THESIS FOR THE DEGREE OF MASTER OF SCIENCE IN CHEMICAL ENGINEERING

A Molecular Dynamics Simulation Study of the Protein Stability of Reteplase in Formulation with Various Excipients during Freeze-Drying Processing

BY GABRIELLA BJÖRKENGREN

LUND UNIVERSITY

Supervisors: Pär Söderhjelm, Günther H.J. Peters, Suk Kyu Ko
Examiner: Kristofer Modig

Division of Biophysical Chemistry
Faculty of Engineering
Lund University

June 2021

PREFACE

This Master's Thesis Project was performed by Gabriella Björkengren, as part of exchange studies between the Faculty of Engineering at Lund University and the Technical University of Denmark. The project was supervised by Associate Professor Günther H.J. Peters and Ph.D. student Suk Kyu Ko at DTU Chemistry, and Associate Professor Pär Söderhjelm and Assistant Professor Kristofer Modig at the Division of Biophysical Chemistry, Lund University.

ABSTRACT

A Molecular Dynamics Simulation Study of the Protein Stability of Reteplase in Formulation with Various Excipients during Freeze-Drying Processing

The purpose of this project was to investigate protein stabilization by formulation excipients during the process of freeze-drying using Molecular Dynamics (MD) simulation. A recombinant therapeutic protein, Reteplase, was used as a model and the study focused on a comparative analysis of two excipients, arginine and tranexamic acid (TXA), as included in commercially available formulations of Reteplase. The simulation of freeze-drying was divided into five consecutive steps based on the conditions applied to the system: room temperature, freezing, primary drying, secondary drying and reconstitution at room temperature. Protein conformational changes and the existence of aggregation-prone regions (APRs) were investigated, as to assess the conformational stability during freeze-drying. The relevant stabilization mechanisms were analyzed with respect to protein-excipient interactions and the nature of preferential interaction. The study found a modest degree of conformational instability of Reteplase, which was mainly attributed to the tertiary structure level and effectively reduced in the presence of either arginine or TXA. In this regard, arginine was effective at a lower concentration than TXA. The stabilization mechanism was associated with preferential binding to exposed protein residues, mainly mediated via ionic interactions. A higher strength of interaction was observed for TXA,

which was suggested as a consequence of its high accessibility to the protein surface and its chemical character, which enables multiple types of interactions. A correlation was suggested between the stabilization of certain APRs and selective protein-excipient interactions. The results from a parallel study of another protein, Granulocyte Colony-Stimulating Factor (G-CSF), was used provide basic level support for the impact of protein surface charge on the proposed modes of stabilization. The conformational space of Reteplase was explored through the application of cluster analysis to Accelerated Molecular Dynamics (aMD) simulation, which provided further support for the good stability of secondary structure elements. The colloidal stability of Reteplase was investigated in a preliminary small-scale study using Coarse-grained (CG) MD simulation. Initial results suggest a tendency of arginine to reduce the strength of unique protein-protein interactions, as to prevent the experimentally observed aggregation-propensity of Reteplase.

TABLE OF CONTENTS

DEDICATION

This Master's Thesis Project is dedicated to my grandfather

Börje Georg Kvist (23 April 1925 – 6 September 2020),

who always encouraged me in my academic studies.

Chapter 1. INTRODUCTION

The development and production of protein-based biopharmaceuticals is a growing industry with a great potential to improve the specificity and safety of conventional small molecule drugs. However, many proteins suffer from an inherent instability in solution, which limits the shelf life of the pharmaceutical products. Over the last few decades, the process of freeze-drying has emerged as promising method with a pharmaceutical application in the manufacturing of solid-state biological therapeutics, which may significantly improve their stability, and thus the ease of storage and transportation (1).

Nevertheless, freeze-drying is a challenging process, which subjects the protein to various stress factors during the consecutive steps of freezing and dehydration. The low temperature applied, freeze-concentration effects on pH or the ionic strength, and the removal of water pose a risk for denaturation, which might trigger protein aggregation and ultimately cause destruction of the product. Reconstitution of the generated solid-state formula is necessary for administration and analysis during product development, which is a non-trivial task (2).

Pharmaceutical formulation design is a key concept in order to stabilize the protein against unfolding and aggregation. In this regard, excipients such as amino acids and sugars are commonly used for cryo- and lyoprotection of the protein through different stabilization mechanisms. The mode of stabilization is a complex balance, which is a function of the specific nature of the protein and the excipient, and of the physical state of the formulation (3). It has been proposed that the preferential exclusion of the excipient molecules from the protein surface is highly relevant for stabilization during freezing. In the dried state, hypotheses suggested for protection rely on the replacement of water molecules being removed by dehydration or the formation of a rigid glass matrix (4). In addition, selective protein-excipient interactions might play a crucial role in improving the stability of susceptible regions of the protein and in reducing the aggregation-propensity (5).

Within the framework of this project, a two-domain recombinant therapeutic protein, Reteplase, was used to study the protein stability aspects during freeze-drying using Molecular Dynamics (MD) simulation approaches. The project aimed to investigate the effect of formulation excipients on the conformational and colloidal stability of the protein, and to explore

the relevant stabilization mechanisms at a molecular level. A comparative analysis was conducted with respect to the performance of two different excipients, arginine and tranexamic acid (TXA), as included in commercially available formulations of Reteplase. To represent some characteristics of a virtually different and commonly used excipient, sucrose, a universal protein stabilizer, was adopted as a model example. The conformational stability was assessed at a secondary and tertiary structure level through analysis of some frequently employed protein structural descriptors, supplemented by graphical interface inspections, and investigations of aggregation-prone regions. The basis for protein stabilization by excipients was analyzed using quantitative and qualitative methods with respect to the frequency and type of protein-excipient interactions. Accelerated MD simulation was used to allow for a more extensive exploration of the protein conformational space. The colloidal stability with respect to protein-protein interactions was investigated in a preliminary study using coarse-grained (CG) MD simulation.

To provide basic support for the hypothesized stabilization mechanisms, some of the results were compared to a parallel study of another therapeutic protein, Granulocyte Colony-Stimulating Factor (G-CSF), which could serve as a starting point for further investigations within the subject.

Chapter 2. FREEZE-DRYING THEORY

2.1 THE PROCESS OF FREEZE-DRYING IN THE PHARMACEUTICAL INDUSTRY

Lyophilization (freeze-drying) serves as an important method in the production of biopharmaceuticals due to its advantages in improving storage time in terms of physical and chemical stability of colloidal systems (6). In this regard, lyophilization is used to produce a solid-state protein formulation through the steps of freezing the aqueous protein solution and subsequently drying the frozen product through accurate control of temperature, pressure and time (Figure 2.1). Through this process a protein powder is generated, which may be subjected to reconstitution before administration (2).

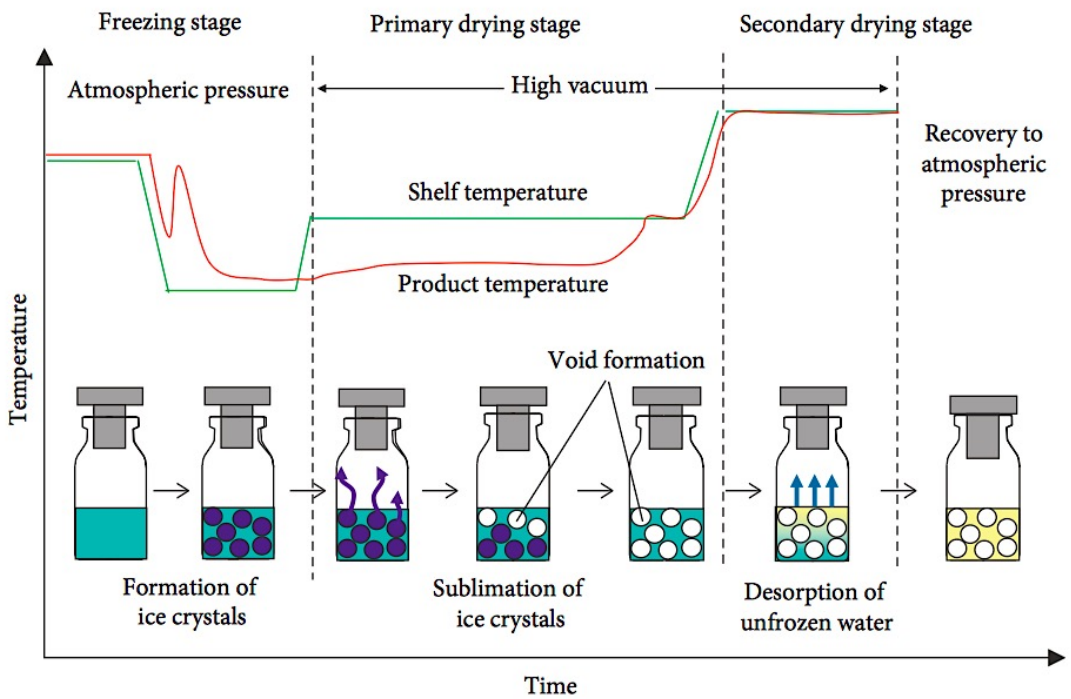


Figure 2.1. Schematic illustration showing the major steps of the lyophilization process: freezing, primary drying and secondary drying. During the freezing step, the temperature is decreased below the freezing point of ice. The formed ice crystals are removed by sublimation in the primary drying stage by applying a high vacuum and raising the shelf temperature. The non-frozen bound water is desorbed from the protein surface during the secondary drying phase by further increasing the temperature (2). The process steps are explained in more detail in section 2.2. Figure adopted from (7).

2.2 THE FREEZE-DRYING CYCLE

2.2.1 *Freezing*

During the freezing stage, the temperature is decreased to below the freezing point, typically using liquid nitrogen, and a supercooled state of the aqueous bulk solution is obtained. During ice nucleation, the product temperature is raised due to the release of the heat of crystallization (Figure 2.1). The growth of ice crystals is observed as the heat is gradually removed. Meanwhile, the protein concentration is increased in the remaining solvent (7). At the collapse temperature (T_c), the mixture is completely saturated with solute, and the protein will be retained as a non-crystalizing component in the surrounding ice matrix. T_c is closely related to the glass transition temperature of the formulation (T_g'), at which the mixture changes from a viscous liquid state into a brittle glassy state as the temperature is decreased further (8). The rate of freezing is another crucial factor, which affects the size, number and the surface area to volume ratio of the ice crystals that are formed. This has a great impact on the following steps of the freeze-drying process (9).

2.2.2 *Annealing*

During a subsequent annealing step, the temperature is kept above the freezing temperature but below T_g' of the formulation for a certain time to allow proper recrystallization of an excipient (9). Removing this component from the amorphous phase may increase T_g', which can provide an advantage to the efficiency of the drying process in terms of a higher maximal temperature and thereby less time needed (2). This thermal treatment has an additional benefit in removing the ice surface layer that might prevent effective water evaporation (9).

2.2.3 *Primary Drying*

The drying phase is conducted under an optimized degree of vacuum, and is divided into two steps: a primary and a secondary drying. During the primary drying, the pressure is lowered below the vapour pressure of ice and the shelf temperature is increased in order to remove the frozen water by sublimation (2). The ice-water interface is gradually moved from top to bottom when the generated water vapour disperses by diffusion through pores in the uppermost layer of

the dry structure. A rapid freezing rate results in many small ice crystals that form tight pores, which obstruct mass transfer of vapour and consequently slow down the primary drying step (9). To increase the rate of drying, the difference between the temperature of the product and the condenser should be maximized. However, it is essential that the temperature is maintained below T_c to avoid destruction of the product cake structure during manufacturing or storage (6). It has also been reported that product collapse might increase the propensity of protein aggregation due to increase in molecular mobility (8).

2.2.4 *Secondary Drying*

The secondary drying is initiated as the product temperature has reached the shelf temperature. The aim of the secondary drying is to remove the non-frozen bound water, which is approximately equal to the hydration shell of the protein. This step is conducted for a time period as to achieve the desired moisture level of the product, which should be optimized for protein stability (2). During the secondary drying, water is removed from the surface of the solute through desorption by further increasing the difference between the shelf temperature and the condenser temperature. In opposite to the primary drying, a high freezing rate is desired for optimal efficiency due to the increased surface area of the dry powder particles from which water is desorbed (9).

2.3 DENATURATION STRESSES

During the lyophilization process, the protein is subjected to numerous freezing and drying stresses, which may cause destabilization or unfolding. The mechanisms resulting in protein instability may be divided into two groups based on its origin: physical and chemical degradation. Physical degradation covers the processes by which the secondary and tertiary structure of the protein is altered in a way as to induce denaturation. Protein unfolding facilitates aggregation of monomer units, which ultimately may cause precipitation. In addition, protein unfolding can also cause the loss of activity. Chemical degradation pathways include reactions that modify the protein structure by formation or breakage of covalent bonds. Such processes might have a destructive effect on disulphide bridges, amide bonds or specific amino acid residues, and might initiate protein unfolding (10).

2.3.1 *Low Temperature Stress*

In terms of physical instability pathways, a crucial factor is cold denaturation due to low temperature stress, which is an enthalpy driven process in contrast to heat denaturation, which is governed by an increase in conformational entropy (11). During cold denaturation, exposure of non-polar residues optimizes the hydrogen bond network in the hexagonal ice (Ice I_h) phase. The corresponding decrease in enthalpy effectively counteracts the entropic penalty of inducing increased order to the water structure, and hence acts to disrupt the hydrophobic effect (12). The increase in hydration of non-polar groups weakens hydrophobic interactions within the core of the protein, which destabilizes the folded state (2).

2.3.2 *Formation of Ice-water Interface*

During lyophilization, the formed ice-water interface may decrease the free energy barrier of unfolding, which increases the rate of denaturation. Some studies suggest that the diffusional rate of water molecules in proximity of the ice surface is reduced, which enables an increased level of hydrogen bond formation to certain protein regions. Similarly to the process of cold denaturation, an increase in the hydration of the protein, and an accompanying disruption of secondary structure elements, results in a partial loss of the native protein fold (11). Elsewhere in literature, a direct surface adsorption of the protein onto the ice surface, resulting in a destructive conformational change, is proposed as a mechanism of unfolding (10).

2.3.3 *Freeze-concentration Effects*

Another important aspect is the gradual freeze-concentration of protein and other solutes. This event has an impact on the ionic strength of the solution and may further alter the composition of the formulation due to crystallization of certain excipients (2). During the freezing step, either component in a buffer system might be subjected to selective crystallization, which can cause a sudden change in the pH value and alter the protonation state of titratable groups in the protein. Such an event might have a detrimental effect on the stabilizing interactions within the protein that are essential for correct conformation, or might induce colloidal instability (2). Extreme pH conditions might pose a further challenge by accelerating chemical degradation by deamidation or methionine oxidation (10).

2.3.4 *Dehydration Stresses*

The drying process exposes the protein to several dehydration stresses. During this step, part of the hydration shell might be removed, which is important to protein stability and might shift the unfolding equilibrium away from the native state (10). Dehydration can also cause irreversible inactivation of proteins functioning as enzymes due to removal of crucial water molecules that are comprised in the active site (2).

2.4 THE ROLE OF EXCIPIENTS IN CRYO- AND LYOPROTECTION

Excipients might be added to the formulation in order to stabilize the protein to denaturation during the lyophilization process caused by stresses from freezing (cryoprotection) or dehydration (lyoprotection). The efficiency of different stabilization mechanisms is usually dependent on the nature of the specific protein, and has to be carefully investigated in each case for optimal performance (3) .

2.4.1 *The Stabilization Mechanism of Cryoprotection*

Considering cryoprotection, the stabilization mechanism is commonly explained by the theory of preferential exclusion. In accordance with this idea, the excipient is excluded to the bulk phase, and the local domain around the protein surface is preferentially hydrated by water molecules (Figure 2.2). This phenomenon leads to an increase in free energy for transferring the protein to the bulk environment due to the thermodynamically less favourable interactions with the excipient. The increase in free energy is greater for the unfolded state due to the larger surface area in contact with the solvent, which contributes to the stabilization of the native state (13). A crucial aspect in the case of cryoprotection is the preservation of an amorphous phase formed by the excipient in the freeze-concentrate of the protein. The amorphous matrix promotes close molecular contact, which is a prerequisite for protein-excipient interaction (14).

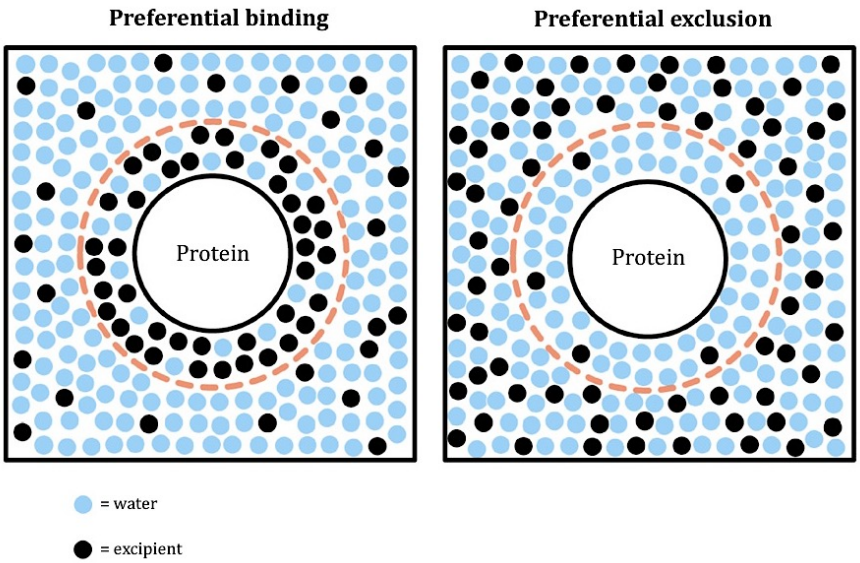


Figure 2.2. Schematic illustration showing the two concepts of preferential binding and preferential exclusion. The stabilization mechanism by excipient during cryoprotection is commonly explained by the theory of preferential exclusion, where the local domain (within red dashed line) around the protein surface is preferentially hydrated by water molecules and the excipient is excluded to the bulk domain (outside red dashed line). The opposite mechanism holds true for the theory of preferential binding (13). Figure adopted from (15).

