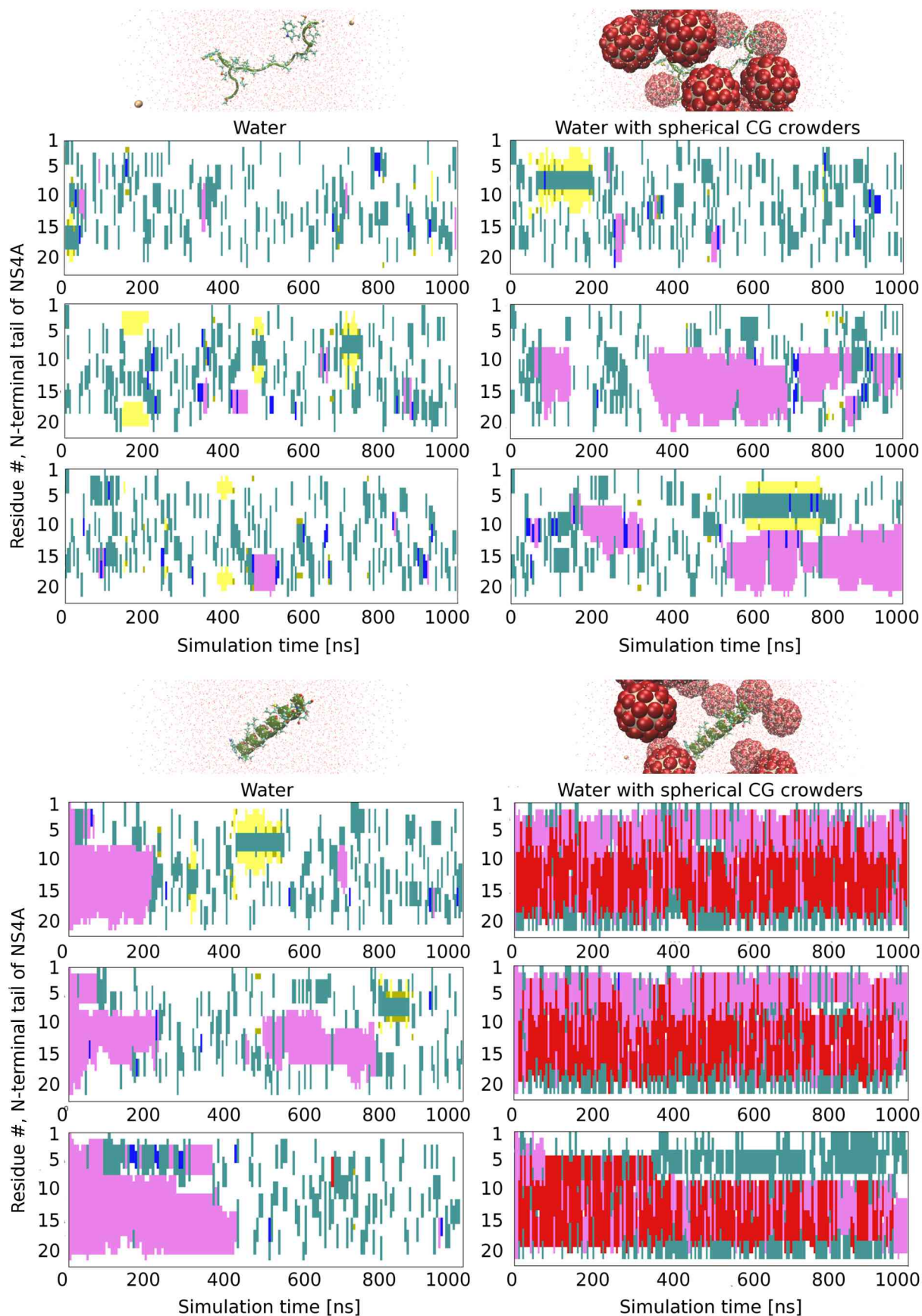


Figure 4.21: Secondary structure of the peptide representing the N-terminal tail of NS4A in various surroundings. Top: extended starting structures, bottom: helical starting structures. The helical fragments are marked with pink (α -helix) or blue (3-10 helix), β -strands in yellow and turns in green.

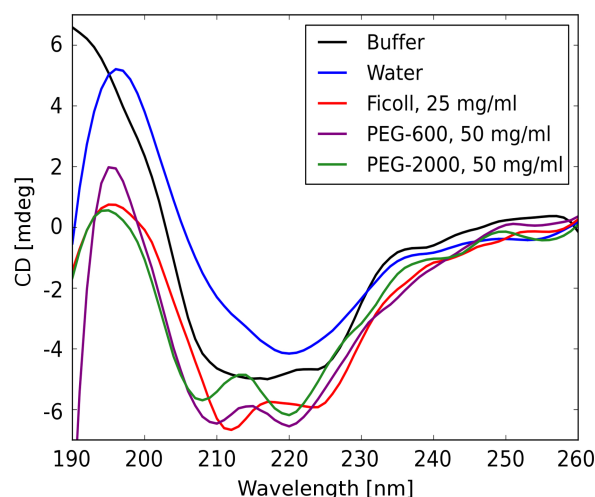


Figure 4.22: CD spectra of the N-terminal tail of NS4A obtained for the protein in water and in crowded environment. Figure from published work⁴⁴.

4.4 Crowding might contribute to membrane association of the NS3/4A complex

The helical fragments that folded into NS3/4A trajectories in a crowded environment are all located within the predicted transmembrane helix that anchors the protease to the ER membrane and is involved in the formation of the HCV replication complex⁴⁰. The observed α -helices contain amino acids that interact with polar lipid headgroups in the protease-membrane complex and several further hydrophobic residues. Considering the mechanism of complex formation it is very likely that these fragments function as precursors of the trans-membrane helix, and form just before insertion of the helix into the membrane. The hypothesis is supported by the fact that the amino acids comprising the predicted precursor are among the most conserved residues in the NS4A sequence⁴⁰, which suggests their important role in the functioning of the protein.

According to the work of Brass et al.⁴⁰, upon the contact with the ER membrane, the NS4A transmembrane helix undergoes spontaneous insertion preceded by an insertion of the 12-24 amphipathic helix of the NS3 protein. In the obtained trajectories, the helical fragments are located right next to this helix (Figure 4.23), which may suggest that as a precursor they interact with the membrane first. Given the sequence of the fragment which in each case is a part of the precursor - Ala-Leu-Ala-Ala-Tyr - its sequence allows it to coil in such a way that the resulting helix is also somewhat amphipathic, with the Tyr residue interacting with the polar groups and the hydrophobic alanines oriented towards the interior of the membrane. Following this mechanism, the rest of a helix would likely fold spontaneously on the membrane-water interface, as proposed in the study by Ulmschneider et al.¹³⁵. In their simulation of polyleucine, the peptide rapidly adhered to the membrane interface, and folded inside the membrane, spontaneously flipping to a conformation perpendicular to the membrane.

In summary, the proposed mechanism of TM helix insertion would consist of:

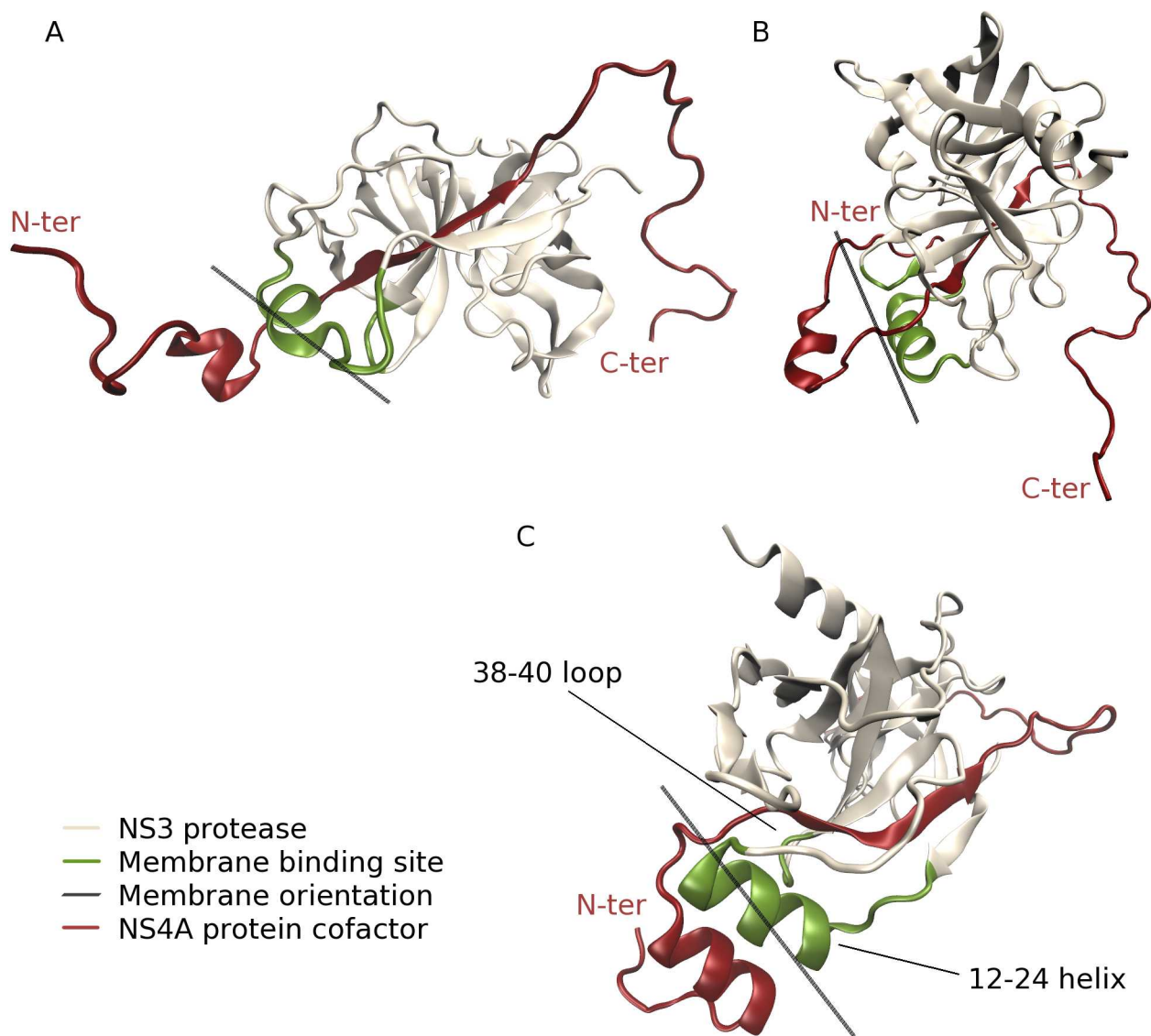


Figure 4.23: Helical structures formed in the trajectories of NS3/4A in the presence of A. PEG, B. CG crowdres and C. polysucrose crowdres. In each case the helix folded in the N-terminal tail of NS4A, in the region supposed to fold into a transmembrane helix upon the contact with the ER membrane of the host cell. Structures show that all helical fragments formed in a close vicinity of the NS3 fragments taking part in the membrane binding process: the 12-24 helix and 38-40 loop. Those fragments, together with the 3-19 TMH of NS4A form a three-point membrane contact interface, to stably anchor the complex to the membrane.

- Adsorption of the NS3 12-24 helix onto the membrane interface,
- Formation of the helix precursor consisting of at least 4 amino acids - NS4A:Ala12 to Tyr16,
- Insertion of the precursor following the mechanism similar to the amphipathic helix insertion,
- Spontaneous insertion of NS4A:1-11 residues and transition of the helix to the transmembrane conformation.

If indeed the presence of a crowded environment induces the formation of the transmembrane helix precursor, this would imply that crowding plays an important role in the biological function of NS4A. One of its main functions is to anchor the NS3/4A protease to the membrane of the endoplasmic reticulum, and the formation of the precursor could significantly assist in this process. Moreover, the positioning of the protease with the attached NS3 helicase domain is an essential step in the construction of the HCV replication complex, in which several proteins bind at the membrane interface and interact to produce elements of new virions. Taking this into account, it can be said that the presence of a crowded environment inside the cell may enhance the entire infection process.

4.5 Discussion

4.5.1 Test-tube versus MD, in numbers

The activity assay performed to study the effects of crowding on the reaction rate of NS3/4A peptide cleavage was carried out using Corning[®] Thermowell PCR 96 well plates, with 200 μ l wells, half-filled. The enzymatic reaction mixtures contained NS3/4A protease in the concentration of 2 nM, and the peptide substrates in the concentrations ranging from 2.5 to 40 μ M. This means that when EDANS fluorescence measurements were conducted, each well contained approximately 10^{11} protein molecules and 10^{14} to 10^{15} substrates.

Regarding the time-scales of the *in vitro* and *in silico* experiments, the activity assay with fluorescence measurements conducted every 50 s for two hours recorded data on a 14 billion longer fragment from the life of NS3/4A proteases than the molecular dynamics simulations, which generated only 500 nanosecond trajectories.

In the experiment, what we observe as results is the collective action of these hundreds of billions of proteins, whereas in the simulation we observe the dynamics of a single protein for a brief moment, but with the precision of single atoms. In the case of protein motions, time-scales of up to microseconds are sufficient to observe conformational changes, movements of smaller structural elements of the protein and to estimate diffusion rates of proteins and smaller molecules, thus allowing characterisation of processes that at the scale of molecules are responsible for changes in enzymatic activity. For example, movements of surface amino acid side chains take place on a scale of tens of picoseconds, loop movements on a scale of nanoseconds, and formation of secondary structure elements, like helical fragments, can take a minimum of tens of nanoseconds¹³⁶.

4.5.2 Explaining experimental results through MD simulations

In this thesis, I describe an attempt to represent the same experimental system - a protease in dilute solution and in a crowded environment - at two scales, in order to be able to look at how crowding affects both enzymatic cleavage of the peptides it catalyses, and the dynamic processes involving a single protein molecule. The two aspects are closely connected - changes in enzymatic activity result directly from alterations in the diffusion of the enzyme itself or its ligands, and from changes within the protein structure. This is why it is so valuable to be able to look at the experimental setup up close and analyse the molecule-scale causes of the phenomena observed in laboratory experiments.