2.4.2 *The Stabilization Mechanism of Lyoprotection*

The relevant stabilization mechanism of lyoprotection is frequently assigned to either of two principles: the water substitute hypothesis or the glass dynamics hypothesis (10). In the former case, the hydroxyl groups of the excipient molecules may form hydrogen bonds to residues located at the protein surface, and thereby act to replace the stabilizing water molecules that in part are removed during dehydration (Figure 2.3) (1). In this way, enhanced protein stability in conferred by increasing the free energy of unfolding by stabilizing the conformation of the native state (10).

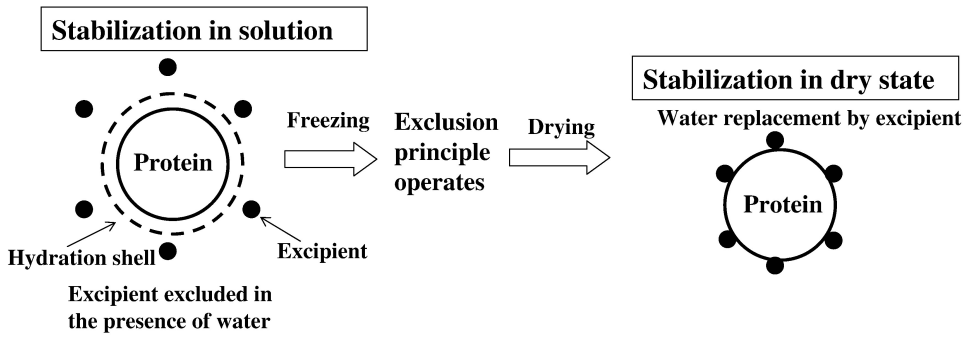


Figure 2.3. Schematic illustration showing the stabilization mechanism by excipient during the drying phase according to the water substitute hypothesis. In the dry state, the excipient acts to replace water molecules in the hydration shell that are removed during dehydration (1). Figure adopted from (15).

According to the glass dynamics hypothesis, increased protein stability is provided by the rigid amorphous glass matrix formed by the excipient, which prevents large-scale molecular motion, as required for complete unfolding. Thus, the glass dynamics hypothesis, which is also called the vitrification hypothesis, is based on a kinetic mechanism rather than on thermodynamical stabilization (10).

2.5 Case Studies of Stabilizing Excipients

Sugars and polyols are widely used as excipients, and improve protein stability in a non-specific manner by acting as both cryo- and lyoprotectants. Non-reducing disaccharides such as sucrose and trehalose are among the most common additives in this regard. Mannitol and sorbitol belong to the group of polyols that are also frequently used as stabilizers (2). Although sugars might confer lyoprotection through the water substitution hypothesis via the presence of multiple hydroxyl groups, preferential exclusion is suggested as the dominant stabilization mechanism (16).

Polymers might be applied as excipients due to their advantageous abilities in increasing T_g and the viscosity of the formulation (10). Polymers might also be used as surfactants in order to reduce protein adsorption at ice-water interfaces and to prevent protein-protein interactions, as involved in aggregation. It has been proven that certain amino acids such as arginine and histidine provide beneficial properties in terms of improved protein stability (10).

It is widely accepted that the action and performance of excipients vary for different proteins. Moreover, the applied physical and chemical conditions have an impact on the effectiveness of the formulation additives. Therefore, a thorough investigation of case studies is essential in terms of providing a knowledge basis for analysis and discussion of stabilization mechanisms.

The commercially available lyophilized formulations of Reteplase include either arginine or tranexamic acid as stabilizing excipients, as sold under the trade names of Retavase™ (17) and Rapilysin® (18), respectively. Therefore, it is logical to investigate and compare their effects on the stability of Reteplase during simulations of the freeze-drying process. Sucrose is a commonly used excipient in pharmaceutical formulations (2), and studies thereof are well documented in literature. However, its chemical nature differs from that of amino acids and their derivatives, and a comparison of characteristic features related to protein stabilization is interesting from an analytical point of view. Based on these arguments, case studies of arginine, tranexamic acid and sucrose as stabilizers are presented within this project. To the extent of existing documentation, hypotheses based on data from experiments as well as MD simulations, are covered. The molecular structures of the chosen excipients are displayed in Figure 2.4.

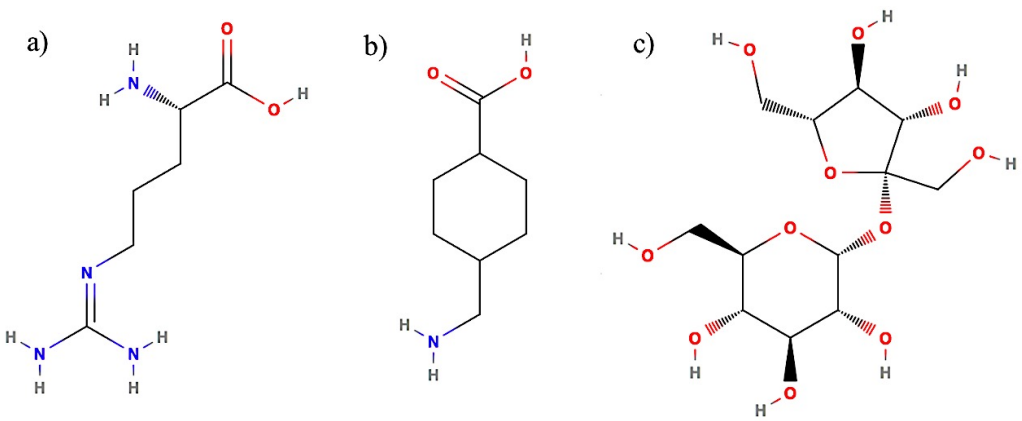


Figure 2.4. The molecular structures of a) L-Arginine, b) Tranexamic acid and c) Sucrose.

2.5.1 *Arginine*

Arginine has been reported to increase solubility of partially unfolded intermediates, and thus to prevent potential protein aggregation. According to current knowledge, this effect is not related to improved conformational stability of the native state (14).

In general, arginine is neither strongly bound nor excluded from the surface of the protein. The extent and strength of arginine-protein interaction relies on the overall composition of the formulation, the pH and the nature of the specific protein. However, studies have revealed an affinity for aromatic side chains via cation–π interactions. The guanidinium group of arginine has been associated to its ability to form clusters, especially around aromatic residues, thereby conferring stability against protein aggregation (15). Arginine may also act as a hydrogen bond donor, and its charged groups enables the formation of ionic interactions. The simultaneous repulsion of arginine from the protein surface due to the excluded volume effect accounts for its relatively low protein affinity. This dual behaviour is thought to be crucial for the ability to suppress protein aggregation (14).

Schneider and Trout studied the suppression of protein aggregation by arginine using vapour pressure osmometry in order to compute the preferential interaction (19). The study revealed a trend of increased preferential exclusion with increasing arginine concentration. The observation was proposed as a consequence of protein surface saturation (19). Shukla and Trout further investigated these results by performing molecular dynamics simulations of lysozyme and α-chymotripsinogen A in aqueous arginine solutions (20). It was proposed that the concentration of arginine in the local domain (Figure 2.2) relies on a relative dependence of the interaction of arginine with residues located at the protein surface compared to the tendency of arginine to self-assemble into clusters in the bulk phase. It was found that arginine binds favourably to charged and aromatic side chains, with the former type being the main contribution due to greater solvent-accessible surface area of charged residues. The authors suggested that the stronger interaction with the protein surface, as calculated for lysozyme, is a result of the greater fraction of exposed charged residues compared to α-chymotripsinogen A. It was concluded that at low arginine concentration, below 0.5 m, attraction to the protein surface dominates. At high concentrations, above 1.5 m, all the arginine-binding sites are occupied and there is a high probability of self-interaction in the bulk solution. Consequently, arginine molecules were excluded from the protein surface (20).

In a previous study, Shukla and Trout investigated the types of interactions of arginine in the local and bulk domains respectively, using a similar molecular dynamics simulation setup of α-chymotripsinogen A and melittin (21). They proposed multiple routes of arginine self-association to form clusters in the bulk phase. In this regard, the presence of the guanidinium group has an important role through its ability to interact with a functional group in another arginine molecule, either via hydrogen bonding to the carboxylate group or by guanidinium–guanidinium stacking (Figure 2.5) (21).

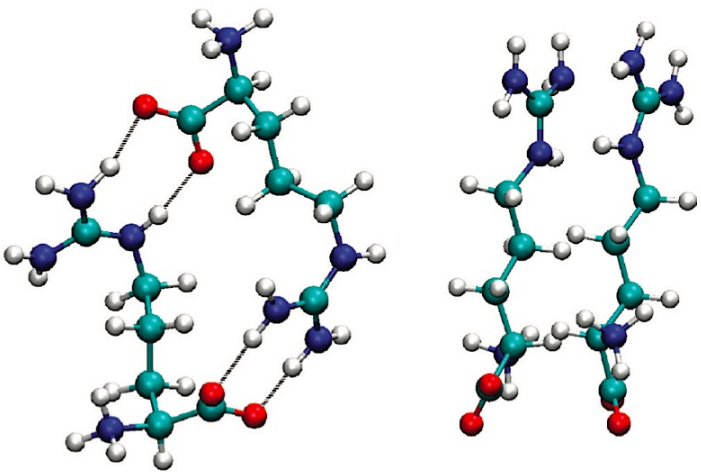


Figure 2.5. Snapshots from an MD simulation showing two possible configurations of arginine self-interaction: hydrogen bonding between the guanidinium group and the carboxylate group (left), and guanidinium stacking (right). Figure adopted from (21).

The interaction between the guanidinium groups of arginine has been investigated elsewhere (22) (23). It is suggested that the planar geometry of the moiety and its nonhomogeneous aromatic charge distribution contribute to the weak hydration of the guanidinium faces, where hydrogen bonding to water molecules is favoured only in the coplanar directions. In non-aqueous environment, the guanidinium dimer is unstable as a consequence of electrostatic repulsion between the like-charged cations. However, the observed parallel guanidinium–guanidinium stacking behavior is stabilized in a conformation where each cation is hydrogen-bonded to six water molecules, as illustrated in Figure 2.6. The stacked dimer formation requires the absence of any steric hindrance that would prevent the two moieties from approaching each other in the favoured relative positions (22).

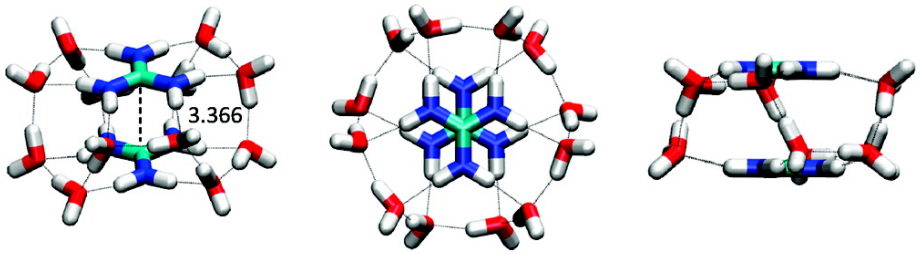


Figure 2.6. Schematic representation of the optimized stacked guanidinium dimer in a cluster with 12 water molecules (22). A top view of the conformation is shown in the middle of the figure and a side view is shown to the right. Figure adopted from (22).

Shukla and Trout found that the self-interaction of arginine is more enthalpically favourable than the interaction with water. There is also an entropic contribution through the release of the solvating water molecules. Thus, minimizing the surface area, which is exposed to the water phase, promotes cluster formation. It is proposed that the interaction with charged side chains of the protein is mainly due to hydrogen bonding to the guanidinium group. Furthermore, the study concluded that the limited degree of interaction with the protein prevents arginine from binding too strongly, and thereby acting as a denaturant (21).

2.5.2 *Tranexamic Acid*

Tranexamic acid (TXA) has a proven effect on increasing the solubility of folding intermediate structures (24). TXA is a cyclic Ω-amino acid, which is synthesized from L-lysine. It has been demonstrated that TXA improves the refolding yield of Reteplase to a greater extent than arginine. According to the suggested hypothesis, TXA stabilizes the native state of Reteplase via a high affinity interaction with a lysine-binding site located in the kringle-2 domain. The specific type of interaction is subject of controversy. Some studies suggest that the aliphatic part of TXA binds to a hydrophobic pocket, while others propose that the positively charged Lys40 residue interacts with the carboxyl group of TXA (24).

2.5.3 *Sucrose*

Bayat *et al.* investigated the stability of sc-tentceplase, which is a mutant of t-PA, in the presence of three different excipients: sucrose, trehalose and mannitol (16). The study revealed that the thermal stability was improved for all three systems, and to a greater extent for sucrose and mannitol. It was also shown that increasing the concentration within the range 50 – 200 mM, gradually increased protein stability. The secondary structure content was not significantly affected by the addition of excipient. However, a change in the tertiary structure, as characterized by a loosening of the protein, was observed. Molecular docking studies showed that hydrogen bonding and hydrophobic interactions constituted the major excipient-protein interactions. Some critical hot spot residues were identified. The catalytic domain and the kringle domains were particularly prone to interactions. Trehalose displayed the highest binding affinity. It was proposed that the greatest extent of preferential exclusion, and thus the least number of hydrophobic interactions, resulted in maximal stabilization of the native conformation (16).

Cloutier *et al.* performed molecular dynamics simulations to investigate the preferential interaction of sorbitol, sucrose and trehalose with various IgG1 antibodies (5). For all three excipients, negative preferential interaction coefficients were computed, as indicative of preferential exclusion from the local domain of the protein. To support the experimentally determined preferential interaction coefficients, partial charges and non-bonded parameters of the carbohydrate alcohol groups were adjusted according to the Kirkwood–Buff parameter set. They found that trehalose displayed the highest degree of preferential exclusion, followed by sucrose and sorbitol in decreasing order. It was proposed that this observation was a consequence of the greater hydrated volume of trehalose, which sterically restricts its proximity to the protein surface. The group also investigated the local excipient-protein interactions at solvent-exposed residues. They found that carbohydrates of suitable size are able to fit into hydrophobic pockets and exclude water molecules from these sites, or they may shield hydrophobic surfaces that are prone to aggregation. As a result, the excipients may reduce the solvent-accessible surface area of the protein by accumulating in certain regions of the protein, and thereby conferring improved stability. This effect was the most notable near aromatic and highly aggregation prone residues. In this regard, a reduction of the spatial aggregation propensity (SAP) score was used as a measurement of the stabilization induced by the excipient. Trehalose and sucrose had the greatest

impact on reducing SAP, as suggested due to the more pronounced higher than average concentration of these excipients in high-SAP residues (5).

Arsiccio and Pisano investigated the mechanisms providing improved stability during a molecular dynamics simulated freeze-drying process of two model proteins, human growth hormone (hGH) and lactate dehydrogenase (LDH), in the presence of some frequently used excipients, namely sucrose, trehalose, sorbitol, lactose, cellobiose and glycine (3). In the absence of excipient, the study revealed an increase in the non-polar surface area of the proteins during the freezing simulation, as indicative of partial unfolding. The authors related this observation to the denaturation stress induced by the existence of an ice-water interface. Specific sequences of the proteins displayed a significantly higher increase in the non-polar surface area, and these regions also demonstrated a higher root-mean-square fluctuation (RMSF), as suggesting a higher propensity of unfolding. The increase in the non-polar surface area was significantly reduced in the presence of excipient. The most effective stabilizers in this regard, sucrose and cellobiose in the cases of hGH and LDH respectively, were found to be the most preferentially excluded from the residues that were confirmed specifically prone to unfolding for each protein. The authors emphasize the importance of selective protein-excipient interactions for the purpose of optimizing stability (3).

It was observed that the drying simulation induced a greater degree of denaturation stress than the freezing simulation, as correlated to the crucial step of water removal during dehydration. For hGH, it was observed that the regions displaying an increase in non-polar surface area coincided with the locations of the α–helices of the protein. Therefore it was suggested that the preservation of secondary structure content might be an important parameter in identification of unfolding. It was futher proposed that the conformational change was connected to the formation of a protein-protein interaction, as the initial step of an aggregation process. The study revealed a clear difference in the ability of the excipients to act as either cryo- or lyoprotectants. Trehalose and cellobiose were found to be the most effective stabilzers of hGH in terms of lyoprotection. The authors explained this difference in performance by the unique stabilization mechanisms during freezing and drying respectively. In this study, sucrose was to great extent preferentially excluded from the protein surface, which promoted its ability to act as a cryoprotectant. However, trehalose has a higher propensity of forming hydrogen bonds, as to

replace the water molecules that are removed during dehydration, which explained its superior performance as a lyoprotectant according to the water substitute hypothesis (3).

In a subsequent study, Arsiccio *et al.* investigated cold denaturation of a model protein, peptostreptococcal protein L, at the ice-water interface by performing molecular dynamics simulations at freezing conditions, incorporating a metadynamics enhanced sampling technique, and using glucose as a cryoprotectant (11). In bulk water below the freezing point, the study found that the sampled fractions of partially unfolded conformations were significantly reduced in favour of the native conformation when adding glucose to the system. The free energy barrier of unfolding the α-helix of the protein was effectively increased and the β-sheet content was preserved to great extent. In contrast, when introducing an ice-water interface to the system without excipient, the partially unfolded conformations displayed a higher sampling and the free energy barriers of unfolding were reduced. However, in the presence of glucose, the authors observed the same stabilizing trend as for the bulk phase (11).

As a consequence of the enthalpic driving force of cold denturation, the conformational ensemble of the denatured state is composed of compact and not completely unfolded structures. Analysis of RMSD of the protein backbone revealed that certain sequences displayed an increased propensity of unfolding, and that the locations of these regions coincided with secondary structure elements. Arsiccio *et al.* proposed that direct adsorption of the protein at the ice-water interface is prevented by the non-frozen hydration shell. As reported previously, the authors instead correlated the destabilizing effect of the interface with the reduced diffusitivity of nearby water molecules, which promotes the formation of strong hydrogen bonds to the protein. The accompanying increase in hydration on non-polar groups accounts for the induced denaturation stress. The study revealed that glucose was preferentially excluded from hydrophobic unfolding prone regions, while it interacted more favourably with a hydrophilic part of the protein that demonstrated a higher RMSD in the presence of glucose, thereby slightly altering the native state. These specific protein-excipient interactions were suggested as the stabilization mechanism of glucose. In contrast, there were no proof of the excipient being adsorbed to the ice-water interface (11).