A laboratory experiment in which the activity of NS3/4A protease was assessed in dilute buffer solution and in crowded environment showed that the presence of polysucrose enhances the NS3/4A-catalysed reaction and PEG slows it down⁹¹ (see Figure 4.2 and Table 4.1). As the analysis of the results of the experiment itself was not sufficient to investigate the reason for such a difference in crowder impacts, the possibility to zoom in and look at the protein dynamics in the surroundings of analogous crowdiers proved to be an invaluable help.

Observations of protein trajectories in different environments enabled a hypothesis to be made on the mechanism of lowering and increasing cleavage rates according to the following inference:

- The simulations excluded interactions of both PEG and polysucrose polymers with the protease active site - the crowdiers did not localize in the substrate binding site or influence its conformation
- The presence of PEG crowdiers slows down protease diffusion to a greater extent than the presence of polysucrose, PEG also significantly limits the diffusion of peptide substrates (Table 4.2 and 4.5)
- Polysucrose crowdiers have the highest translational diffusion coefficient of all studied crowdiers, moving faster even than spherical coarse-grained crowdiers (Table 3.3)
- Polysucrose, compared to PEG, interacts more readily with acidic residues of the protease (Table 4.6).

Based on these observations, I concluded that the most likely reason for the different effects of the presence of PEG and polysucrose crowdiers on protease activity is due to the different effects on diffusion of both the protein and its substrates. While PEGs significantly limit their diffusion, Ficolls represented by polysucrose in the simulations diffuse relatively fast, and they could in effect also increase diffusion rate of the substrates. Due to their high content of acid residues, the peptide substrates used in the activity assay have the potential to additionally form transient clusters with the Ficoll crowdiers.

The reasons for both the restriction of diffusion by PEGs and the likely acceleration of substrate diffusion by polysucrose are the very different physical properties of the crowdiers themselves. PEGs form a kind of network of variously shaped and flexible chains, while polysucrose forms more

globular structures that do not attract each other as PEGs do (Tables 3.2, 3.4); on the contrary, it is possible that they exhibit repulsive interactions with the other crowdiers.

The way to test this hypotheses is to conduct a simulation of a protease surrounded with Ficoll-like crowdiers and the peptide substrates - similar to this already performed for PEG crowdiers. Based the trajectories of such system, the effect of polysucrose crowdiers on the diffusion of peptide substrates can be measured by comparing the diffusion of substrates in water and in crowded environment. These simulations are planned within the project being a continuation of work described in this thesis. In addition to the introduction of substrates, the planned simulations will include an improved Ficoll model, in which sucrose molecules will be linked by glycerol linkers described with new force field patches.

4.5.3 PEG and Ficoll as crowdiers, in laboratory experiments and MD

Typically, in laboratory experiments containing a crowded environment, both PEG and Ficoll are used as crowdiers showing minimal interactions with the protein, and mainly occupying space around the protein limiting the room available to it. Ficoll is considered a more inert polymer, which in water forms spherical structures that do not interact with proteins^{113,6}. For PEG however, some protein-crowdier interactions have been detected; binding to hydrophobic amino acids was detected with experimental methods^{101,100}. In later studies, PEG molecules have been found to show size-dependent properties: longer PEG chains were shown to form prevalently hydrophobic interactions, whereas PEG of the mass around 8000 Da and smaller to also show electrostatic interactions and hydrogen bonds with protein residues^{104,108}.

My simulations of the NS3/4A protease surrounded with polysucrose showed that polysucrose crowdiers are able to make contacts and interact with the protein surface (Tables 3.4, 3.5 and 3.6). Moreover, they form more hydrogen bonds than PEG polymers. Despite the presence of numerous interactions, analyses of how polysucrose affects the protein dynamics show that the effects exerted by these crowdiers are much more similar to those of the non-interacting CG crowdiers than to those of interacting PEG molecules:

- Polysucrose restricts the translational diffusion of the protein to a lesser extent than the other crowdiers and has the least effect on its rotations. While theoretically interacting crowdiers should have a reducing effect on protein rotations - such effect is seen with PEG crowdiers, but to much lesser extent with polysucrose (Tables 4.2, 4.3 and 4.4).
- The presence of polysucrose affects the structure of the protease in the same way as the presence of CG crowdiers. The unstructured NS4A tails stick to the NS3 surface as in simulations with CG crowdiers, and the NS3/4A complex adopts the compact, globular conformation characteristic of proteins surrounded by non-interacting crowdiers. Polysucrose-protein interactions are insufficient to detach the NS4A tails (Figures 4.16, 4.17).

PEG, on the other hand, affects the NS3/4A structure similarly to what one would expect from protein crowders. Their presence makes NS4A conformations more diverse, and the unstructured N-terminal tail is able to detach from the NS3 surface after it comes into contact with it. This indicates that in the case of PEG crowders, the dominant effect is that caused by crowder-protein interactions - and not, as in the case of polysucrose, the excluded volume effect^{10,11}.

4.5.4 Stabilisation of the NS4A helix in a crowded environment

One of the most apparent effects of the presence of a crowded environment is the stabilisation of the helical fragment on the N-terminal tail of NS4A (Figure 4.20, 4.23). This fragment is located within a part of NS4A that remains unstructured in water, while in contact with the ER membrane it folds into a trans-membrane helix (TMH), containing the amino acids from Trp3 to Thr19⁴⁰. The helical fragment folding in crowded environment consists of at least four NS4A residues - Ala12 to Tyr16, which in the predicted TMH locates it within the membrane core, interacting with the hydrophobic lipid tails. The most common length of the observed helical fragment is 5 residues, but during some parts of the simulation with polysucrose crowders it reaches lengths of up to ten, which is half of the N-terminal tail of NS4A (Figure 4.20).

In water, brief folding of short helical fragments is also observed, but they are unstable and quickly unfold. In a crowded environment, however, a definite stabilisation of the helix occurs - the helical fragments that form in the presence of crowders maintain their structure for times on the order of hundreds of nanoseconds, often until the end of performed simulations (Figure 4.20). Stable helical fragments form with similar probability both in the presence of crowders that do not interact with the protein surface, and in the presence of interacting ones. For this reason, helix stabilisation is attributed to volume exclusion - in the case of the NS3/4A protease, it is sufficient to limit the space available to the protein to produce this kind of effect.

It may be concerning that helical fragments folded for all crowders in only one simulation, with the total number of simulation copies between three and five. I am not sure that with more simulation copies I would get a statistically significant difference between the degree of NS4A helicity in water and in the presence of crowders. However, there are significant indications that the observed helix formation was not a random event:

- Simulations of a peptide with the sequence of the NS4A N-terminus confirmed that the initially unfolded peptide folds into a helix in the crowder environment, while in water it remains unfolded. Additionally, the initially helical peptide undergoes unfolding in water, but not in the crowded environment (Figure 4.21).
- To confirm the results obtained from the simulations, a CD spectroscopy experiment was performed for peptides of a sequence representing the N-terminal tail of NS4A. The experimental conditions corresponded to those modelled in the simulations - CD spectra were recorded for the peptide in water, and surrounded by PEG and Ficoll crowders. The experiment showed

a clear increase in the degree of helicity in the presence of both PEG and Ficoll molecules, compared to the helicity in water (Figure 4.22).

4.5.5 The role of crowding in biological function of the protease - potential consequences for the viral life cycle

Given the spatial positioning of the helical fragment that has formed on the N-terminal tail of NS4A, it can be presumed that its location is not accidental, but has a specific significance for the biological function of the protein. The helix locates in the immediate vicinity of the other membrane-binding elements - the NS3:12-24 helix and the NS3:38-40 loop (Figure 4.23) - with which the NS4A:3-19 transmembrane helix forms a three-point contact site with the interface of host membrane⁴⁰.

The sequence of the helical fragment - Ala-Leu-Ala-Ala-Tyr - allows to obtain a kind of amphipathic helix, i.e. a helix which has hydrophobic amino acids on one side and polar amino acids (in this case one Tyr residue) on the other. This is a typical property of helices that come into contact with a lipid membrane parallel to its surface, so that hydrophobic residues insert into the membrane and polar residues interact with lipid head groups. This is the nature of helix 12-24 of the NS3 protease. This sequence lends credence to the model proposed in this work for anchoring TMH in the ER membrane, according to which the NS4A:12-16 fragment comes into contact with the membrane first, followed by binding of the the rest of the tail and spontaneous flipping across the membrane. According to this hypothetical model, the NS4A:12-16 fragment acts as a TMH precursor that initiates NS4A binding and facilitates the formation of a stable complex of the NS3/4A protease with the membrane (see section 4.4).

While attempting to establish the biological role of N-terminal NS4A binding to the membrane, one must try to reconstruct the chronology of events involving the entire protease (for detailed information on NS3/4A functions see section 1.4). It is known that the two-domain NS3 protein, when released from the polyprotein, first performs a protease function and then switches to a helicase function to perform its role in producing parts for new viruses³⁹. Initially, the N-terminal thirty amino acid fragment of NS3 remains unstructured, to fold into helix and beta-strand after forming a complex with NS4A⁴³. NS4A acts as a protease cofactor - NS3 alone is only able to cut the NS4A/NS4B junction, whereas the NS3/4A complex cuts also the NS4A/NS4B and NS4B/NS5A cleavage sites⁴³. The last information, which is not an experimental fact but a theoretical hypothesis, is that after binding to the membrane, the helical domain of NS3 rotates around the inter-domain linker, to move away from the membrane interface⁴⁰ - such movement would also expose the binding site for peptide substrates, which I noticed to be located at the meeting point of two domains, blocked by the helicase domain (Figure 4.24).

In such case, NS4A binding and TMH formation would both have to happen before cleaving the NS4A/NS4B, NS4A/NS4B and NS4B/NS5A junctions, and freeing NS4B, NS5A and NS5B proteins. Therefore, since crowded environment facilitates transmembrane helix formation, it also facilitates the formation of a complete replication complex of the HCV virus. Given these arguments,

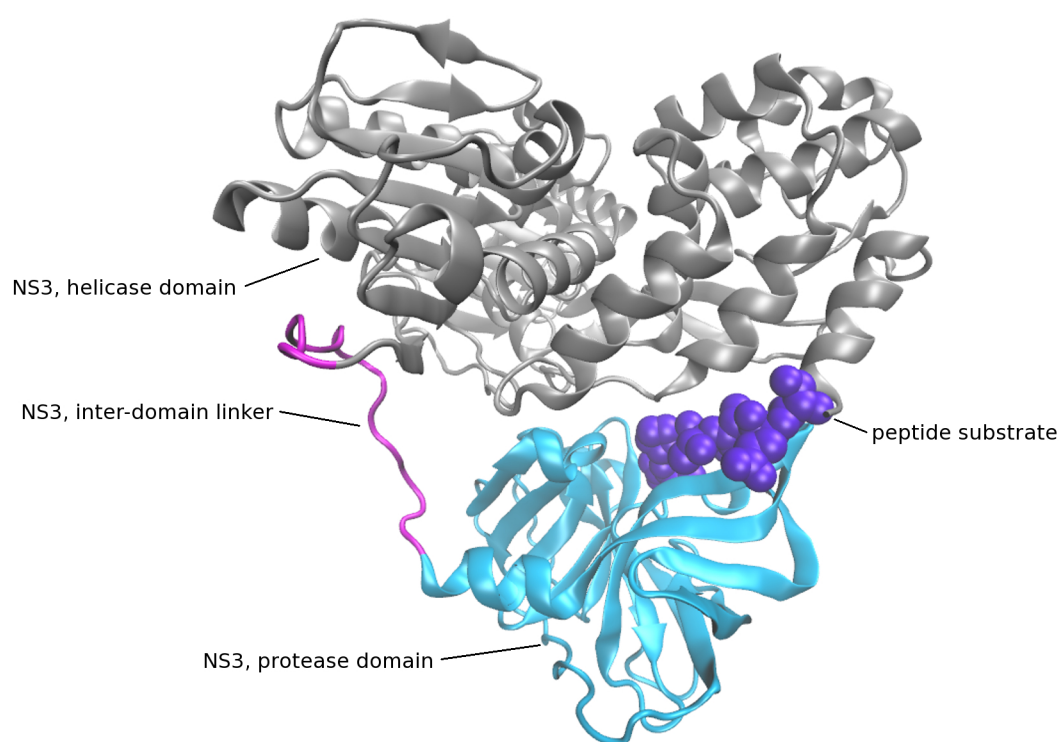


Figure 4.24: The full-length NS3 protease, composed of the protease domain (in blue), the helicase domain (in gray) and the linker between domains (in magenta). The protease is bound with a peptide substrate, shown in all-atom representation (in violet). The figure was prepared based on superposition of two experimentally-determined structures, using a model of a full length protease to show the location of helicase and linker domains (PDB:4A92), and the second structure to show the location of a protease domain bound with a peptide substrate (PDB:3M5N).

the presence of crowdors definitely works in favour of the virus.

Chapter 5

Summary and future perspectives

Studying the effects of molecular crowding on protein activity and dynamics is a field that is constantly evolving. We know that there are proteins that do not perform their catalytic function in water, recovering it only in the environment of crowders^{14,15}. There are also proteins that are not affected by crowding at all¹³⁷. The physical principles of the influence of crowding on macromolecules are known, and it is known that its effects result from a combination of volume exclusion and interactions with crowder molecules¹³⁷. However, it is still problematic to estimate how crowding will affect a given protein and whether it is worth conducting laboratory experiments in crowded solution, or whether a standard diluted buffer will suffice.

For such estimation to be possible, we need data on how crowding affects different types of proteins; we also need a deeper understanding of the mechanism of how crowders used in experiments affect macromolecules - knowledge about the properties of these crowders, and a decision on what model of crowded environment to choose for a particular experiment are crucial for the interpretation of its results. Research described in this thesis contributes to addressing these needs.

While studying the effects of molecular crowding on the NS3/4A protease from HCV, I attempted to combine observations from laboratory experiments and simulations, both carried out under similar conditions. In both cases, the protease was placed in water and in solutions crowded with non-protein polymers - PEG and polysucrose. These polymers are used as standard to mimic crowded environments in the laboratory^{6,46}, but they are extremely rarely incorporated into molecular dynamics simulations. In this work, however, they were also used in simulations, so that by observing the system on two scales it was possible to link the results of the experiment with their causes on a molecular scale. The experiment provides information about the activity of the enzyme and simulations about its internal dynamics and diffusion. Both complement each other and help to describe the enzymatic process in both dilute and crowded conditions. In the case of the protein studied here it was particularly important, since it turned out that the crowders used in the activity assay have different effects on the activity of NS3/4A - PEG causes a decrease in peptide cleavage rate, while Ficoll causes its increase⁹¹.