Chapter 3. THE BIOPHARMACEUTICALS OF INTEREST

3.1 RETEPLASE

3.1.1 *The Function of Reteplase*

Reteplase (r-PA) is a recombinant therapeutic protein, which is used in the treatment of acute myocardial infarction (AMI). It is a deletion mutant variant of tissue-type plasminogen activator (t-PA). In the human body, t-PA functions to catalyze the cleavage of plasminogen into plasmin (25). In the active form, plasmin acts to degrade the fibrin network of blood clots in plasma. This is the rate-limiting step in fibrinolysis (26). A simplified sketch of the fibrinolysis pathway is shown in Figure 3.1.

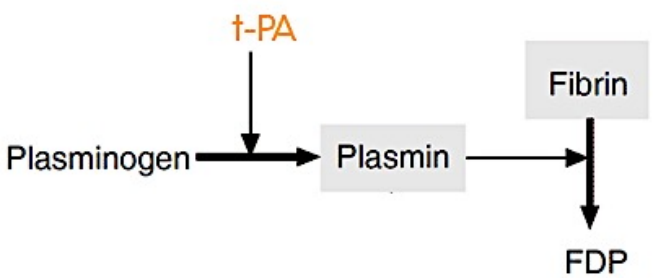


Figure 3.1. Simplified illustration of the fibrinolysis pathway. Tissue-type plasminogen activator (t-PA) catalyzes the rate-limiting step of plasminogen activation to form plasmin, which in turn cleaves fibrin into fibrin degradation products (FDP) in order to dissolve blood clots.

Over the past few decades multiple fibrin-specific thrombolytic agents have been launched to the market. Among these, Reteplase was the first next generation therapeutic to be approved. Reteplase exhibits several benefits over its parent drug t-PA such as a reduced rate of clearance and a prolonged half-life (25). Reteplase has a reduced binding affinity to fibrin compared to t-PA, which facilitates penetration into clots (27).

3.1.2 *The Structure of Tissue-type Plasminogen Activator*

Native t-PA is a serine protease consisting of five domains: an N-terminal fibronectin-like finger domain (1 – 49), an epidermal growth factor-like domain (50 – 87), two kringle domains (88 – 175 and 176 – 261) and a C-terminal trypsin-like serine protease catalytic domain (262 – 527) (Figure 3.2). Reteplase contains only the kringle-2 domain and the catalytic domain. In total, these two domains correspond to 355 amino acid residues. The protein structure of Reteplase includes nine disulphide bonds. Unlike t-PA, Reteplase is non-glycosylated (28). The structure of Reteplase is displayed in Figure 3.3.

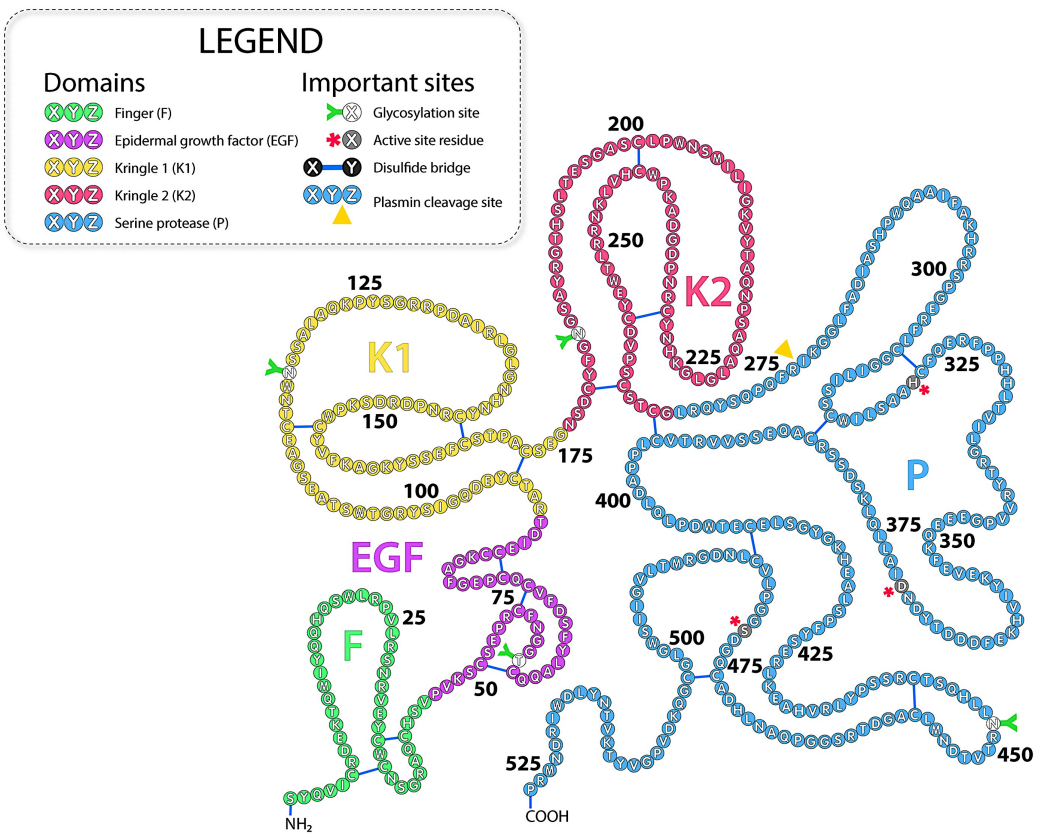


Figure 3.2. Schematic drawing of the primary structure of tissue-type plasminogen activator (t-PA). The illustration displays the five domains of t-PA: fibronectin-like finger domain (green, F), epidermal growth factor-like domain (purple, EGF), kringle-1 domain (yellow, K1), kringle-2 domain (pink, K2) and trypsin-like serine protease catalytic domain (blue, P). Disulphide bridges are indicated by blue lines. The active site residues are assigned by red stars. The plasmin cleavage site is displayed by a yellow triangle. Figure adopted from (29).

In the text further, all residue numbers are given for the relevant primary structure of Reteplase. t-PA is secreted as a zymogenic single-chain (sc) protein but might undergo cleavage by plasmin at the Arg103 – Ile104 peptide bond into an active two-chain (tc) form during attachment to the fibrin network. Thereof produced heavy and light chains are connected via an internal disulphide bond. tc-t-PA displays a tenfold increase in catalytic activity compared to sc-t-PA, which is considerably lower than for most other trypsin-like serine proteases (26).

3.1.3 *The Catalytic Domain*

The catalytic domain of Reteplase (Ser90 – Pro355) comprises a spherical structure and is composed of two six-stranded β-barrels, which are kept in position by three strap segments. Moreover, the catalytic domain contains several mechanistically important surface loops surrounding the active site and two α-helices, of which one is located in the C-terminal end (30). Charged residues form positively and negatively charged patches on the surface of the catalytic domain. The surface close to the active site is dominated by a negative potential (26).

The proteolytic activity of Reteplase is mediated via a catalytic triad consisting of His150, Asp199 and Ser306. These active site residues are located in a cleft in between the two β-barrels. In the active conformation of tc-t-PA, the ammonium group in the N-terminal end of the light chain is relocated to an activation pocket comprising the active site in the interior of the protein, where it forms a salt bridge to the carboxylate group of Asp305. In chymotrypsin, which belongs to the same family of serine proteases as t-PA, the side chain of the corresponding asparginine residue is rotated in a manner that opens a substrate specificity pocket in the active site cleft and a functional oxyanion hole (30).

It is assumed that the heavy chain portion of the loop region connecting the two domains of Reteplase, referred to as the activation peptide (Ser90 – Arg103), is stretched out along the surface of the catalytic domain in the active state, and that the rest of protein adopts the conformation of tc-t-PA (26).

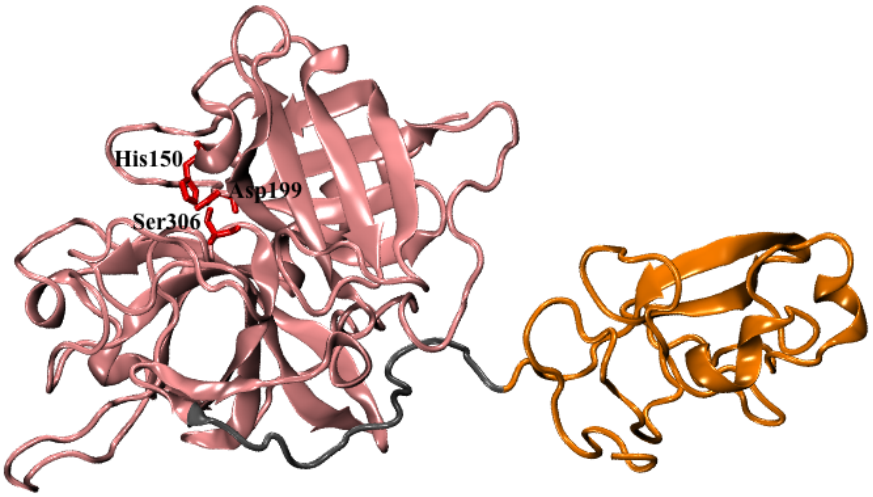


Figure 3.3. Visual representation of the structure of Reteplase, which was used as input structure for the MD simulations, here displayed in NewCartoon drawing style. The kringle-2 domain (Ser1 – Cys89) is coloured in orange. The catalytic domain (Ser90 – Pro355) and the incorporated activation peptide (Ser90 – Arg103) are coloured in pink and grey respectively. The residues of the catalytic triad (His155, Asp199 and Ser306), located in the activation domain, are highlighted in red and displayed in licorice drawing style. The model is derived from homology modeling using the structures of 1TPK chain B (kringle-2 domain) and 1BDA chain A (catalytic domain) as templates. The figure is rendered in VMD.

3.1.4 *The Kringle-2 Domain*

The kringle-2 domain of Reteplase (Ser1 – Cys89) is folded into two antiparallel β–strands, which are held together by two disulphide bridges. The kringle-2 domain and the catalytic domain are connected via a flexible linker segment composed of Ser90 – Thr91 (26). The kringle-2 domain contains a lysine-binding site located around Asp64, which binds to fibrin and thereby stimulates the catalysis of plasminogen cleavage (29).

3.2 GRANULOCYTE COLONY-STIMULATING FACTOR

The effect of formulation excipients in terms of improving conformational and colloidal stability during lyophilization, as investigated for Reteplase in this project, was compared to a parallel study of Granulocyte Colony-Stimulating Factor (G-CSF) performed by Ph.D. student Suk Kyu Ko.

3.2.1 *The Function and Structure of G-CSF*

G-CSF belongs to a family of glycoproteins called Colony-Stimulating Factors (CSFs), which have a crucial role as cytokines functioning in the proliferation and differentiation of hematopoietic precursor cells. In this regard, G-CSF is specifically involved in stimulation of the granulocytic cell lineage to produce mature neutrophils. G-CSF exerts its function through binding to a specific receptor (G-CSF-R), which initiates a signal transduction cascade into the cell nucleus (31). In its recombinant non-glycosylated form, G-CSF is used in the treatment of neutropenia (32).

Human G-CSF consists of 174 amino acid residues and contains two disulphide bonds, which are assumed essential to activity. The protein structure comprises an antiparallel four-α-helix bundle motif similar to growth hormone. The helices are commonly denoted by letters A-D. The only additional secondary structure elements include a 4-residue 3_{10}-helix followed by a 6-residue α-helix, both of them located within the AB loop, and short segment in the BC loop constituting approximately one turn of a left-handed helix (33). The structure of G-CSF is displayed in Figure 3.4.

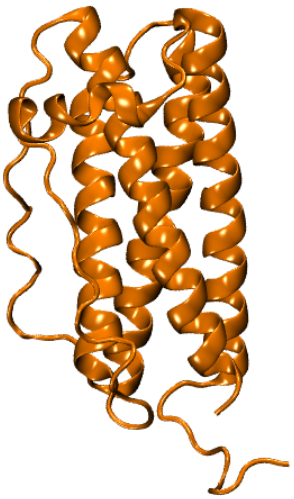


Figure 3.4. Visual representation of the structure of Granulocyte Colony-Stimulating Factor (G-CSF) (PDB ID: 1CD9 chain A), as displayed in NewCartoon drawing style. The figure rendered is in VMD.

3.2.2 *Case Study of the Stabilization of G-CSF by Excipients*

The commercially available formulation of recombinant G-CSF Filgrastim contains sorbitol, Tween-80 and sodium acetate buffer. G-CSF is commonly stored at low pH due to its extensive instability at physiological pH. The aggregation kinetics of G-CSF suggests partial unfolding of the protein monomer as the initial step of aggregation, followed by dimerization and ultimately oligomerization to form complex species (34).

Wood *et al.* investigated the stabilization to protein unfolding and aggregation of recombinant G-CSF (PDB ID: 2D9Q), as induced by formulation excipients, using the methods of *in silico* molecular docking and hydrogen deuterium exchange-mass spectrometry (HDX-MS) (32). The following excipients were included for study: arginine, glutamic acid, histidine, phenylalanine, mannitol, sorbitol, sucrose and trehalose. The short-term stability of G-CSF was assessed through studies of the thermal-unfolding transition midpoint (T_m) and the aggregation onset temperature (T_{agg}). When increasing the concentration of excipient, T_{agg} was observed to increase for the case of phenylalanine, as suggesting improved stability, while the opposite observation was made for the remaining amino acids. In this regard, the effect of sugar excipients was modest. For all amino acids except arginine, T_m was observed to reach a

maximum in the low concentration range. By contrast, arginine displayed a continuous decreasing trend. The sugars were shown to increase T_m up to a plateau at intermediate concentration (32).

The molecular docking studies revealed that sucrose and trehalose provided the strongest total binding energy, while sorbitol displayed the weakest interaction. The order of excipient binding energies correlated well with the observed ability of sugars to improve protein stability in terms of increasing T_m. Two interaction hotspots were identified, as located in surface pockets, which enables multiple protein-excipient interactions (32).

The deuterium exchange rates observed during the HDX studies indicated slow binding kinetics for the establishment of protein-excipient interactions. The interaction sites observed in HDX corresponded well to those predicted through the process of molecular docking. The results from using arginine as an excipient suggested a more rapidly established and global resistance to deuterium uptake, which in combination with its protein destabilization in terms of decreasing T_m, indicates the existence of multiple weak interaction sites. The authors propose that the low pH used in their study might have an influence in this regard due to the vastly neutralized protonation state of carboxylate groups, which might weaken surface interactions with arginine and prevent arginine self-assembly, as suggested a crucial factor to its stabilization by preferential exclusion (32).

In agreement with current research performed by others within this field, the authors emphasize the importance of selective protein-excipient interactions at hotspot sites located on the protein surface, in addition to the stabilization mechanism by preferential exclusion (32).

Chapter 4. MOLECULAR DYNAMICS SIMULATION THEORY

Molecular dynamics simulation is a powerful tool in the structural and dynamical study of biomolecular systems at the atomic level, which provides an important complement to experimental observations. MD simulation has the ability to capture biological processes and events, which occur on a wide range of time scales such as protein folding, transportation across ions channels or the local flexibility of individual amino acid side chains (35).

In the theory of MD simulation, the time-dependent progression of a system containing *N* atoms, which interacts via a potential energy function $U(\vec{r}_i)$, $i = 1, \ldots, N$, is simulated by simultaneously solving Newton's equations of motion for all atoms in three dimensions, as defined in equation (4.1), where m_i is the mass of atom *i* and $\ddot{\vec{r}}_i$ is the acceleration of that atom. The force acting upon each atom ($\vec{F}_i$) is computed using equation (4.2), where the atomic position as a function of time is denoted by $\vec{r}_i$ (36). An overview of the concept of numerical integration techniques to solve the system of equations is presented in section 4.2.

$$\vec{F}_i = m_i \ddot{\vec{r}}_i \quad (4.1)$$

$$\vec{F}_i = -\frac{\partial U(\vec{r}_i)}{\partial \vec{r}_i} \quad (4.2)$$

4.1 MOLECULAR MECHANICS

In the theory of molecular mechanics, each atom within a molecule is considered to be a sphere, and the intramolecular bonds are represented by springs. In the molecular mechanics model, the electrons are ignored. Due to the space-filling attribute of the atoms, their proximity in space is restricted, and any modification of the geometry of the molecule results in a change of the potential energy (37).

The total potential energy (U_{tot}) is defined as the sum of several terms, that each corresponds to a specific contribution to the overall energy, according to equation (4.3). U_{stretch} originates from stretching the bond length, U_{bend} is related to the change of the angle between two bonds, U_{torsion} refers to the torsional rotation around a single bond and $U_{\text{non-bond}}$ results from

intermolecular interactions. The set of mathematical expressions of these energy terms, as defined below, including their associated empirical parameters compose a force field (FF). The additions to the total energy are summed up over all covalent bonds, all angles between three sequential atoms, all dihedral angles defined by four sequential atoms and all pairs of interacting non-bonded atoms respectively (37). The visual interpretations of the energy contributions from bond stretching, angle bending and torsional rotation are illustrated in Figure 4.1.

$$U_{tot} = \sum_{bonds} U_{stretch} + \sum_{angles} U_{bend} + \sum_{dihedrals} U_{torsion} + \sum_{pairs} U_{non-bond} \qquad (4.3)$$

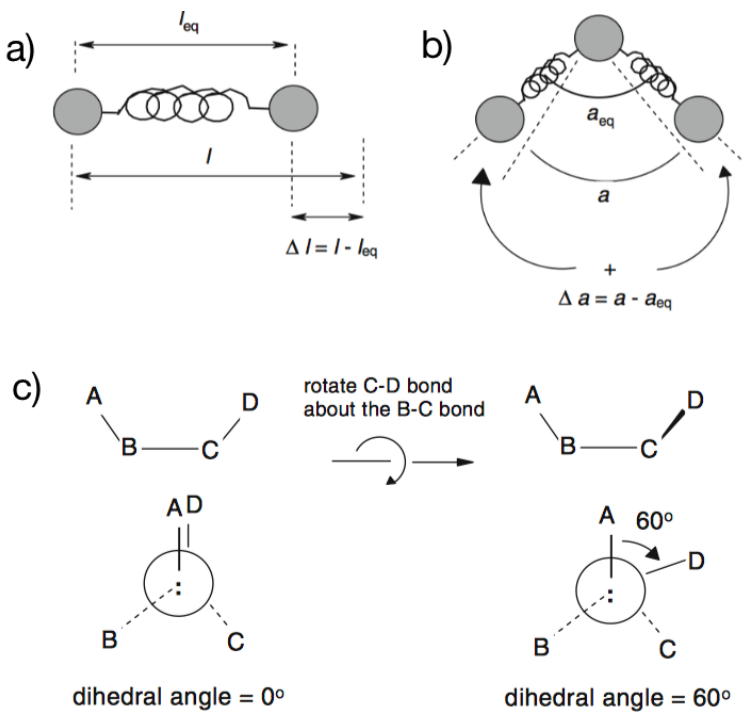


Figure 4.1. Visual interpretation of the different energy contributions to the total energy potential. A) The stretching of covalent bonds, as modelled by a harmonic spring of length *l*, which connects the two atoms being represented by spheres. B) The bending of bond angles comprising the angle *a* between two covalent bonds modelled by springs, which connects three sequential atoms represented by spheres. C) The torsional rotation of the C –D bond about the B – C single bond by the dihedral angle *Φ*, which is defined by the angle between the A – B bond and the C – D bond when inspected along the B – C bond in the system connecting four sequential covalently bound atoms A, B, C and D (37). Figure adopted from (37).

The bond stretching term is expressed by equation (4.4), where k_{stretch} is the proportionality constant, which equals half the force constant of the bond as modelled by a harmonic spring, l_{eq} is the equilibrium bond length and l is the stretched bond length (37).

$$U_{stretch} = k_{stretch}\left(l - l_{eq}\right)^2 \quad (4.4)$$

The angle bending term is expressed by equation (4.5), where k_{bend} is the proportionality constant, which equals half the angle bending force constant, a_{eq} is the equilibrium bond angle and a is the distorted bond angle (37).

$$U_{bend} = k_{bend}\left(a - a_{eq}\right)^2 \quad (4.5)$$

The torsional terms is expressed by equation (4.6), which is a cosine function where k_{torsion} is the energy barrier to rotation about the dihedral angle Φ, m is the periodicity and Φ_{offset} is the offset angle of the minimum energy for the staggered conformation (38).

$$U_{torsion} = \frac{1}{2}k_{torsion}\left[1 + cos\left\{m\left(\Phi + \Phi_{offset}\right)\right\}\right] \quad (4.6)$$

The non-bonded interaction term is expressed in equation (4.7) as a combination of the van der Waals interaction U_{vdW}, as represented by the Lennard-Jones potential in equation (4.8), and the electrostatic interaction U_{el} according to Coulomb's law in equation (4.9), which are valid for two atoms separated by a distance r (38). In the Lennard-Jones potential, ε corresponds to the depth of the energy minimum, and σ is the equilibrium distance where the potential equals zero (39). In Coulomb 's law, q_{A} and q_{B} are the partial charges of the two atoms, and ε_0 is the dielectric constant (38).

$$U_{non-bond} = U_{vdW} + U_{el} \quad (4.7)$$

$$U_{vdW} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad (4.8)$$

$$U_{el} = \frac{q_A q_B}{4\pi\varepsilon_0 r} \quad (4.9)$$

4.2 NUMERICAL INTEGRATION TECHNIQUES

The total potential energy is computed at all spatial positions (3*N*) of the complete three-dimensional system containing *N* number of atoms. A numerical integration technique is required to solve the set of Newton's equations of motion due to the lack of an analytical solution for this complicated system. There are multiple numerical algorithms available, which are based on the concept of repeating computation of the atomic forces using the defined FF in order to solve for the positions and velocities of atoms for each cycle corresponding to a short discrete time step δt. The initial atomic positions and velocities have to defined prior to the first computational cycle, as typically obtained from the protein crystal structure and the Maxwell-Boltzmann distribution of velocities at the relevant temperature, respectively (39).

In this context, the leap-frog algorithm will be presented. For this method, the velocities (v) are calculated at a time point $t + \delta t/2$ using the velocities at time $t - \delta t/2$ and the forces ($\vec{F}_i$) at time t, according to equation (4.10). Subsequently, these velocities, and the positions (r) at time t, are used to update the positions at a later time point $t + \delta t$ according to equation (4.11). The velocitites at time t are calculated using equation (4.12) (36).

$$\vec{v}_i\left(t + \frac{\delta t}{2}\right) = \vec{v}_i\left(t - \frac{\delta t}{2}\right) + \frac{1}{m_i}\vec{F}_i(t)\delta t \quad (4.10)$$

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \vec{v}_i\left(t + \frac{\delta t}{2}\right)\delta t \quad (4.11)$$

$$\vec{v}_i(t) = \frac{1}{2}\left[\vec{v}_i\left(t - \frac{\delta t}{2}\right) + \vec{v}_i\left(t + \frac{\delta t}{2}\right)\right] \quad (4.12)$$

4.3 Boundary Conditions

When setting up the simulation system, the boundary conditions at the edges of the simulation box have to be defined. In simulations of biomolecular systems, periodic boundary conditions (PBC) are frequently employed to avoid surface artifacts (35). Through the application of PBC, the central simulation box is replicated infinitely in all spatial directions. Thus, when a molecule leaves the simulation box at one edge, it will reappear on the opposite face via its periodic image box (36), as illustrated in Figure 4.2. To avoid interaction between the periodic copies of a particle, the non-bonded interactions have to be smoothly cut off at a specified distance. One such method relies on the Particle-Mesh-Ewald (PME) summation, where electrostatic interactions are divided into short-range and long-range interactions. The latter group of interactions is assigned through a charged grid, which is computed as a reciprocal sum using Fourier transformation to allow for rapid convergence in the associated space (35).

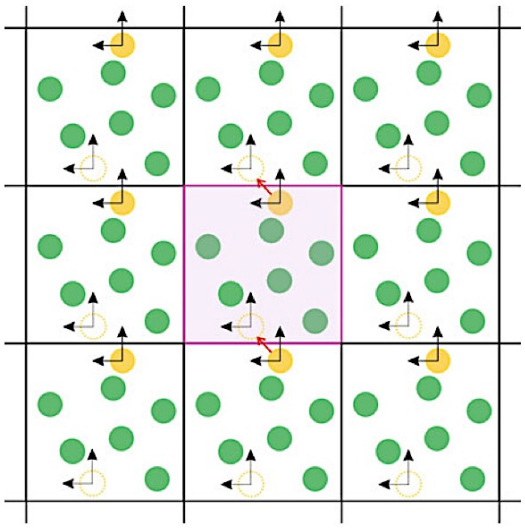


Figure 4.2. Schematic illustration of the concept of periodic boundary conditions in two dimensions. The yellow particle leaving the central simulation box will enter on the opposite edge from the periodic image box (36). Figure adopted from (36).

4.4 THERMODYNAMICAL ENSEMBLES

A statistical ensemble comprising a defined set of fixed thermodynamical variables is used to describe the conditions at which the system is simulated. The method of ensemble averages connects the equilibrium physical observables to the collection of individual microscopic configurations. Common thermodynamical ensembles include the micro-canonical (*NVE*) ensemble, the canonical (*NVT*) ensemble and the isothermal-isobaric (*NPT*) ensemble. In the *NVE* ensemble, the number of particles (*N*), the volume (*V*) and the total energy (*E*) are kept constant. In the *NVT* ensemble, the energy is allowed to change through exchange with an external imaginary heat bath, which is coupled to a thermostat algorithm of choice, in order to keep the temperature (*T*) fixed. In the *NPT* ensemble, the volume of the system is able to change by the movement of an imagined piston that function to maintain the pressure (*P*) at a constant level through the coupling to a barostat algorithm (39).

In the simulations of the freeze-drying process investigated in this project, the Langevin thermostat and the Monte-Carlo (MC) barostat were employed for control of the temperature and pressure respectively. The corresponding algorithms are further explained in sections 4.4.1 and 4.4.2.

4.4.1 *The Langevin Thermostat*

A thermostat algorithm is a modification of Newton's equation of motion as to obtain a constant temperature thermodynamical ensemble through the coupling of the system to an external heat bath at a specified reference temperature, with which it may exchange energy. The temperature of the system depends on the velocities of atoms via the correlation to kinetic energy. To maintain the temperature at a constant level, the atomic accelerations have to be restrained within some limits (40). Using the Langevin thermostat, such control is imposed by adding a friction term and a stochastic force ($\vec{R}_i$) to Newton's second law, according to equation (4.13), where ζ is the friction coefficient (40) (41).

$$\vec{a}_i(t) = \frac{\vec{F}_i(t)}{m_i} - \frac{\zeta}{m_i}\vec{v}_i(t) + \frac{\vec{R}_i(t)}{m_i} \quad (4.13)$$

The stochastic noise term is a model of Brownian motion involving random particle collisions with solvent molecules (36). The random force has a Gaussian distribution with an average of zero and a variation that is related to the target temperature (41). The friction coefficient affects the relaxation rate of the system (36), and is related to the collsion frequency (γ) by equation (4.14) (41). Furthermore, there is a correlation between the friction coefficient and the diffusion coefficient (D_i) according to the Einstein relation in equation (4.15), which is an example of the fluctuation-dissipation theorem. The temperature is denoted by T, and k is the Boltzmann constant (42).

$$\gamma_\iota = \frac{\zeta}{m_i} \quad (4.14)$$

$$D_i = \frac{kT}{\zeta} \quad (4.15)$$

The momentum is not conserved using the Langevin thermostat. This fact has the implication of not providing a correct description of long-range hydrodynamic interactions (36).

4.4.2 *The Monte-Carlo Barostat*

Employing the isothermal-isobaric (*NPT*) ensemble, the pressure is likewise fixed at a constant value. This intention is achieved by allowing the volume of the system to change during the simulation through the coupling to a barostat (36). Using the MD/ MC algorithm, a trial volume change (ΔV) of the periodic box is computed according to equation (4.16), which is based on a random number generator to obtain a scaling factor (n_{rand}) within the range [-1, 1]. The maximal volume move (ΔV_{max}) is set in accordance with a 40 – 50 % acceptance ratio following the standard MC probability, as defined in equations (4.20) and (4.21) (43).

$$\Delta V = n_{rand} \Delta V_{max} \quad (4.16)$$

The new volume of the periodic box (V') is calculated according to equation (4.17), where the original volume is denoted by V. The adjusted box lengths (l_i') are determined from equation (4.18), where l_i is the original box length. The adjusted coordinates (r_i') corresponding to the center of mass of each molecule in the system are calculated using equation (4.19), where the original coordinates are defined by r_i, and the coordinates of the center of the box are assigned by c_i (43).

$$V' = V + \Delta V \quad (4.17)$$

$$l_i' = l_i \sqrt[3]{\frac{V'}{V}} \quad i = x, y, z \quad (4.18)$$

$$r_i' = (r_i - c_i) \sqrt[3]{\frac{V'}{V}} + c_i \quad i = x, y, z \quad (4.19)$$

The potential energy of the adjusted configuration (U_{pot}') is assessed according to the Metropolis algorithm in equation (4.20), as to determine whether the trial volume move will be accepted or rejected. If $\Delta W > 0$, the probability of acceptance (P) is determined from equation (4.21). If $\Delta W \leq 0$, the volume move is always accepted. In the referred equations, U_{pot} is the potential energy of the original configuration, P_0 and T_0 are the desired pressure and temperature respectively, N is the number of particles in the system and k is the Boltzmann constant (44).

$$\Delta W = U_{pot}' - U_{pot} + P_0 \Delta V - N k T_0 ln\left(\frac{V'}{V}\right) \quad (4.20)$$

$$P(\Delta V) = \begin{cases} e^{-\frac{\Delta W}{kT_0}} & \Delta W > 0 \\ 1 & \Delta W \leq 0 \end{cases} \quad (4.21)$$

4.5 Molecular Dynamics Simulation of the Freeze-Drying Process

The simulation of the freeze-drying process conducted in this project was in part adopted from a number of molecular dynamics protocols of similar studies, as developed by Arsiccio *et al*. The group has investigated the effect of excipients on the stabilization of biopharmaceuticals during lyophilization (3) (4) (11) (45) (46) (47).

4.5.1 *TIP4P and TIP4P/Ice Water Models*

The simulation of the freezing process stage is based on inserting hexagonal ice (I_h) blocks into a simulation box also containing liquid water, as to accurately resemble the formation of ice and the generation of an ice-water interface. In this regard, the TIP4P/Ice water model is suitable for representing water molecules due to its authentic prediction of the ice melting temperature and the ice-water phase diagram (3). In preceding and consecutive simulations, where freezing of the system is not intended, the TIP4P water model makes a good choice owing to its compatibility with the TIP4P/Ice model and its ability to simulate liquid water at low temperatures. To avoid artificial phase separation, the different models should not be mixed in the same simulation (3).

A variety of different water models have been developed over the years. However, many of them suffer from poor predictive ability of solid-phase properties. In the majority of cases, more complex models offer little advantage, and the use of these representations has a critical drawback in increasing computational time. In this regard, the TIP4P water model has been successfully implemented in computational chemistry. The geometry of the TIP4P model is based on four interaction sites. Three of these are placed at the positions of the constituting atoms, i.e. the oxygen and the two hydrogen atoms respectively. The remaining *M* site is placed at the position of the bisector of the H–O–H angle in the same plane as the other interaction sites. The O–H bond length and the H–O–H bond angle are specified according to the experimental values of 0.9572 Å and 104.52 ° respectively (48). The TIP4P water model is illustrated in Figure 4.3.

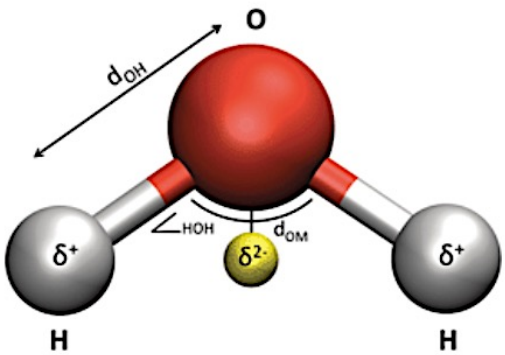


Figure 4.3. Visual representation of the TIP4P water model, which comprises four interaction sites. Positive partial charges (δ^+) are placed at the positions of the hydrogen atoms and a neutralizing negative partial charge (δ^{2-}) is placed at the position of the *M* site, which is located at the bisector of the H–O–H bond angle at a specified distance (d_{OM}) from the position of the oxygen atom. The values of the O–H bond length (d_{OH}) and the H–O–H bond angle are based on experimental data (48). Figure adopted from (49).

In the TIP4P water model, the partial charges are located at the positions of the hydrogen atoms and at the *M* site. The oxygen site constitutes the sole contribution to the Lennard-Jones term in the non-bonded pair potential, as corresponding to the interaction between oxygen atoms in two different water molecules. In the mathematic expression of the non-bonded interaction term, valid for either of the TIP4P or TIP4P/Ice models, there are four unknown parameters: the depth of the energy minimum in the Lennard-Jones pair potential, the equilibrium distance between two oxygen sites, the hydrogen site partial charge and the distance separating the oxygen site and the M site. The parameterization procedure is performed by a nonlinear fit to experimental data of chosen physical properties, which are assigned suitable weighting factors. The parameters of the TIP4P and TIP4P/Ice water models are displayed in Table 4.1. The distinguished application areas of the two models with respect to simulations of liquid water and hexagonal ice respectively, account for a good estimation of the relevant physical state (48).

The TIP4P and TIP4P/Ice models have been employed successfully in studies similar to this project (3). The use of the TIP4P/2005 water model might be an alternative approach since it is able to substitute as a general model for describing the condensed phases of water (50). However, due to the limited data of using the TIP4P/2005 model for simulating freeze-drying of biomolecules in solution, the TIP4P and TIP4P/Ice models were preferred for this project.

Table 4.1. Comparison of the parameters of the TIP4P and TIP4P/Ice water models. In the table, ε and σ are the parameters of the Lennard-Jones potential of the oxygen atom, q_H (= -0.5q_M) is the partial charge located at each of the positions of the hydrogen atoms and d_{OM} is the distance between the *M* site and the position of the oxygen atom (48).

Model	Parameters			
	ε/k [K]	σ [Å]	q_H [e]	d_{OM} [Å]
TIP4P	78.0	3.154	0.520	0.150
TIP4P/Ice	106.1	3.1668	0.5897	0.1577

4.5.2 *Simulations of Freezing*

Matsumoto *et al.* investigated the spontaneous process of ice formation during freezing using the TIP4P water model in a molecular dynamics simulation study performed at the condition of constant pressure and temperature (51). They found that the inception of ice nucleation requires the formation of a sufficient amount of long-lived hydrogen bonds in the same local environment. The freezing process was preceded by rapid quenching from 400 to 230 K and a dormant period, as characterized by random hydrogen bond rearrangement, at which water exists in a supercooled liquid state. Following the formation of an initial stable ice nucleus, the system enters a phase of successive ice growth, as associated with the extension of a three-dimensional hydrogen bond network composed of six-membered rings. The process of crystallization occurs at gradually increasing rate, which is correlated to a corresponding decrease in potential energy, and eventually levels off during ice structure completion. These observations are in agreement with the widely accepted theory of ice nucleation (51). However, freezing of an aqueous system in complex with solutes requires the addition of preformed ice blocks, as to provide the initial configuration of the hexagonal ice phase. This strategy has been used successfully elsewhere (11) (45) by application of the GenIce algorithm (52) for generation of ice structures.

4.5.3 *Simulations of Drying*

Steinberg *et al.* investigated the dynamics of water evaporation from cytochrome *c* using molecular dynamics simulation during vacuum conditions and at constant energy (53). The procedure of creating a vacuum and covering the protein surface with a monolayer of water molecules, as to generate the hydration shell, was performed by means of two different strategies. The first method was initiated with a short relaxation phase in a simulation box containing a very small number of water molecules, following the removal of all waters that were dislocated from the surface. This approach produced a non-uniform distribution with water aggregates gathered at charged surface patches. Using the second strategy, the protein was placed in a bulk water box and all water molecules located beyond a certain cut-off distance from the protein surface, as specified to render an equal number of water molecules, were removed. This method produced a more uniform but less stable distribution in regard of water escaping tendency from the surface. During the coarse of evaporation, a gradual temperature decreased was monitored as a consequence of the kinetic energy cost of breaking of protein-water interactions. The study revealed that the second strategy resulted in an increased rate of evaporation. In addition, setting the initial temperature to a higher value promoted the rate of evaporation and produced a lower residual moisture level, as corresponding to a fewer number of water molecules attached to the surface. The fraction of evaporated water was determined to 10 – 50 %. Under the condition of constant energy, the conformational stability of the protein was not affected during desolvation. However, lysine residues displayed a high degree of flexibility and a tendency to stretch out from the protein surface with water molecules clustering at the positively charged end of the side chain (53).