While performing the molecular dynamics simulations of the protease in crowded environment, I aimed to separate the observed effects into those caused by excluded volume effect, and those

caused by protein-crowder interactions. For this reason, three different models of crowders were used - all-atom PEG and polysucrose molecules, and the CG crowder with the interactions with NS3/4A reduced to minimum. Both polymers were found to interact with the protease, engaging in contacts and hydrogen bonds with the residues of NS3/4A. While it is commonly believed that PEG show minimal interactions with the proteins^{101,100}, and Ficoll is an inert crowder^{113,6}, in my trajectories polysucrose that was used to model Ficoll was found to form more than twice as many contacts and hydrogen bonds with NS3/4A than PEG. Nevertheless, it turned out that the effects of polysucrose on protein dynamics more closely resemble those caused by spherical CG crowders, while the way PEG affects it, is dominated by effects typical of interacting crowders, such as protein crowders.

Polysucrose molecules were found to diffuse very fast, faster even than the coarse-grained crowders, and to limit translational diffusion of the protease to a smaller extent than other crowders. They were also found to more readily bind acidic amino acids, which make almost a half of the residues of peptide substrates used in the activity assay. This set of results was used to explain the different effects of PEG and Ficoll on the rate of peptide cleavage reaction, that was observed in the laboratory experiment, according to which PEG slows down the reaction, and Ficoll speeds it up. Simulation results showed, that PEG significantly slows down the diffusion of both the protease and its substrates, limiting the frequency of enzyme-substrate contacts, whereas polysucrose has the potential to accelerate diffusion of substrates and increase this frequency. This hypothesis will be investigated in the planned continuation of the project.

Another important phenomenon observed in the simulations in water and with crowders is a stabilisation of helical fragment formed on the N-terminal tail of NS4A - the tail that folds into transmembrane helix (TMH) to anchor NS3/4A to the host ER membrane. Helix formation was observed in all performed simulations, including simulations in water, although without the crowders the helix quickly unfolds, while in the presence of all studied crowders it remains stable. Location and sequence of observed helical fragment suggest, that it could be a TMH precursor, forming to facilitate insertion of about 20 amino acids of NS4A into the membrane. This means, that molecular crowding inside a human cell might actually support the virus, by making it easier for it to construct a membrane-bound replication complex inside the human cell.

In conclusion, by studying the effects of crowding in simulation and in an experiment using the same crowder models, it was possible to propose the likely role of crowding in the HCV life cycle, and to discover unknown facts about polysucrose polymers, used as a representation of Ficoll which is one of the crowders most often used in laboratory experiments.

As the simulations with a simplified model provided interesting and promising results, the project will be continued to develop and parameterise an improved model of Ficoll, with branched structure and accurate representation of sucrose-sucrose linkages. Upgraded model will be used in a set of molecular dynamics simulations, studying the effects of Ficoll on the diffusion of NS3/4A and its peptide substrates, as well as the interactions and dynamics of the crowders themselves. The aim of those simulations will be to investigate how, despite the presence of interactions with the protein,

the effects of polysucrose resemble those of inert CG crowders more than those of interacting PEGs. New simulations will also test the hypotheses made on the basis of the simulations performed so far.

Bibliography

- [1] R. J. Ellis and A. P. Minton. Cell biology: Join the crowd. *Nature*, 425:27–28, 2003.
- [2] M. Feig, I. Yu, P. Wang, G. Nawrocki, and Y. Sugita. Crowding in cellular environments at an atomistic level from computer simulations. *J. Phys. Chem. B*, 121(34):8009–8025, 2017.
- [3] B10NUMB3R5. <https://bionumbers.hms.harvard.edu>.
- [4] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. Molecular biology of the cell, fifth edition. *Energetics of Biological Macromolecules Part B*, pages 123–136, 2008.
- [5] M. Tabaka, T. Kalwarczyk, J. Szymanski, S. Hou, and R. Holyst. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Frontiers in Physics*, 2:54, 2014.
- [6] I. M. Kuznetsova, K. K. Turoverov, and V. N. Uversky. What macromolecular crowding can do to a protein. *Int. J. Mol. Sci.*, 15(12):23090–23140, 2014.
- [7] G. Rivas, F. Ferrone, and J. Herzfeld. Life in a crowded world. *PNAS*, 5(1):23–27, 2004.
- [8] A. P. Minton. Molecular crowding: analysis of effects of high concentrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion. *Methods Enzymol.*, 295: 127–149, 1998.
- [9] K. A. Sharp. Analysis of the size dependence of macromolecular crowding shows that smaller is better. *PNAS*, 112(26):7990–7995, 2015.
- [10] G. B. Ralston. Effects of "crowding" in protein solutions. *J. Chem. Educ.*, 67(10):857, 1990.
- [11] M. Sarkar, C. Li, and G. J. Pielak. Soft interactions and crowding. *Biophys. Rev.*, 5(2): 187–194, 2013.
- [12] S. von Bülow, M. Siggel, M. Linke, and G. Hummer. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *PNAS*, 116:9843–9852, 2019.
- [13] G. Nawrocki, P.H. Wang, I. Yu, Y. Sugita, and M. Feig. Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J. Phys. Chem. B.*, 121:11072–11084, 2017.

- [14] S. B. Zimmerman and B. H. Pfeiffer. Macromolecular crowding allows blunt-end ligation by dna ligases from rat liver or escherichia coli. *PNAS*, 80(19):5852–56, 1983.
- [15] Y. Qu and D.W. Bolen. Efficacy of macromolecular crowding in forcing proteins to fold. *Biophysical Chemistry*, 101-102:5852–56, 2002.
- [16] J. Martin and F. U. Hartl. The effect of macromolecular crowding on chaperonin-mediated protein folding. *PNAS*, 94(4):1107–1112, 1997.
- [17] K. Maximova and J. Trylska. Kinetics of trypsin catalyzed hydrolysis determined by isothermal titration calorimetry. *Analytical Biochemistry*, 486:24–34, 2015.
- [18] K. Maximova, J. Wojtczak, and J. Trylska. Enzymatic activity of human immunodeficiency virus type 1 protease in dilute and crowded solutions. *Eur. Biophys. J.*, 48:685–689, 2019.
- [19] L. Stagg, S. Q. Zhang, M. S. Cheung, and P. Wittung-Stafshede. Molecular crowding enhances native structure and stability of alpha-beta protein flavodoxin. *PNAS*, 104:18976–18981, 2007.
- [20] M. S. Cheung, D. Klimov, and D. Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. *J. Phys. Chem. B*, 102:4753–4758, 2005.
- [21] G. Wieczorek and P. Zielenkiewicz. Influence of macromolecular crowding on protein-protein association rates-a brownian dynamics study. *Biophys. J.*, 95:5030–5036, 2008.
- [22] Y. C. Kim, R. B. Best, and J. Mittal. Macromolecular crowding effects on protein - protein binding affinity and specificity. *J. Chem. Phys.*, 133(20):205101, 2010.
- [23] Y. C. Kim, A. Bhattacharya, and J. Mittal. Macromolecular crowding effects on coupled folding and binding. *J. Phys. Chem. B*, 118:12621–12629, 2014.
- [24] R. Harada, N. Tochio, T. Kigawa, Y. Sugita, and M. Feig. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J. Am. Chem. Soc.*, 135(9): 3696–3701, 2013.
- [25] I. Yu, T. Mori, T. Ando, R. Harada, J. Jung, Y. Sugita, and M. Feig. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife*, 5:e19274, 2016.
- [26] A. Bille, K. S. Jensen, S. Mohanty, M. Akke, and A. Irböck. Stability and local unfolding of SOD1 in the presence of protein crowders. *J. Phys. Chem. B*, 123:1920–1930, 2019.
- [27] D. D. L. Minh, C. Chang, J. Trylska, V. Tozzini, and J. A. McCammon. The influence of macromolecular crowding on hiv-1 protease internal dynamics. *J. Am. Chem. Soc.*, 128: 6006–6007, 2006.