Another study investigated the evaporation of water from model protein ubiquitin using molecular dynamics simulation during vacuum conditions (54). Two structures of partially solvated protein with different hydration shell thickness were generated by removing all water molecules located beyond 6 or 3 Å respectively. The fraction of evaporated water was greater for the thin shell structure, while there was no difference in the evaporative cooling rate. The conformational change was evaluated for the two different structures, and compared to the cases of fully solvated protein in bulk water and isolated protein in vacuum environment respectively. A trend of increased protein compactness, as measured by the radius of gyration, was revealed upon gradually reducing the hydration level. Meanwhile, the exposure of hydrophobic or

solvent-excluding surface area was found to increase, as correlated to the loss of the hydrophobic effect when removing entropically disfavoured hydration water. In comparison to protein in bulk water, the number of hydrogen bonds towards the solvent was found to decrease significantly when reducing the hydration shell thickness. In contrast, there was an approximately equal number of protein-water hydrogen bond interactions for the thick shell structure, as further indicating the higher extent of preservation of native protein conformation (54).

4.6 Accelerated Molecular Dynamics Simulation

Enhanced sampling techniques such as accelerated molecular dynamics (aMD) presents an additional option of tackling the issue of simulating biological processes where the relevant time scale is inaccessible to conventional molecular dynamics (cMD). Protein conformational change frequently involves a large number of functional transition states, and thus the associated free energy landscape contains an equal number of energy minima separated by high-energy barriers, where the system might be trapped for prolonged time periods. Crossing these energy barriers relies on rare events to occur, which might require excessive computational time (55).

The aMD approach allows for a full exploration of the protein conformational space. In the method introduced by the McCammon group, a bias boost potential $\Delta U(r)$ is added to the true potential $U(r)$ when this is below a defined threshold energy E, as to produce a modified potential $U(r)^*$ according to equation (4.22) (56).

$$U(r)^* = \begin{cases} U(r), & U(r) \geq E \\ U(r) + \Delta U(r), & U(r) < E \end{cases} \quad (4.22)$$

This modification has the effect of reducing the height of the energy barriers by raising the local energy minima, as illustrated in Figure 4.4. The outcome of this approach facilitates the progression of the system through the energy landscape by increasing the escaping rate from the energy minima, and thus significantly decreases the required simulation time span. aMD offers an advantage over many other enhanced sampling techniques in that a single copy of the system is sufficient. Moreover, there is no requirement of prior knowledge of the energy barriers and minima or of the corresponding conformational states. (56).

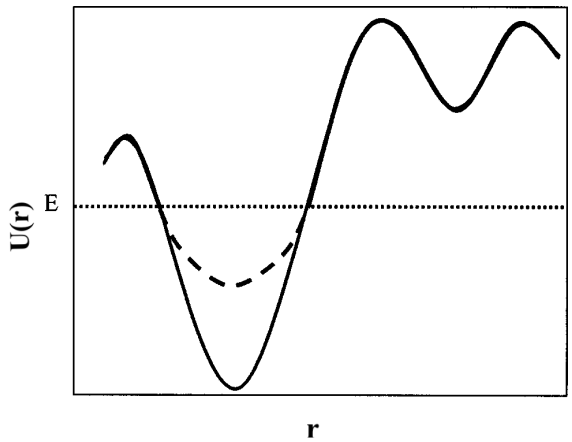


Figure 4.4. Graphical representation showing the principle idea of the aMD approach, which relies on raising the local energy minima of the true potential *U(r)* (solid line) to produce a modified potential (dashed line) through the addition of a boost potential, if below a specified threshold energy *E*. Figure adopted from (56).

Protein conformational change greatly involves torsional rotation of the dihedral angles of the protein backbone. Therefore, the dihedral energy constitutes a main contribution to the total potential energy in this regard. Thus, a dual boosting approach, where an extra boost is added to the torsional terms of the potential, might be effective to explore the conformational space of the protein (56). In agreement with this idea, the boost potential $\Delta U(r)$, as defined in equation (4.23), is divided into separate terms for the total potential and the dihedral potential. In this equation, α_p and α_d are the boost factors applied to the total potential and dihedral potential respectively. Similarly, E_p and E_d are the threshold energies for the total potential and the dihedral potential respectively (57). The boost factors affect the depth of the energy minima of the modified potential and the values of the boost parameters α and E have a crucial impact on the level of acceleration of the molecular dynamics simulation (56). Using the approach described by Salomon-Ferrer *et al.* (58), the computation of the boost parameters follows equations (4.24 – 4.27), as based on empirical findings. In these equations, n_{units} is the total number of protein residues and excipient molecules, and n_{atoms} is the total number atoms in the system. The values of the threshold energies depend on the average total potential energy U_p^0 and the average dihedral energy U_d^0 respectively, as typically obtained from a cMD simulation (58).

$$\Delta U(r) = \frac{\left(E_p - U(r)\right)^2}{\alpha_p + \left(E_p - U(r)\right)} + \frac{(E_d - U_d(r))^2}{\alpha_d + (E_d - U_d(r))} \quad (4.23)$$

$$E_p = U_p^0 + 0.16 n_{atoms} \quad (4.24)$$

$$E_d = U_d^0 + 4 n_{units} \quad (4.25)$$

$$\alpha_p = 0.16 n_{atoms} \quad (4.26)$$

$$\alpha_d = \frac{1}{5} \cdot 4 n_{units} \quad (4.27)$$

To further accelerate the simulation, a high-boost mode might be applied to the parameters E_p, E_d and α_p according to equations (4.28 – 4.30) (58).

$$E_p = U_p^0 + 0.2 n_{atoms} \quad (4.28)$$

$$E_d = U_d^0 + 4 n_{units} + 10 \cdot \frac{1}{5} \cdot 4 n_{units} \quad (4.29)$$

$$\alpha_p = 0.2 n_{atoms} \quad (4.30)$$

4.7 COARSE-GRAINED MOLECULAR DYNAMICS SIMULATION

Coarse-grained (CG) modelling offers an option of simulating diverse processes of large biomolecular systems that occur on lengthy time scales, and which are computationally challenging, by replacing the all-atom molecular description with a lower resolution model. The reduction of the atomistic degrees of freedom (DOFs) allows for an extension of the spatial and temporal limits of conventional molecular dynamics (cMD) (59). The CG modelling approach is based on a united atom representation of the system, where individual atoms are grouped into a reduced number of interactions sites. For instance, hydrogen atoms might be integrated with their associated heavy atoms and a bond constraint might be added to a group of four adjacent heavy atoms, which are ultimately depicted as one coarse-grained bead particle. However, the grouping of atoms can be performed in many ways with the aim of providing an acceptable representation of the chemical and physical properties of the system. Subsequently, a parameter set of choice, as defined within a specified CG force field, is employed in order to compute the effective interaction potentials (60).

Using the CG system simplification strategy, the number of particles and pairwise interactions are reduced. In addition, high frequency motions are effectively eliminated, which permits a larger integration time step to be used in the equations of motion. The combined effect is a reduction of the computational demand, which allows for simulation of a longer time interval using an equal number of time steps (60).

4.7.1 *SIRAH Force Field*

Darré *et al.* developed a CG force field (SIRAH) specifically adapted to simulation of biomolecules (61). The parameterization procedure of the SIRAH force field was performed according to a top-down approach, where parameters were fitted to available structural data as obtained from the Protein Data Bank (PDB). Unlike many other CG force fields, SIRAH models long-range electrostatic interactions, which is accomplished using the Particle Mesh Ewald (PME) method. The mapping scheme relies on a match of the coarse-grained beads with the positions of distinguished molecular units with various degree of granularity. For instance, individual beads model the nitrogen, α-carbon and oxygen respectively in the protein backbone.

In contrast, the side chains are treated with a lower atomistic detail, where the mapping of interaction sites is related to the characterization of the specific residue according to size, charge, aromaticity etc. The performance of the SIRAH force field was tested in a number of trial simulations of different proteins and complexes thereof. Phenomena such as protein-solvent interactions, self-aggregation and conformational change related to temperature or ionic strength, were analysed in this context (61).

The group previously developed an explicit solvent model of water, named WatFour (WT4), which was successfully implemented with SIRAH force field (62). The basis of the WT4 model is the tetrahedral hydrogen bond structural arrangement, as observed temporarily in liquid water (Figure 4.5). Thus, WT4 consists of four connected beads, as representing 11 water molecules with respect to the density of water, where each bead is assigned one partial charge of ± 0.41e. This charge distribution offers an advantage in its ability to produce a dielectric permittivity on its own, and thus avoiding the issue of a using a continuum dielectric medium, as in the case of implicit solvation. The WT4 model has proven to provide a good approximation of water characteristics at the relevant temperature range of biomolecules (62).

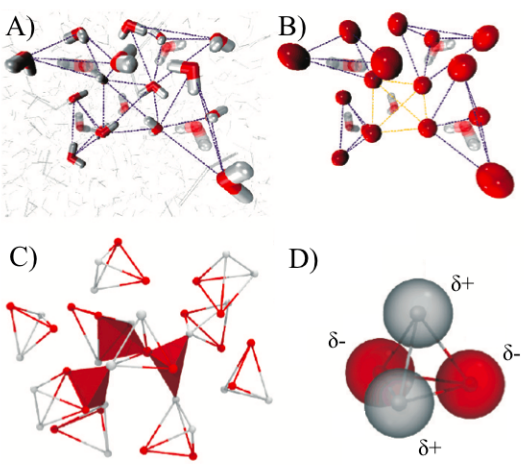


Figure 4.5. Illustrations showing the CG mapping of water molecules with respect to the WatFour (WT4) water model. A) Ordering of liquid bulk water showing the positions of water molecules in the corners of transiently formed irregular tetrahedrons. B) The positions of the oxygen atoms are visualized by the red beads, which compose the basic concept of the WT4 model. C) Structural arrangement of liquid bulk water using the WT4 model. D) The tetrahedral geometry of a WT4 molecule comprising four coarse-grained beads, which are assigned positive (white) and negative (red) partial charges respectively (62). Figure adopted from (62).

An update of the SIRAH force field (SIRAH 2.0) was launched in 2019, as to improve the structural representation of proteins (63). The modifications included a redefined set of bonded and non-bonded parameters of amino acids. The torsion angles of the protein backbone were adjusted as to allow for more flexibility and a wider range of side chain motion was permitted for selected residues. In regard of the non-bonded parameters, changes were made as to improve the hydrophobic/ hydrophilic balance. This aim was sought to be achieved through the classification of coarse-grained beads into five main groups: hydrophilic, aromatic, hydrophobic, backbone and water or ions. A subdivision was made within each group based on charge module and Lennard-Jones parameters. A further subdivision was performed within each subgroup as to cover the sign of the charge and Lennard-Jones parameters that are not computed according to the Lorentz-Berthelot combination rules, such as hydrogen bonds, salt-bridges and cation-π interactions. In addition, a uniform mass assignment of the coarse-grained beads was introduced. SIRAH 2.0 was implemented with the AMBER molecular dynamics suite. In a comparative test to the previous force field version, the modifications included in SIRAH 2.0 were observed to provide enhanced robustness in terms of a number of structural descriptors, such as root mean square deviation (RMSD), solvent-accessible surface area (SASA), radius of gyration (R_G) and native contacts. The capacity of SIRAH 2.0 was further evaluated in a study of calcium-binding protein Calmodulin (CaM) and its ligand peptide, as to explore the structural and dynamical behaviour in terms of metal-ion coordination, conformational change and protein-peptide interaction. The results from this analysis were in good agreement with experimental data, and further displayed a satisfactory resolution level (63).

Chapter 5. MOLECULAR DYNAMICS SIMULATION PROCEDURE

The system set-up files and simulation input files were prepared by Ph.D. student Suk Kyu Ko, who also developed the scripts for analysis. The freeze-drying simulation protocol was in part adopted from similar studies performed by Arsiccio *et al.* (3) (4) (11) (45) (46) (47). The analysis was performed using Tcl and Python scripting, Visual Molecular Dynamics (VMD) (64), PyMOL (65) and MATLAB (66).

5.1 MODEL STRUCTURE

The input structure of Reteplase was generated using homology modeling. The FASTA format of the protein sequence was downloaded from the DrugBank, which is a freely accessible online pharmaceutical database (67). The BLAST algorithm (68) was used to align the query sequence of Reteplase to sequences available at the Protein Data Bank (PDB) in order to identify regions of local similarity. Based on the results from this search, one structure of the catalytic domain of human sc-t-PA (PDB ID: 1BDA(A)) and three structures of the kringle-2 domain of human t-PA (PDB IDs: 1PK2(A), 1TPK(A) and 1TPK(B)) were chosen for homology modeling due to high percentage sequence identity and high significance of the matches.

The MODELLER 9.20 software (69) was used for sequence alignment of each template and the target to confirm the validity, and for building the associated homology models, as displayed in Table (5.1). Ten different submodels were generated for each homology main model. The submodels were submitted to the RAMPAGE webserver tool (69) for analysis of the corresponding Ramachandran plots and evaluation of outlier regions comprising amino acids with non-favourable dihedral angles. The best models in this regard, were visually analyzed in VMD to assess the separation between the two protein domains, and to exclude models with potential structure overlap. Based on these arguments, one of the submodels of 1TPK(B) – 1BDA(A) was chosen as the optimal model to be used as the input structure for simulations. This submodel had a sequence identity of 100 %, and the percentage of Ramachandran outliers, excluding glycine and proline residues, was determined to 0.6 %. The sequence alignment of 1TPK(B), 1BDA(A) and Reteplase is displayed in Figure (5.1). The first three N-terminal residues in the sequence of Reteplase (Ser1–Tyr2–Gln3) were not covered by the structure of 1TPK(B), which were therefore automatically modelled by MODELLER.

Table 5.1. The three different homology main models, as generated from available structures of the kringle-2 domain (K2) and the catalytic serine protease domain (P) of human t-PA, using MODELLER software. An alignment of the target sequence to each template was performed prior to model construction. The templates are denoted by their PDB IDs and the chosen chain IDs for modelling.

Template 1 (K2)	Template 2 (P)	Homology main model
1PK2(A)	1BDA(A)	1PK2(A) - 1BDA(A)
1TPK(A)	1BDA(A)	1TPK(A) - 1BDA(A)
1TPK(B)	1BDA(A)	1TPK(B) - 1BDA(A)

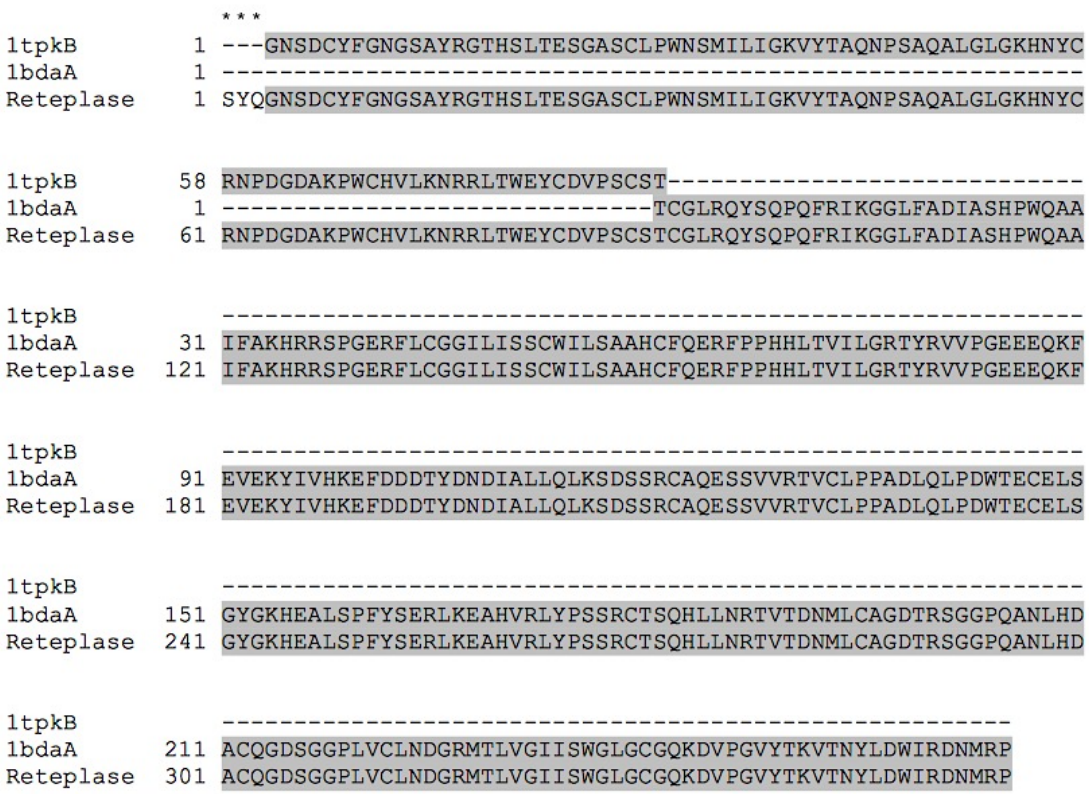


Figure 5.1. Sequence alignment of 1TPK(B), 1BDA(A) and Reteplase using MODELLER 9.20. Identical sequences are highlighted in grey and gaps in the primary structures are marked by " - ". Residues for which no crystal structure is available are marked by " * ".

5.2 SYSTEM PREPARATIONS

5.2.1 *Setup of Full-atom Molecular Dynamics Simulations*

The topology and coordinate files used for the simulations were generated using the tLEaP preparatory program included in the AmberTools19, as part of the AMBER molecular dynamics suite (70). The Amber18 ff14SB force field developed for proteins and nucleic acids, was used to model the protein (70).

The Antechamber software package, as included in AmberTools19 (70), was used in conjunction with tLEaP to create library files, as well as topology and coordinate files for the excipients investigated in complex with protein in this project. Three-dimensional molecular structures to be used as input files were downloaded from the ZINC database. The GAFF2 AMBER force field (71), specifically intended for small organic molecules within rational drug design, was used to model the relevant excipient molecules. Antechamber functions to define atom and bond types, assign atomic charges and add missing force field parameters. It was found that excipient molecules modeled by GAFF2 interacted too strongly with protein, which resulted in inaccurate preferential interaction coefficients. In order to better reproduce the experimental evidence, Ph.D. student Suk Kyu Ko performed a modification to the non-bonding parameters obtained from the CHARMM force field (72) (73) (74) (75) (76), as based on approaches published by others in this subject (77) (78) (79). In the case of sugar excipient molecules, the Kirkwood-Buff parameter set of the CHARMM force field was applied, which has been used successfully elsewhere (5).

The model protein was inserted into a cuboid periodic boundary box, where the minimum distance between the protein and the edge of the box was set to 15 Å, and 27 433 water molecules were subsequently added to the system. For simulations at room temperature the TIP4P Ewald water model (80) was used. Simulations at freezing conditions were conducted using the TIP4P/Ice water model (48). Excipient molecules were introduced into the system through replacement of randomly selected water molecules. The number of excipient molecules (n_{ex}) was calculated using equation (5.1), where c_{ex} is the concentration of excipient expressed as the weight percentage of excipient per total weight (w/w %), n_w is the number of water molecules, M_w is the molar mass of water and M_{ex} is the molar mass of excipient. Molar concentrations were converted to weight percentages prior to using equation (5.1). The system

was neutralized by adding either Na^+ or Cl^- counterions. For arginine and tranexamic acid, the effect of excipient was investigated at three different concentration levels: low (10 mM), intermediate (100 mM) and high (10 w/w %) concentration. For sucrose, the high concentration level was considered the most relevant based on data from experimental studies within this field, and therefore the lower concentrations were not investigated in this project. The concentrations of excipient and the corresponding number of molecules are presented in Table (5.2).

$$n_{ex} = \frac{c_{ex} n_w M_w}{M_{ex}} \quad (5.1)$$

Table 5.2. The investigated concentrations of excipient and the corresponding number of excipient molecules, as calculated using equation (5.1), of the systems including arginine (ARG), tranexamic acid (TXA) and sucrose (SUC) respectively.

Excipient	Concentration	Number of excipient molecules
ARG	10 mM	5
ARG	100 mM	51
ARG	10 w/w % (~ 570 mM)	284
TXA	10 mM	5
TXA	100 mM	51
TXA	10 w/w % (~ 640 mM)	314
SUC	10 w/w % (~ 290 mM)	144

5.2.2 *Setup of Coarse-grained Molecular Dynamics Simulations*

CG MD was used to study the protein-protein aggregation dynamics of Reteplase in the absence and presence of arginine excipient. The SIRAH 2.0 force field (version 19-08) (63) was used to model the system, which was set up and simulated using the GROMACS molecular dynamics software program (version 2018.5) (81) (82) (83) (84) (85) (86) (87). The simulation protocol was developed in collaboration with the Technical University of Munich (TUM).

To generate topology and coordinate files, the system was set up as described in the following. As required for mapping to the CG representation, the file format of the initial structure of Reteplase was converted from PDB to PQR using standard non-protonated residues, and subsequently translated to GROMACS format. A total of eight protein monomers were inserted into a cubic periodic boundary box, where the minimum distance between a protein monomer and the edge of the box was set to 7.5 Å. The distance between the centers of mass of two adjacent monomers was set to 8.5 nm. The corresponding protein concentration was optimized during preceding trail studies in order to sample maximum protein-protein interactions with an acceptable simulation time. To generate different initial conformations, the monomers were randomly rotated by a maximum of 180º. The system was solvated by adding approximately 36 600 WT4 molecules (62) to the simulation box, and those within 3 Å from the protein were removed. A total of 4115 arginine molecules were added to the system to reproduce a concentration of 10 w/w %, as calculated using equation (5.1) and recognizing that each WT4 molecule corresponds to 11 water molecules. The system was neutralized by the addition of either Na^+ or Cl^- SIRAH counterions.

5.3 Molecular Dynamics simulations of the Freeze-drying process

The complete freeze-drying simulation was divided into five consecutive parts: simulation at room temperature, freezing, primary drying, secondary drying and reconstitution. The settings of the input files for each part are described in the following subsections. In total, the freeze-drying process was simulated for 2 μs. Duplicate simulations were performed for each system as defined in Table (5.2).

5.3.1 *Simulations at Room Temperature*

The system was subjected to energy minimization using the steepest descent algorithm for the first 5000 cycles and the conjugate gradient algorithm for the next 5000 cycles. In all instances, the cutoff distance of non-bonded interactions was set to 12 Å.

The system was subsequently heated during 300 ps by raising the temperature from 10 K to 300 K under conditions of the canonical ensemble (NVT). In all cases, the temperature was controlled using the Langevin thermostat and the associated collision frequency was assigned a value of 5 ps^{-1}. Throughout the simulations of the freeze-drying process, the time step was set to 2 fs. In all occasions, the SHAKE algorithm (88) was used to constrain all hydrogen bonds.

System equilibration was performed in two separate steps using the isothermal-isobaric ensemble (NPT) at 300 K and 1 atm. The equilibration steps were conducted for 200 ps and 2 ns respectively. For all cases when pressure control was intended, the MC barostat was applied. The production was divided into two separate parts performed under the NPT ensemble, each of them being 200 ns long. At the end of each of the five steps of the freeze-drying simulation, the protein was centered in the periodic box in all frames of the trajectory by means of re-imaging the atomic coordinates. The last frame was saved as a PDB file to serve as the initial structure for the freezing simulation.

5.3.2 *Simulations of Freezing*

Before initiating the simulations of the freezing stage, the GenIce open source software (52) was used to generate three hydrogen-disordered hexagonal structure (I_h) ice lattices. The TIP4P/Ice water model was employed. For each ice block, the unit cell was repeated in *x*, *y* and *z* directions as to obtain the dimensions required to cover the three repeating surfaces of the periodic boundary box comprising the simulation system.

A short minimization was conducted in order to remove bad contacts by means of slow repulsion between atoms that had been positioned in too close proximity. The steepest descent algorithm was used for the first 500 cycles and the conjugate gradient algorithm for the next 500 cycles. Subsequently, the system was heated slowly during 300 ps from 10 K to 100 K under conditions of the NVT ensemble. A second heating was conducted for an additional 300 ps employing the NPT ensemble, as to further raise the temperature to 233 K. The equilibration and production were each performed in a two-step manner using the same approach as for the room temperature simulation, although at the lower temperature of 233 K. At the end of the freezing simulation, the protein and all excipient and water molecules within 3 Å of the protein surface, as corresponding to the hydration shell, were saved as a PDB file to be used as the initial structure for the simulation of the primary drying.

5.3.3 *Simulations of Primary Drying*

Before initiating the simulation of the primary drying stage, a cuboid periodic box without solvent molecules was generated around the protein and its associated hydration shell and included excipient molecules, as specified to enclose entire atoms. The minimum distance between the protein and the edge of the simulation box was set to 50 Å in order to create the vacuum space required for the drying process.

The system was subjected to heating for 300 ps using the NVT ensemble to raise the temperature from 10 K to 233 K. As before, the production was divided into two parts, each 200 ns long. However, for this case, the simulation was performed under the conditions of the NVT ensemble at 233 K. At the end of the simulation of the primary drying, the protein and all excipient and water molecules within 2.4 Å of the predefined crystal water molecules were saved as a PDB file to be used as the initial structure for the simulation of the secondary drying.

5.3.4 *Simulations of Secondary Drying*

Before initiating the simulation of the secondary drying stage, a vacuum box was generated around the protein and its bound water and excipient molecules, using an identical approach as described for the primary drying phase.

The system was heated from 10 K to 300 K during a time span of 300 ps under application of the NVT ensemble. The production was performed in the same manner as for the case of the primary drying, although at 300 K. At the end of the simulation of the secondary drying, the entire system was saved as a PDB file to be used as the initial structure for the reconstitution simulation.

5.3.5 *Simulations of the Reconstitution Phase*

Before initiating the simulation of the reconstitution stage, the protein and its bound excipient and water molecules were placed in a cuboid periodic water box, where the minimum distance between the protein and the edge of the box was set to 15 Å.

An energy minimization was performed using the steepest descent algorithm for the first 500 cycles and the conjugate gradient algorithm for the next 500 cycles. The heating, equilibration and production of the system were performed using the same settings as described for the initial simulation at room temperature.

5.4 Accelerated Molecular Dynamics Simulations

The input file for aMD simulations was obtained from the AMBER Advanced Tutorial 22, which is a part of the Amber 2020 Reference Manual (57) and follows the approach described by Salomon-Ferrer *et al.* (58). The aMD simulations were performed using normal boost mode and high-boost mode respectively, as defined in section 4.6. The simulated systems contained Reteplase in the absence and presence of arginine (10 w/w %). The employed boost parameters α and E for the total potential and the dihedral potential respectively, are displayed in Table 5.3.

Table 5.3. The calculated values of the boost factor α and the threshold energy E for the total potential (p) and the dihedral potential (d) employed for normal boost and high-boost mode aMD simulations of Reteplase in the absence and presence of arginine (ARG).

aMD mode	System	Parameter			
		E_p [kcal mol^{-1}]	E_d [kcal mol^{-1}]	α_p [kcal mol^{-1}]	α_d [kcal mol^{-1}]
Normal	Reteplase	- 294920.6	5825.9	18426.2	284
	Reteplase + ARG	- 408032.4	8776.5	19516.5	511.2
High-boost	Reteplase	- 290314.0	8665.9	23032.8	284
	Reteplase + ARG	- 403153.2	13888.5	24395.6	511.2

The values of the boost parameters were calculated using equations (4.24 – 4.27) for normal boost mode, and equations (4.28 – 4.30) for high-boost mode respectively. The values of the average total potential energy U_p^0 and the average dihedral energy U_d^0 were obtained from the output of cMD simulations at room temperature, which also provided the topology and coordinate files required to initiate the aMD simulations.

All aMD simulations were conducted for 500 ns using a time step size of 2 fs. The conditions of the NVT ensemble were applied at 300 K. The temperature was controlled using the Langevin thermostat and the associated collision frequency was assigned a value of 5 ps^{-1}. The SHAKE algorithm was used to constrain all hydrogen bonds, and for these interactions forces were not calculated. The cutoff distance of non-bonded interactions was set to 12 Å.

5.5 Coarse-grained Molecular Dynamics Simulations

An energy minimization of the system was conducted using the steepest decent and the conjugate gradient algorithms, where the maximum number of time steps was set to 50 000. The cutoff distance of non-bonded interactions was set to 12 Å throughout the steps of the CG simulation. The system was subsequently heated for 2 ns using a time step of 20 fs. The conditions of the NPT ensemble at 300 K and 1 atm were applied. The temperature and the pressure were coupled to the Berendsen thermostat and barostat respectively. For the case when arginine was included in the system, a slow pre-heating process was performed for 10 ns to a temperature of 10 K under the conditions of the NVT ensemble using the Berendsen thermostat.

An equilibration procedure was performed for 500 ns using a time step of 10 fs. The NPT ensemble was applied at 300 K and 1 atm. The temperature was controlled using the velocity rescale thermostat and the pressure was coupled to the Parrinello-Rahman barostat. The production was conducted for 1.5 μs using the same settings as for the equilibration. Duplicate simulations were performed for each of the two systems including Reteplase either in the absence or in the presence of arginine.

Chapter 6. RESULTS AND DISCUSSION

6.1 Initial Protein Structure and Stability Investigations

Initial investigations of the electrostatic potential molecular surface of the input structure of Reteplase using the Adaptive Possion-Boltzmann Solver (APBS) calculation program (89) of PyMOL 2.0 (65) revealed the existence of positively and negatively charged surface patches, as displayed in Figure (6.1). The same observation has been documented elsewhere (26), and provides a foundation for potential protein aggregation events by electrostatic attraction between areas of opposite charge located on the surface of individual monomer units. The protein surface of the catalytic domain is predominantly negatively charged, whereas the surface of the kringle-2 domain is characterized by a greater extent of positively charged sites.

The increase in protein-protein interaction by electrostatic surface forces might explain the low colloidal stability as observed for Reteplase in the absence of excipient.

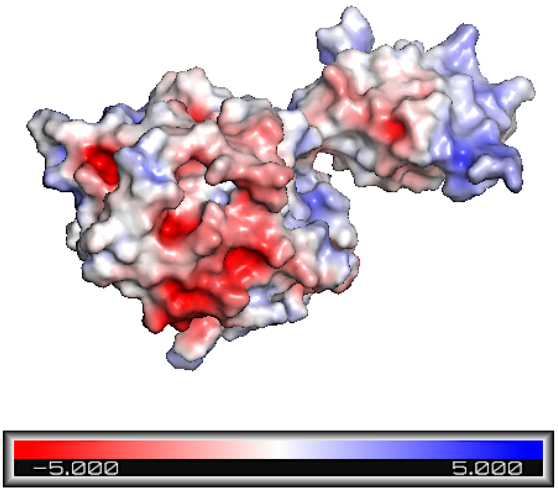


Figure 6.1. The electrostatic surface potential map for the PQR input structure of Reteplase, as calculated using the Adaptive Poisson-Boltzmann Solver (APBS) software plugin of PyMOL 2.0. Red and blue areas display negatively and positively charged surface patches respectively.

Trial investigations of the aggregation propensity of the protein structure during the course of the freeze-drying process using the Aggrescan3D 2.0 server (90), revealed a number of aggregation prone regions, as displayed in Figure 6.2. Aggrescan3D uses a single three-dimensional structure of the protein, in this case the PDB files derived from the last frame of the trajectories of the simulations at room temperature and the secondary drying respectively, as to identify surface-exposed residues that are prone to aggregation. A positive score corresponds to an aggregation-prone residue, whereas a negative score indicates a hydrophilic residue and a zero value score is identical to a buried residue. A comparison of Reteplase in the absence and in the presence of arginine (10 w/w %), indicated a difference in stability in this regard, which was made most notable in the dried state. Local regions of Reteplase in the absence of excipient experience an increase in aggregation propensity during the drying phase. Thus, conformational instability might be enhanced due to stresses during the freeze-drying process. However, when arginine is added to the system the increase in aggregation propensity is prevented to great extent.

The results from the primary investigations provided initial support for the hypothesis of improved protein stability of Reteplase by excipients, which served as a basis for further stability investigations, as explored to more detail and documented in the sections below.

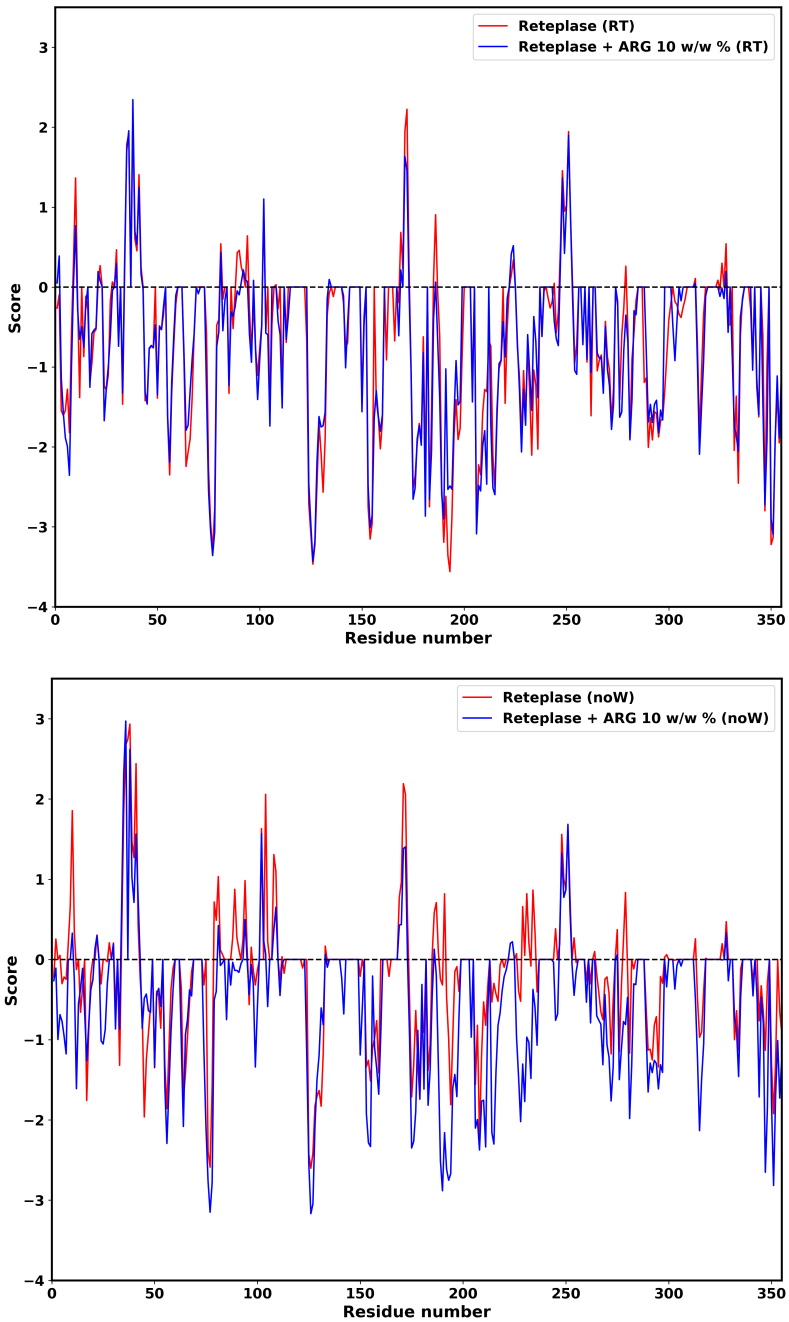


Figure 6.2. Aggregation propensity score per residue of Reteplase in the absence of excipient (red) and in the presence of arginine 10 w/w % (blue), as predicted by the Aggrescan3D server. The upper subplot displays the results from room temperature (RT) simulation and the lower subplot displays the results from the simulation of the secondary drying stage (noW). The value of the score suggests that the associated residue is either aggregation-prone (positive), hydrophilic (negative) or not exposed to the solvent (zero).

6.2 Analysis of Protein Conformational Change during the Simulation of Freeze-drying

The MD simulated freeze-drying process of Reteplase was analyzed with respect to the time evolution of a number of protein structural descriptors, namely the radius of gyration (R_G) and the solvent-accessible surface area (SASA). Furthermore, time-averaged calculations of protein aggregation propensity and protein-excipient interactions were performed. The aim of the analysis was to investigate the conformational stability of Reteplase in the absence and in the presence of excipient molecules respectively. In this regard, the initial cMD simulation at room temperature served as a reference for comparison to the consecutive steps of the freeze-drying simulation. Each step was designed to provide an authentic representation of the real process, as to accurately investigate the effect on protein conformational change induced by the stress factors during freeze-drying.