- [28] 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*, 28:303–325, 2009.
- [29] A. V. Fonin, A. L. Darling, I. M. Kuznetsova, K. K. Turoverov, and V. N. Uversky. Intrinsically disordered proteins in crowded milieu: when chaos prevails within the cellular gumbo. *Cellular and Molecular Life Sciences*, 75:3907–3929, 2018.
- [30] A. Soranno, I. Koenig, M. B. Borgia, H. Hofmann, F. Zosel, D. Nettels, and B. Schuler. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *PNAS*, 111(13):4874–4879, 2014.
- [31] V. NUversky, E. M. Cooper, K. S. Bower, J. Li, and A. L. Fink. Accelerated alpha-synuclein fibrillation in crowded milieu. *FEBS Letters*, 515(1-3):99–103, 2002.
- [32] F. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, and P. Selenko. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Nature*, 530:45–50, 2016.
- [33] J. Bai, M. Liu, G. J. Pielak, and C. Li. Macromolecular and small molecular crowding have similar effects on alpha-synuclein structure. *ChemPhysChem*, 18(1):55–58, 2017.
- [34] A. V. Predeus, S. Gul, S. M. Gopal, , and M. Feig. Conformational sampling of peptides in the presence of protein crowders from AA/CG-multiscale simulations. *J. Phys. Chem. B*, 116: 8610–8620, 2012.
- [35] A. Bille, B. Linse, S. Mohanty, and A. Irböck. Equilibrium simulation of Trp-cage in the presence of protein crowders. *J. Chem. Phys.*, 143:175102, 2015.
- [36] N. El-Hage and G. Luo. Replication of hepatitis c virus rna occurs in a membrane-bound replication complex containing nonstructural viral proteins and rna. *J. Gen. Virol.*, 84(10): 2761–2769, 2003.
- [37] L. Gao, H. Aizaki, JW. He, and M. M. C. Lai. Replication of hepatitis c virus rna occurs in a membrane-bound replication complex containing nonstructural viral proteins and rna. *J. Virol.*, 78(7):3480–8, 2004.
- [38] E. Blanchard and P. Roingeard. The hepatitis c virus-induced membranous web in liver tissue. *Cells*, 7(11):191, 2018.
- [39] S. M. Saalau-Bethell, A. J. Woodhead, G. Chessari, M. G. Carr, J. Coyle, B. Graham, S. D. Hiscock, C. W. Murray, P. Pathuri, S. J. Rich, C. J. Richardson, P. A. Williams, and H. Jhoti. Discovery of an allosteric mechanism for the regulation of hcv ns3 protein function. *Nat. Chem. Biol.*, 8(11):920–925, 2012.

- [40] V. Brass, J. M. Berke, R. Montserret, H. E. Blum, F. Penin, and D. Moradpour. Mechanistic role of ns4a and substrate in the activation of hcv ns3 protease. *PNAS*, 105(38):14545–50, 2008.
- [41] R. A. Love, H. E. Parge, J. A. Wickersham, Z. Hostomsky, N. Habuka, E. W. Moomaw, T. Adachi, and Z. Hostomska. The crystal structure of hepatitis c virus ns3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell*, 87(2):331–342, 1996.
- [42] O. Abian, S. Vega, J. L. Neira, and A. Velazquez-Campoy. Conformational stability of hepatitis c virus ns3 protease. *Biophys. J.*, 99(11):3811–3820, 2010.
- [43] H. Zhu and J.M. Briggs. Mechanistic role of ns4a and substrate in the activation of hcv ns3 protease. *Proteins*, 79(8):2428–2443, 2011.
- [44] N. Ostrowska, M. Feig, and J. Trylska. Crowding affects structural dynamics and contributes to membrane association of the ns3/4a complex. *Biophysical Journal*, 120(17):3795–3806, 2021.
- [45] B. D. Lindenbach, B. M. Pragai, R. Montserret, R. K. F. Beran, A. M. Pyle, F. Penin, and C. M. Rice. The c terminus of hepatitis c virus ns4a encodes an electrostatic switch that regulates ns5a hyperphosphorylation and viral replication. *Journal of Virology*, 81(17):8905–8918, 2007.
- [46] H-X. Zhou, G. Rivas, and A. P. Minton. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.*, 37: 375–397, 2008.
- [47] G. Rivas and A. P. Minton. Macromolecular crowding in vitro, in vivo, and in between. *Trends Biochem. Sci.*, 41:970–981, 2016.
- [48] A. Kessel and N. Ben-Tal. *Introduction to Proteins - Structure, Function and Motion*. CRC Press, 2018.
- [49] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, 337(3):635–645, 2004.
- [50] A. Tarakhovsky and R. K. Prinjha. Drawing on disorder: How viruses use histone mimicry to their advantage. *J. Exp. Med.*, 215(7):1777–1787, 2018.
- [51] M. Born and J. R. Oppenheimer. Zur quantentheorie der molekeln. *Ann. Physik*, 389(20): 457–484, 1927.
- [52] On the Quantum Theory of Molecules - English translation by S. M. Blinder. <http://www.chm.bris.ac.uk/pt/manby/papers/bornop.pdf>, 1998.

- [53] R. S. Mulliken. Spectroscopy, molecular orbitals, and chemical bonding. *Science*, 157(3874): 13–24, 1967.
- [54] R. Ditchfield, W. J. Hehre, and J. A. Pople. Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.*, 54(2):724–728, 1971.
- [55] N. Eddy, U. J. Ibok, and B. Ita. Qsar and quantum chemical studies on the inhibition potentials of some amino acids for the corrosion of mild steel in h₂so₄. *J. Comput. Methods Sci. Eng.*, 11(1-2):25–43, 2011.
- [56] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.
- [57] A. R. Leach. *Molecular Modelling: Principles and Applications, Second Edition*. Pearson, 2001.
- [58] 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.*, 79:926–935, 1983.
- [59] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *The Journal of Computational Physics*, 23(3):327–341, 1977.
- [60] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.
- [61] S. R. LaPlante, H. Nar, C. T. Lemke, A. Jakalian, N. Aubry, and S. H. Kawai. Ligand bioactive conformation plays a critical role in the design of drugs that target the hepatitis c virus ns3 protease. *J. Med. Chem.*, 57(5):1777–1789, 2014.
- [62] Uniprot, HC-J4 HCV strain. <http://www.uniprot.org/uniprot/O92972>.
- [63] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. Ucsf chimera visualization system for exploratory research and analysis. *J Comput Chem.*, 25(13):1605–1612, 2004.
- [64] D. S. Berkholz, M. V. Shapovalov, R. L. Dunbrack Jr, and A. P. Karplus. Conformation dependence of backbone geometry in proteins. *Structure*, 17(10):1278–1279, 2009.