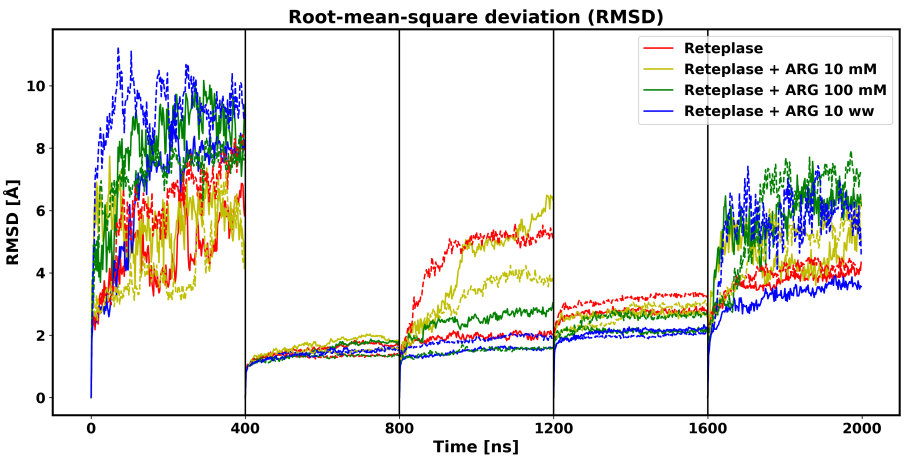
6.2.1 *Root-Mean-Square Deviation*

As a first step, the root-mean-square deviation (RMSD) of the protein was inspected, as displayed in Figure 6.3, in order to assess the convergence of the simulations. For a few systems, it appears that equilibrium has not been reached at the end of specific stages of the freeze-drying simulation, which likely will have an effect on other calculated structural properties. However, due to practical reasons with respect to computational time, each stage was limited to 400 ns. The chosen time span was based on similar simulations of G-CSF. However, the time scale required to model the process of protein conformational change during the stages of freeze-drying might vary for different proteins depending on structural complexity and the degree of protein destabilization.

In particular, the simulations of the reconstitution phase display a high degree of fluctuation in terms of RMSD for systems containing high concentration of arginine or TXA. This observation might be explained by large-scale movements of the protein when refolding from the partly conformationally changed and energetically unfavourable structure of the dried state. The addition of water molecules to the system and its effect on reconfiguration of the protein, is likely another factor contributing to this behavior. At the low temperature of the freezing stage, the protein is effectively immobilized and prevented from extensive movement.

Referring to the analysis of the radius of gyration in the following section, the protein structure is made more compact in the dried state, which reduces the degree of internal fluctuation. For Reteplase without excipient, the effect of conformational reversibility during reconstitution is not observed, and thus less fluctuation is registered. One of the duplicate simulations of Reteplase in the presence of TXA (10 w/w %) display outlier behavior during the reconstitution phase with non-convergence and a very high level of fluctuation in terms of RMSD. This observation will likely have a strong effect on the calculated averages of other properties for this phase of the freeze-drying simulation. By visual inspection in VMD, the connecting loop was observed to adopt a maximally extended conformation for the last 40 ns of the reconstitution, where the side chains of the flexible linker segment Ser90 – Thr91 changed from a cis to a trans configuration.

Similarly, one of the duplicate simulations of Reteplase in the presence of 10 w/w % sucrose displays a distinct increase of RMSD during reconstitution. By observation, this behavior is correlated to a twist of the kringle-2 domain by approximately 90º, as accomplished by the same type of cis-trans reconfiguration of the linker segment.



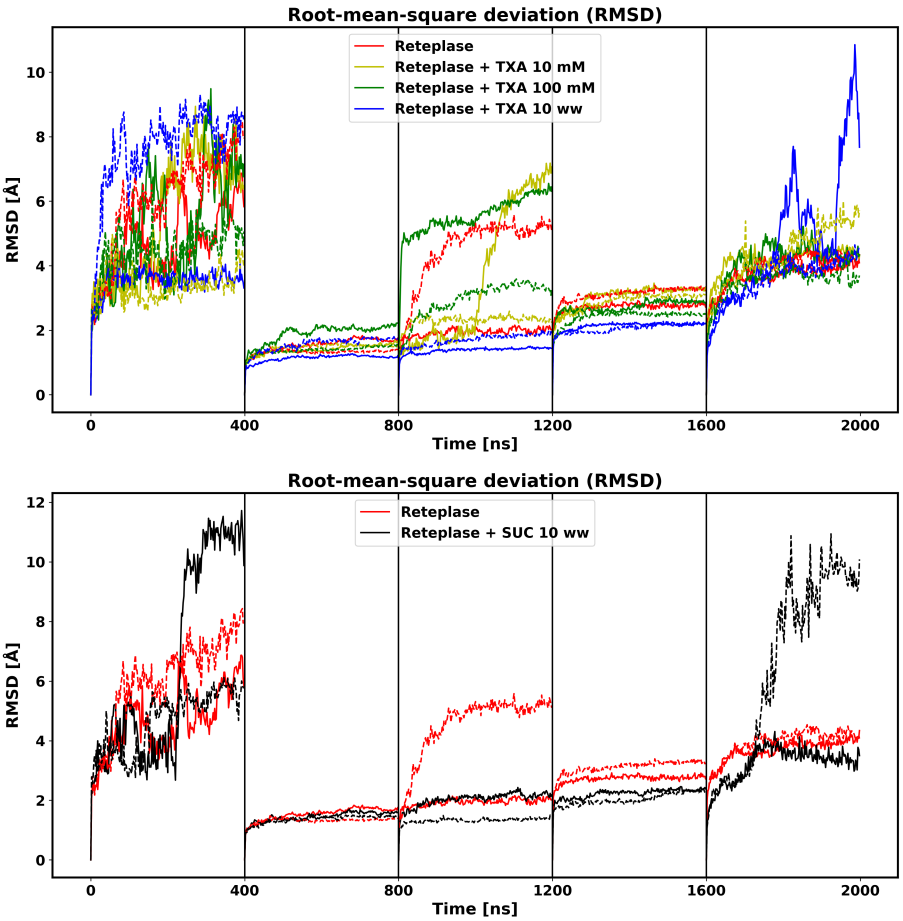
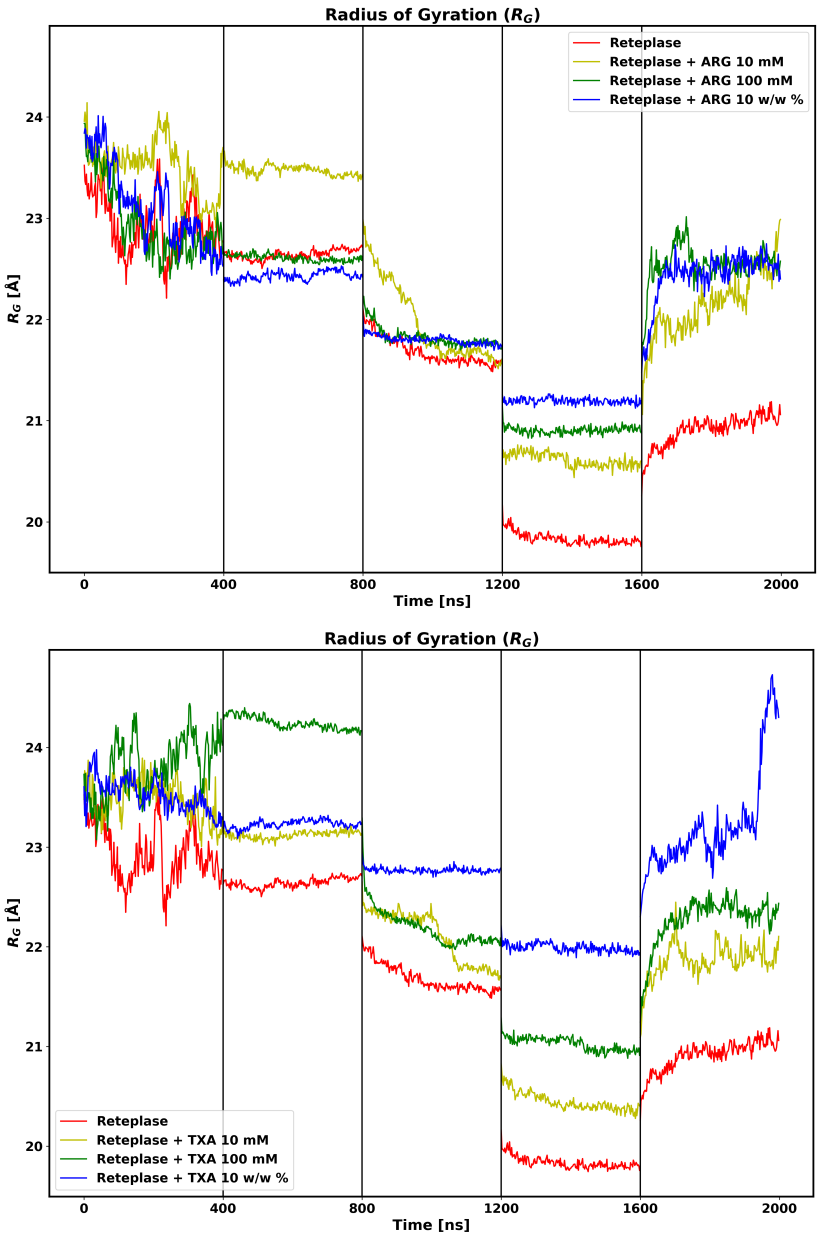


Figure 6.3. Plots displaying the root-mean-square deviation (RMSD) of Reteplase during the different stages of the freeze-drying process, namely room temperature, freezing, primary drying, secondary drying and reconstitution, as separated by vertical lines. The different graphs in each subplot shows the data of Reteplase in a system containing the following concentrations of excipient: 10 mM (yellow), 100 mM (green) and 10 w/w % (blue). The red graphs show the data of Reteplase without excipient. The upper subplot corresponds to the systems containing arginine as an excipient, the central subplot corresponds to the systems containing TXA and the lower subplot corresponds to the system containing 10 w/w % sucrose. The data of the duplicate simulations are represented by dashed lines.

6.2.2 *Radius of Gyration*

The plots of the radius of gyration (R_G) for the investigated systems during the different stages of the freeze-drying process are displayed in Figure 6.4. The results indicate an increased level of compactness of the protein structure in the dried state, which might be a consequence of the removal of space-filling water molecules. In addition, induced dehydration stresses when stepwise removing the hydration shell might play a role in destabilization of the protein.



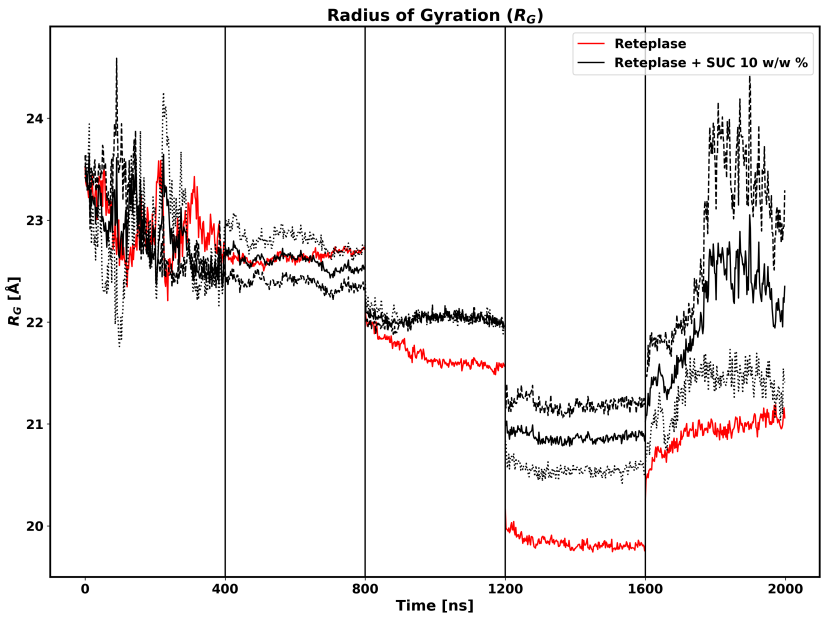


Figure 6.4. Plots displaying the time evolution of the radius of gyration during the different stages of the freeze-drying process. The different graphs in each subplot show the data of Reteplase in a system containing the following concentrations of excipient: 10 mM (yellow), 100 mM (green) and 10 w/w % (blue). The red graphs show the data of Reteplase without excipient. The upper subplot corresponds to the systems containing arginine as an excipient and the central subplot corresponds to the systems containing TXA. Each graph represents the average of the two duplicate simulations of the same system. The lower subplot corresponds to the system containing 10 w/w % sucrose, where average R_G is represented by a solid line and the data of the two duplicates are represented by dotted and dashed lines respectively.

Visual inspection in VMD (Figure 6.5) revealed a tendency of the two protein domains to spatially approach each other in the partly conformationally changed structure. This observation might be a result of a collapse of the connecting loop region into the interior of the protein, and of a twisting motion of the kringle-2 domain. The effect is significantly reduced in the presence of either arginine or TXA. The most pronounced domain separation is obtained through the addition of TXA to the system.

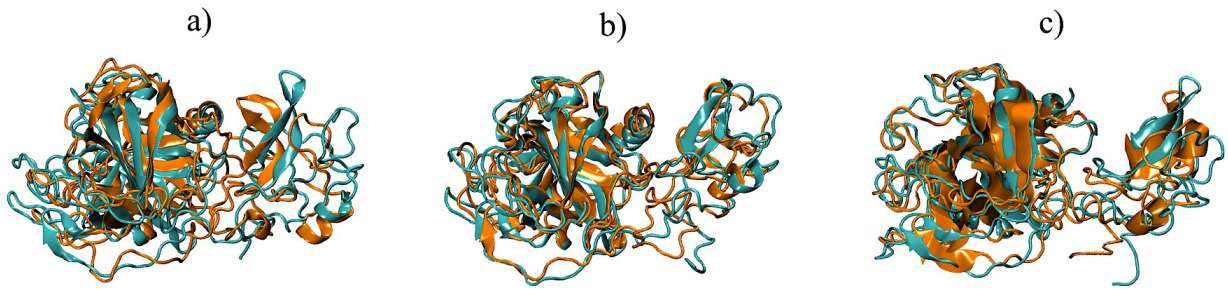


Figure 6.5. Visual representations of the protein conformation of Reteplase comparing the last frames of the simulations at room temperature (blue) and secondary drying (orange). Figures display a) Reteplase without excipient, b) Reteplase with arginine (10 w/w %) and c) Reteplase with TXA (10 w/w %). Figures rendered in VMD.

The systems including arginine and TXA respectively, display the same general trend in terms of limiting the observed shrinkage of the protein structure. This effect is gradually more pronounced when increasing the concentration of excipient, with approximately no impact at the lowest concentration of 10 mM. A concentration of 10 w/w % of either arginine or TXA are equally effective in this regard, whereas 100 mM arginine performs significantly better than the same concentration of TXA. The stretching of the loop region, as observed for one of the duplicate simulations of 10 w/w % TXA, is clearly seen in the data of R_G as a sharp increase during the relevant time span, which indicates a significant loosening of the protein structure.

The high value of RMSD associated with one of the duplicates of 10 w/w % sucrose strongly affects the data of R_G. If disregarding average computation of the two duplicates, reversibility in terms of protein compactness level is not observed, which indicates that sucrose is not as efficient in preserving the native structure and overall protein dimensions of Reteplase.

The observed ability of molecules of TXA to intersect in between the two protein domains, due to its sufficiently small molecular size, provides a steric hindrance to the approaching movement of the domains, which appear to be more spatially separated by inspection of Figure 6.5. This physical property of TXA may act to improve stability in addition to conventional stabilization mechanisms. In contrast to Reteplase without excipient, a minimum concentration of 100 mM arginine appears to induce conformational reversibility in terms of R_G during the reconstitution phase. The same observation is made for a concentration of 10 w/w % TXA. However, large fluctuations are registered for the last few data points, as expected from the results of RMSD.

6.2.3 *Solvent-Accessible Surface Area*

The solvent-accessible surface area (SASA) was calculated based on total surface area, non-polar surface area and the ratio of non-polar to total surface area. The results are displayed in Figures 6.6 – 6.8. In agreement with the data of R_G, the total SASA (Figure 6.6) displays a trend of stepwise decrease during the two drying stages. This observation is significantly more pronounced for Reteplase without excipient, as indicating a higher extent of preservation of the original structure in the presence of either arginine or TXA. The effect is increased with higher concentration of the excipient. At a concentration of 100 mM, arginine performs slightly better in terms of limiting the reduction of total SASA. In contrast to the data of R_G, reversibility in terms of total SASA during the reconstitution phase, is observed for Reteplase both in the absence and in the presence of excipient.