- [65] A. Fiser, R.K. Do, and A. Sali. Modeling of loops in protein structures. *Protein Science*, 9: 1753–1773, 2000.
- [66] M. Shen and A. Sali. Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, 15:2507–2524, 2006.
- [67] A. Bakan, L. M. Meireles, and I. Bahar. Prody: Protein dynamics inferred from theory and experiments. *Bioinformatics*, 27(11):1575–1577, 2011.
- [68] W. Humphrey, A. Dalke, and K. Schulten. Vmd: visual molecular dynamics. *J. Mol. Graph.*, 14:33–38, 1996.
- [69] M. Feig, J. Karanicolas, and C.L. Brooks III. Mmtsb tool set: enhanced sampling and multi-scale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modeling*, 22(5):377–395, 2004.
- [70] R.B. Best, X. Zhu, J. Shim, P. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi1 and chi2 dihedral angles. *Journal of Chemical Theory and Computation*, 8:3257–3273, 2012.
- [71] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. De Groot, H. Grubmuller, and A.D. MacKerell. Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, 2016.
- [72] R. B. Best, N. V. Buchete, and G. Hummer. Are current molecular dynamics force fields too helical? *Biophysical Journal*, 95(1):PL07–L09, 2008.
- [73] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison II, and A. D. MacKerell Jr. Charmm additive all-atom force field for carbohydrate derivatives and their utility in polysaccharide and carbohydrate-protein modeling. *Journal of Chemical Theory and Computing*, 7:3162–3180, 2011.
- [74] E. P. Raman, O. Guvench, and A. D. MacKerell Jr. Charmm additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *Journal of Physical Chemistry B*, 114:12981–12994, 2010.
- [75] H. Lee, R.M. Venable, A.D. MacKerell Jr, and R.W. Pastor. Molecular dynamics studies of polyethylene oxide and polyethylene glycol: Hydrodynamic radius and shape anisotropy. *Biophysical Journal*, 95(4):1590–1599, 2008.
- [76] D. Petrov and D. Zagrovic. Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLOS Computational Biology*, 10(5):e1003638, 2014.

- [77] 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.*, 10: 5113–24, 2014.
- [78] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *Journal of Computational Chemistry*, 26:1781–1802, 2005.
- [79] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, 13(7):e1005659, 2017.
- [80] S.E. Feller, Y. Zhang, R.W. Pastor, and B.R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *J. Chem. Phys.*, 103:4613, 1995.
- [81] D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579, 1995.
- [82] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham. Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.*, 3(6):2312–2334, 2007.
- [83] I.C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B*, 108:15873–15879, 2004.
- [84] A. Ortega, D. Amoros, and J. García de la Torre. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.*, 101: 892–898, 2011.
- [85] V. Wong and D.A. Case. Evaluating rotational diffusion from protein md simulations. *J. Phys. Chem. B*, 112(19):6013–6024, 2008.
- [86] M. Linke, J. Köfinger, and G. Hummer. Rotational diffusion depends on box size in molecular dynamics simulations. *J. Phys. Chem. Lett.*, 9:2874–2878, 2018.
- [87] M. Roos, M. Ott, M. Hofmann, S. Link, E. Rössler, J. Balbach, A. Krushelnitsky, and K. Saalwächter. Coupling and decoupling of rotational and translational diffusion of proteins under crowding conditions. *J. Am. Chem. Soc.*, 138(32):10365–10372, 2016.
- [88] A. van Blaaderen, J. Peetermans, G. Maret, and J. K. G. Dhont. Long-time self-diffusion of spherical colloidal particles measured with fluorescence recovery after photobleaching. *J. Chem. Phys.*, 96:4591, 1992.

- [89] G. H. Koenderink, H. Zhang, D. G. Aarts, M. P. Lettinga, A. P. Philipse, and G. Nägele. On the validity of stokes-einstein-debye relations for rotational diffusion in colloidal suspensions. *Faraday Discuss.*, 123:335–354, 2003.
- [90] M. Taliani, E. Bianchi, F. Narjes, M. Fossatelli, A. Urbani, C. Steinkuhler, R. De Francesco, and A. Pessi. A continuous assay of hepatitis c virus protease based on resonance energy transfer depsipeptide substrates. *Analytical Biochemistry*, 240:60–67, 1996.
- [91] A. Popielec, N. Ostrowska, M. Wojciechowska, M. Feig, and J. Trylska. Crowded environment affects the activity and inhibition of the ns3/4a protease. *Biochimie*, 176:169–180, 2020.
- [92] I. Coin, M. Beyermann, and M. Bienert. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nature Protocols*, 2:3247–3256, 2007.
- [93] E. Kaiser, R.L. Colescott, C.D. Bossinger, and P.I. Cook. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.*, 34:595–598, 1970.
- [94] A. Kessel and N. Ben-Tal. *Introduction to Proteins - Structure, Function and Motion*. CRC Press, 2018.
- [95] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer, 2006.
- [96] Y. Wei, A. A. Thyparambil, and R. A. Latour. Protein helical structure determination using cd spectroscopy for solutions with strong background absorbance from 190-230 nm. *Biochim. Biophys. Acta*, 1844(12):2331–2337, 2014.
- [97] D. M. Rogers, S. B. Jasim, N. T. Dyer, F. Auvray, M. Réfrégiers, and J. D. Hirst. Protein helical structure determination using cd spectroscopy for solutions with strong background absorbance from 190-230 nm. *Chem*, 5(11):2751–2774, 2019.
- [98] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36:1627–1639, 1964.
- [99] C. Özdemir and A. Güner. Solubility profiles of poly(ethylene glycol)/solvent systems, i: Qualitative comparison of solubility parameter approaches. *Eur. Polym. J.*, 43:3068–3093, 2007.
- [100] D. L. Zhang, L. J. Wu, J. Chen, and Y. Liang. Effects of macromolecular crowding on the structural stability of human alpha-lactalbumin. *Acta Biochim. Biophys. Sin.*, 44:703–711, 2012.
- [101] P. B. Crowley, K. Brett, and J. Muldoon. Nmr spectroscopy reveals cytochrome-c-poly(ethylene glycol) interactions. *ChemBiochem*, 9:685–688, 2008.

- [102] E. L. Furness, A. Ross, T. P. Davis, and G. C. King. A hydrophobic interaction site for lysozyme binding to polyethylene glycol and model contact lens polymers. *Biomaterials*, 19:1361–1369, 1998.
- [103] S. Hirota, Y. Hattori, S. Nagao, M. Taketa, H. Komori, H. Kamikubo, Z. Wang, I. Takahashi, S. Negi, Y. Sugiura, M. Kataoka, and Y. Higuchi. Cytochrome c polymerization by successive domain swapping at the c-terminal helix. *PNAS*, 107:12854–12859, 2010.
- [104] L. Bekale, D. Agudelo, and H. A. Tajmir-Riahi. The role of polymer size and hydrophobic end-group in peg-protein interaction. *Colloids Surf. B Biointerfaces*, 130:141–148, 2015.
- [105] S. Lowe, N. M. O’Brien-Simpsonb, and L. A. Connal. Antibiofouling polymer interfaces: poly(ethylene glycol) and other promising candidates. *Polymer chemistry*, 6:198–212, 2015.
- [106] A. P. Chapman. Pegylated antibodies and antibody fragments for improved therapy: a review. *Advanced Drug Delivery Reviews*, 54(4):531–545, 2002.
- [107] D. B. Knowles, I. A. Shkel, N. M. Phan, M. Sternke, E. Lingeman, X. Cheng, L. Cheng, K. O’Connor, and M. T. Record Jr. Chemical interactions of polyethylene glycols (peg) and glycerol with protein functional groups: Applications to peg, glycerol effects on protein processes. *Biochemistry*, 54(22):3528–3542, 2015.
- [108] J. Wu, C. Zhao, W. Lin, R. Hu, Q. Wang, H. Chen, L. Li, S. Chen, and J. Zheng. Binding characteristics between polyethylene glycol (peg) and proteins in aqueous solution. *J. Mater. Chem. B*, 2:2983–2992, 2014.
- [109] A. N. Leonard, R. W. Pastor, and J. B. Klauda. Parameterization of the charmm all-atom force field for ether lipids and model linear ethers. *J Phys Chem B*, 122(26):6744–6754, 2018.
- [110] D. C. Latshaw, M. Cheon, and C. K. Hall. Effects of macromolecular crowding on amyloid beta (16-22) aggregation using coarse-grained simulations. *J. Phys. Chem. B*, 118:13513–13526, 2014.
- [111] H. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: Tip4p-ew. *Journal of Chemical Physics*, 120(20):9665–9678, 2004.
- [112] Namd wiki: Namdtroubleshooting. <https://www.ks.uiuc.edu/Research/namd/wiki/index.html>
- [113] S. Shahid, I. Hassan, A. Islam, and F. Ahmad. Size-dependent studies of macromolecular crowding on the thermodynamic stability, structure and functional activity of proteins: in vitro and in silico approaches. *Biochimica et Biophysica Acta*, 1861(2):178–197, 2017.

- [114] O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor, and A. D. MacKerell Jr. Additive empirical force field for hexopyranose monosaccharides. *J Comput Chem.*, 29(15):2543–2564, 2008.
- [115] O. Guvench E. Hatcher and A. D. MacKerell Jr. Charmm additive all-atom force field for aldopentofuranoses, methyl-aldopentofuranosides and fructofuranose. *Journal of Physical Chemistry B*, 113(37):12466–76, 2009.
- [116] X. Feng, A. J. East, W. B. Hammond, Y. Zhang, and M. Jaffe. Overview of advances in sugar-based polymers. *Polymers for Advanced Technologies*, 22(1):139–150, 2011.
- [117] S. Palit and A. Yethiraj. Dynamics and cluster formation in charged and uncharged ficoll70 solutions. *J. Chem. Phys.*, 147:074901, 2017.
- [118] N. Ostrowska, M. Feig, and J. Trylska. Modeling crowded environment in molecular simulations. *Frontiers in Molecular Biosciences*, 6:86, 2019.
- [119] X. Wang, J. Bowman, D. Nykypanchuk S. Tu, O. Kuksenok, and S. Minko. Polyethylene glycol crowder’s effect on enzyme aggregation, thermal stability, and residual catalytic activity. *Langmuir*, 37(28):635–645, 2021.
- [120] D. Liu, Y. Qiu, Q. Li, and H. Zhang. Atomistic simulation of lysozyme in solutions crowded by tetraethylene glycol: Force field dependence. *Journal of Molecular Biology*, 27(7):2110, 2022.
- [121] V. Nguemaha, S. Qin, and HX. Zhou. Atomistic modeling of intrinsically disordered proteins under polyethylene glycol crowding: Quantitative comparison with experimental data and implication of protein-crowder attraction. *J. Phys. Chem. B*, 122(49):11262–11270, 2018.
- [122] X. You, J. C. Shirley, E. Lee, and C. R. Baiz. Short- and long-range crowding effects on water’s hydrogen bond networks. *Cell Reports Physical Science*, 2(5):100419, 2018.
- [123] S. F. Sousa, J. Peres, M. Coelho, and T. F. Vieira. Analyzing pegylation through molecular dynamics simulations. *ChemistrySelect*, 3:8415–8427, 2018.
- [124] Y. Mao, X. Han, and Y. Zhang. Cryoprotective mechanism of using ficoll for cell cryopreservation at non-cryogenic temperatures: A molecular dynamics study. *International Journal of Heat and Mass Transfer*, 127(A):319–325, 2018.
- [125] H. Holter and K. M. Moller. A substance for aqueous density gradients. *Experimental Cell Research*, 15:631–632, 1958.
- [126] T. C. Laurent and K. A. Granath. Fractionation of dextran and ficoll by chromatography on sephadex-g-200. *Biochimica et Biophysica Acta*, 25732:191–198, 1966.

- [127] PubChem database. <https://pubchem.ncbi.nlm.nih.gov>.
- [128] D. Ridgway, G. Broderick, A. Lopez-Campistrous, M. Ru'aini, P. Winter, M. Hamilton, P. Boulanger, A. Kovalenko, and M. J. Ellison. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophys. J.*, 94:3748–3759, 2008.
- [129] I. Oh, S. Choi, Y. Jung, and J. S. Kim. Unusual size-dependence of effective interactions between collapsed polymers in crowded environments. *Soft Matter*, 10:9098, 2014.
- [130] C. M. Miller, Y. C. Kim, and J. Mittal. Protein composition determines the effect of crowding on the properties of disordered proteins. *Biophys. J.*, 111:28–37, 2016.
- [131] T. Vöpel and G. I. Makhatadze. Enzyme activity in the crowded milieu. *PloS One*, 7:e39418, 2012.
- [132] M. Rotta, L. F. Timmers, C. Sequeiros-Borja, C. V. Bizarro, O. N. De Souza, D. S. Santos, and L. A. Basso. Observed crowding effects on mycobacterium tuberculosis 2-trans-enoyl-acp (coa) reductase enzyme activity are not due to excluded volume only. *Scientific Reports*, 7(1), 2017.
- [133] W. Gtari, H. Bey, A. Aschi, L. Bitri, and T. Othman. Impact of macromolecular crowding on structure and properties of pepsin and trypsin. *Mater. Sci. Eng. C*, 72:98–105, 2017.
- [134] J. Batra, K. Xu, and HX. Zhou. Nonadditive effects of mixed crowding on protein stability. *Proteins Struct. Funct. Bioinf.*, 77:133–138, 2009.
- [135] M. B. Ulmschneider, J. P. Ulmschneider, N. Schiller, B. A. Wallace, G. von Heijne, and S. H. White. Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *J. Am. Chem. Soc.*, 138:10365–10372, 2016.
- [136] A. Kessel and N. Ben-Tal. *Introduction to Proteins - Structure, Function and Motion*. CRC Press, 2018.
- [137] K. Cheng, Q. Wu, Z. Zhang, G. J. Pielak, M. Liu, and C. Li. Crowding and confinement can oppositely affect protein stability. *ChemPhysChem*, 19:3350, 2018.