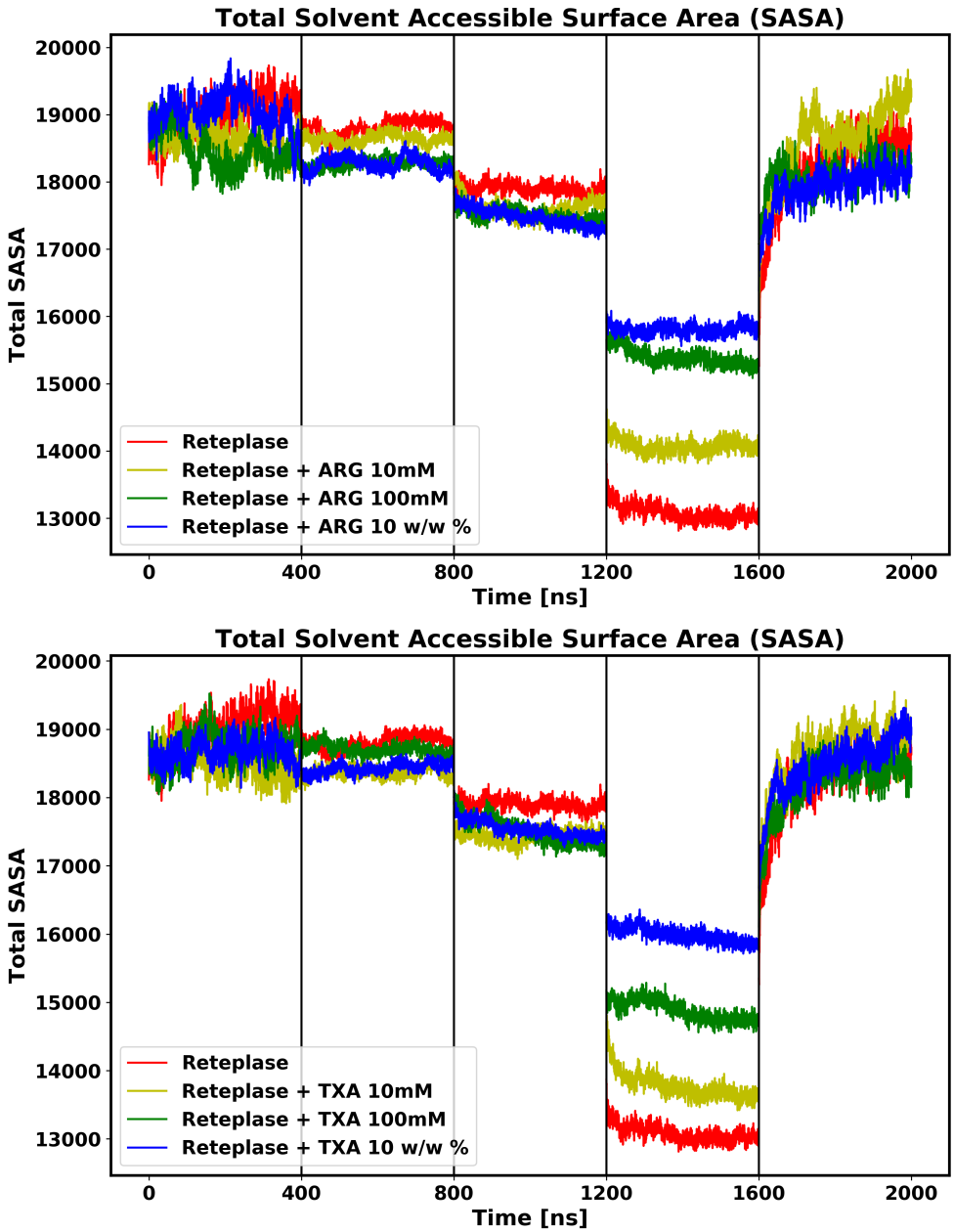


Figure 6.6. Plots displaying the time evolution of the total solvent-accessible surface area during the different stages of the freeze-drying process. The different graphs in each subplot show the data of Reteplase in a system containing the following concentrations of excipient: 10 mM (yellow), 100 mM (green) and 10 w/w % (blue). The red graphs show the data of Reteplase without excipient. The upper subplot corresponds to the systems containing arginine as an excipient and the lower subplot corresponds to the systems containing TXA. Each graph represents the average of the two duplicate simulations of the same system.

The data of the non-polar SASA (Figure 6.7) for arginine and TXA follows the same pattern as for the total SASA, with the exception of an observed increase during the reconstitution phase to above the original level at room temperature. This observation might be a consequence of exposure of non-polar residues, as indicative of an unfolding process. However, the addition of excipient to the system offers conformational stabilization of the protein and reduces the increase of non-polar SASA. At a minimum concentration of 100 mM arginine, non-polar SASA reverts to approximately the baseline level at room temperature. For 10 w/w % TXA, observations during reconstitution might be affected by non-equilibrium behavior.

Theoretically, the non-polar SASA would be expected to increase during dehydration due to the loss of the hydrophobic effect in the absence of water molecules. However, in practice this outcome is not observed. This behavior of the protein might be derived from the physical shrinkage of the protein structure, which is correlated to the decrease of R_G when removing space-filling water molecules.

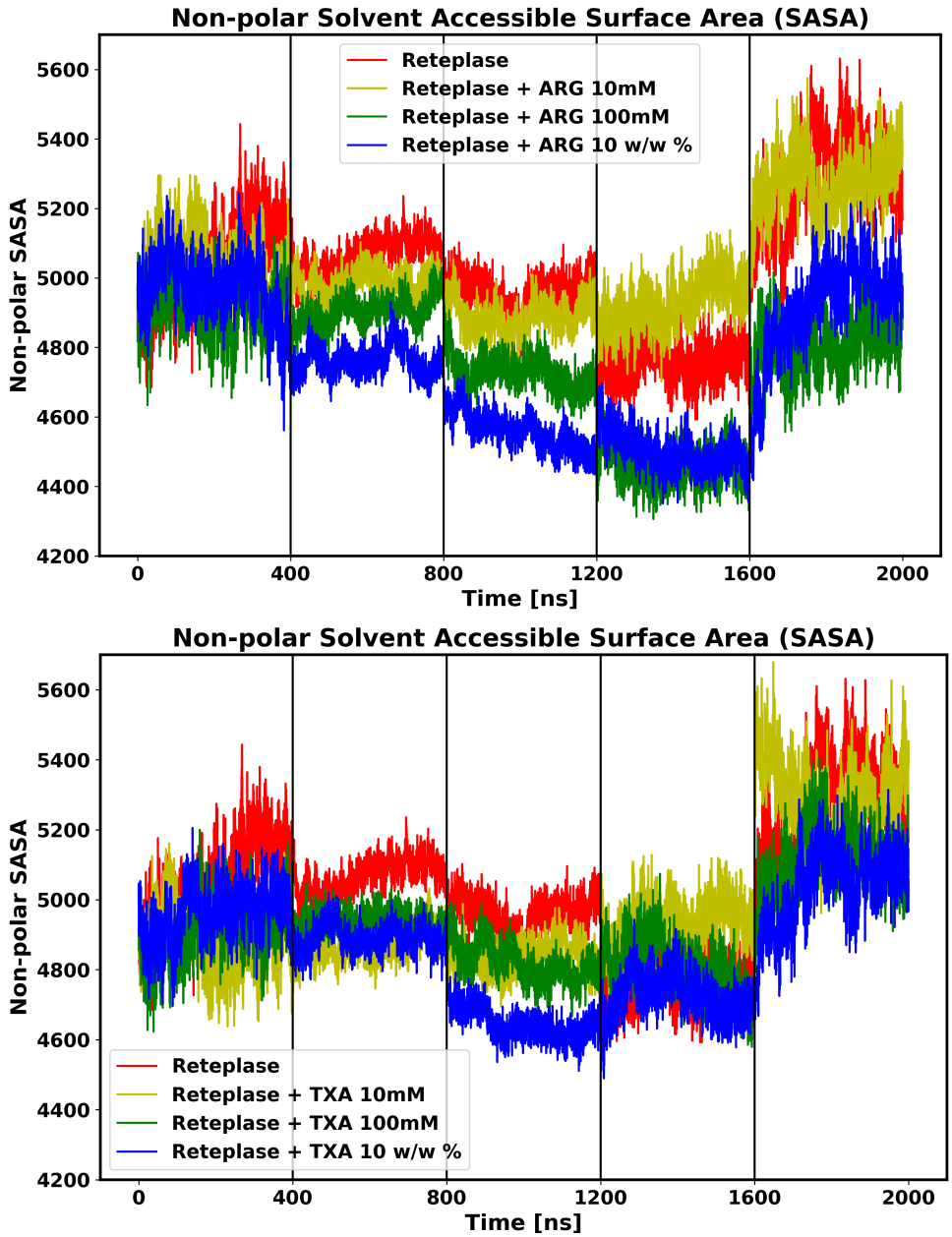


Figure 6.7. Plots displaying the time evolution of the non-polar solvent-accessible surface area during the different stages of the freeze-drying process. The different graphs in each subplot show the data of Reteplase in a system containing the following concentrations of excipient: 10 mM (yellow), 100 mM (green) and 10 w/w % (blue). The red graphs show the data of Reteplase without excipient. The upper subplot corresponds to the systems containing arginine as an excipient and the lower subplot corresponds to the systems containing. Each graph represents the average of the two duplicate simulations of the same system.

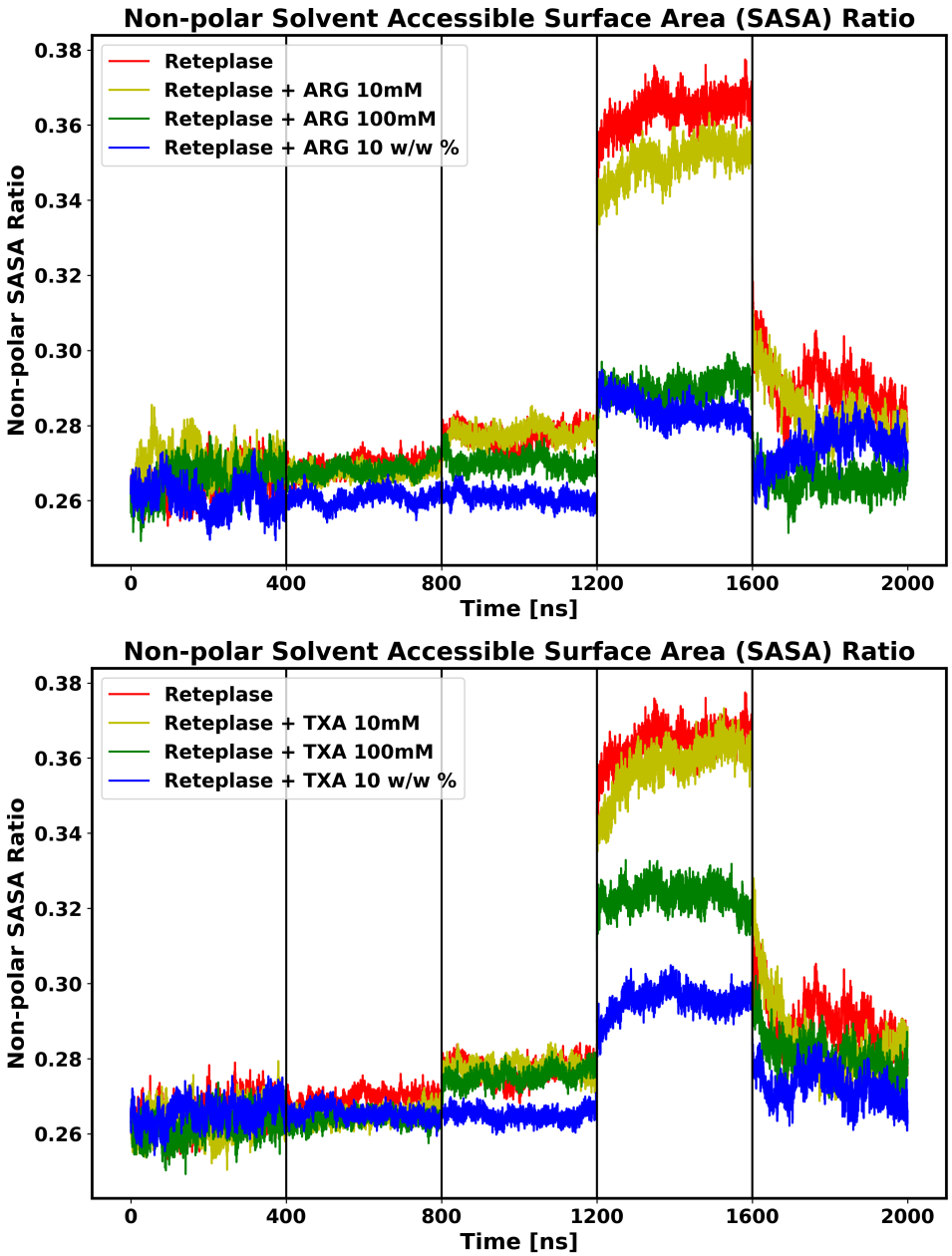


Figure 6.8. Plots displaying the time evolution of the ratio of non-polar to total solvent-accessible surface area during the different stages of the freeze-drying process. The different graphs in each subplot show the data of Reteplase in a system containing the following concentrations of excipient: 10 mM (yellow), 100 mM (green) and 10 w/w % (blue). The red graphs show the data of Reteplase without excipient. The upper subplot corresponds to the systems containing arginine as an excipient and the lower subplot corresponds to the systems containing TXA. Each graph represents the average of the two duplicate simulations of the same system.

The conclusions regarding the trend of total and non-polar SASA are reflected in the data of the non-polar to total SASA ratio (Figure 6.8). Assuming that the total SASA is decreased to greater relative extent compared to the non-polar SASA during the secondary drying, the ratio will increase during this stage. By similar argument, the non-polar SASA increases more than the total SASA during the reconstitution, which is the reason why the ratio is increased. This observation indicates that Reteplase will be more aggregation-prone both during and after the freeze-drying process. However, the destabilization of the protein is significantly reduced in the presence of either arginine or TXA. In this regard, the data implies that arginine is effective at a lower concentration than TXA.

6.2.4 *Spatial Aggregation Propensity*

The spatial aggregation propensity (SAP) was calculated for simulations at room temperature, secondary drying and reconstitution. The SAP algorithm takes into account the solvent-accessible surface area of amino acid side chains weighted by their hydrophobicity factor, and the contribution to instability from structurally nearby residues, as to predict the aggregation propensity score. The computation of SAP was based on the last five frames of the associated trajectories. The plots of SAP are displayed in Figures 6.9 – 6.11, comparing Reteplase without excipient to Reteplase in the presence of either arginine or TXA respectively. For simplicity, only the highest concentration (10 w/w %) is shown. However, it was observed that the stabilizing effect on the protein was improved when increasing the concentration of excipient, as deduced from the increased level of SAP score reduction at certain aggregation prone regions (APRs) marked by vertical lines in the plots. The difference in aggregation propensity between Reteplase in the absence, versus in the presence of excipient, is increased in the dried state of the protein, whereas only a small change is observed at room temperature. Thus, the conformational instability of Reteplase is augmented during the freeze-drying process, which puts emphasis on the necessity of formulation excipients, as to prevent protein unfolding and aggregation. It appears that some APRs are stabilized during reconstitution even in the absence of excipient molecules. However, complete reversibility is not observed, and the benefit of adding excipient in terms of improving stability is greater compared to the initial simulation at room temperature.

In general, arginine performs better in reducing the aggregation propensity of Reteplase during freeze-drying.

A number of common APRs are recognized, where excipient molecules are effective in reducing aggregation propensity. To generalize, these regions comprise the following segments of the protein: Asn5 – Ala15, Lys75 – Ala110, Arg170 – Ile200, Arg220 – Pro250, Thr270 – Asp300 and Leu320 – Cys330. These high SAP regions are recognized by their predominant location in surface loop elements. It is rationale to assume a relationship between structural instability and local APRs, which is investigated in the text further.

By inspection of the plots of SAP and the associated APRs at room temperature (Figure 6.9) and secondary drying (Figure 6.10), it appears that the initial observations from Aggrescan3D (Figure 6.2) are effectively reproduced, which provides support for the validity of the results.

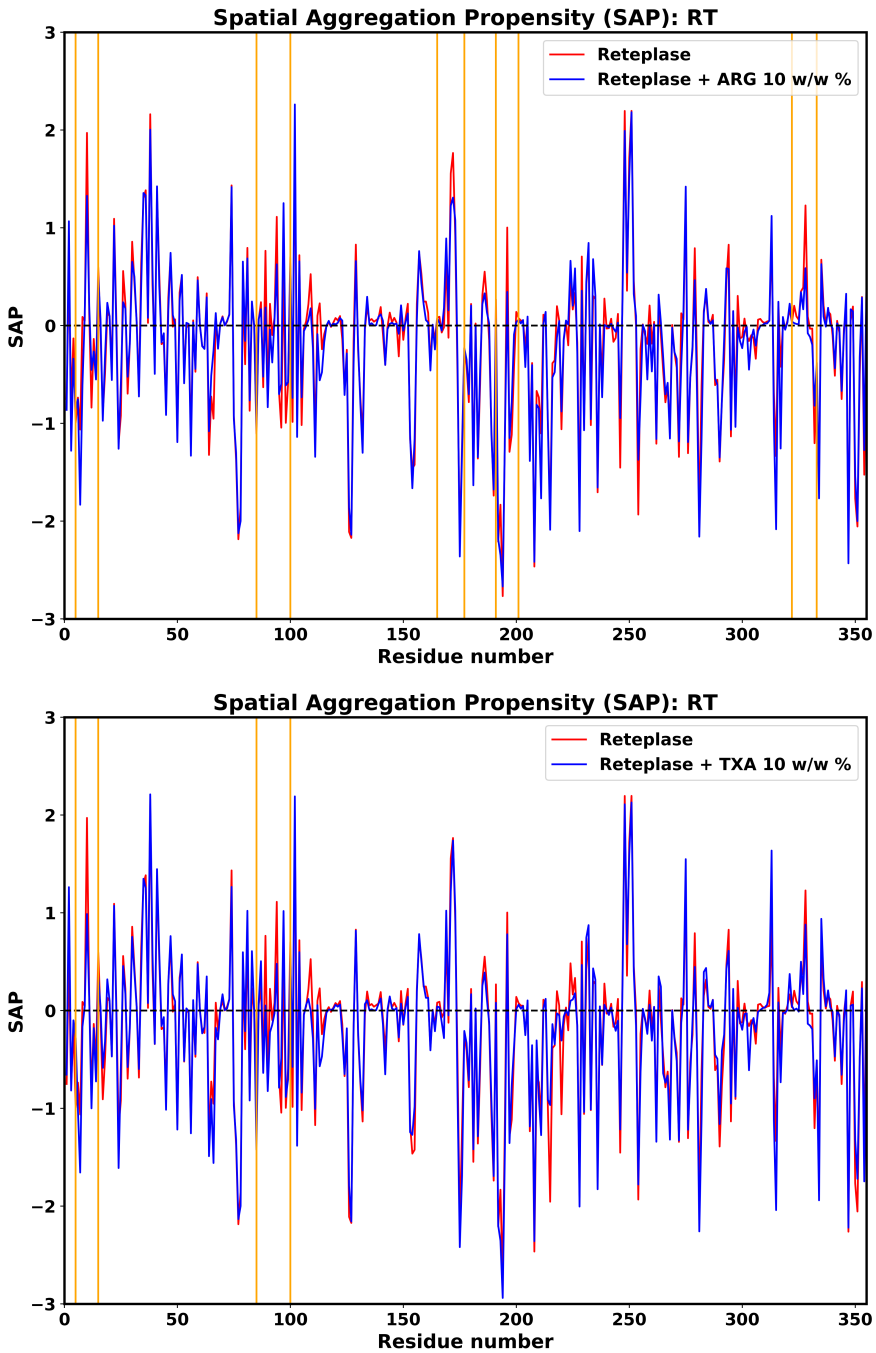


Figure 6.9. Plots displaying the spatial aggregation propensity (SAP) score per residue at room temperature for Reteplase without excipient (red) and Reteplase in the presence of excipient (blue), referring to either arginine (upper subplot) or TXA (lower subplot). Orange vertical lines mark aggregation prone regions where a large difference in SAP score is observed.

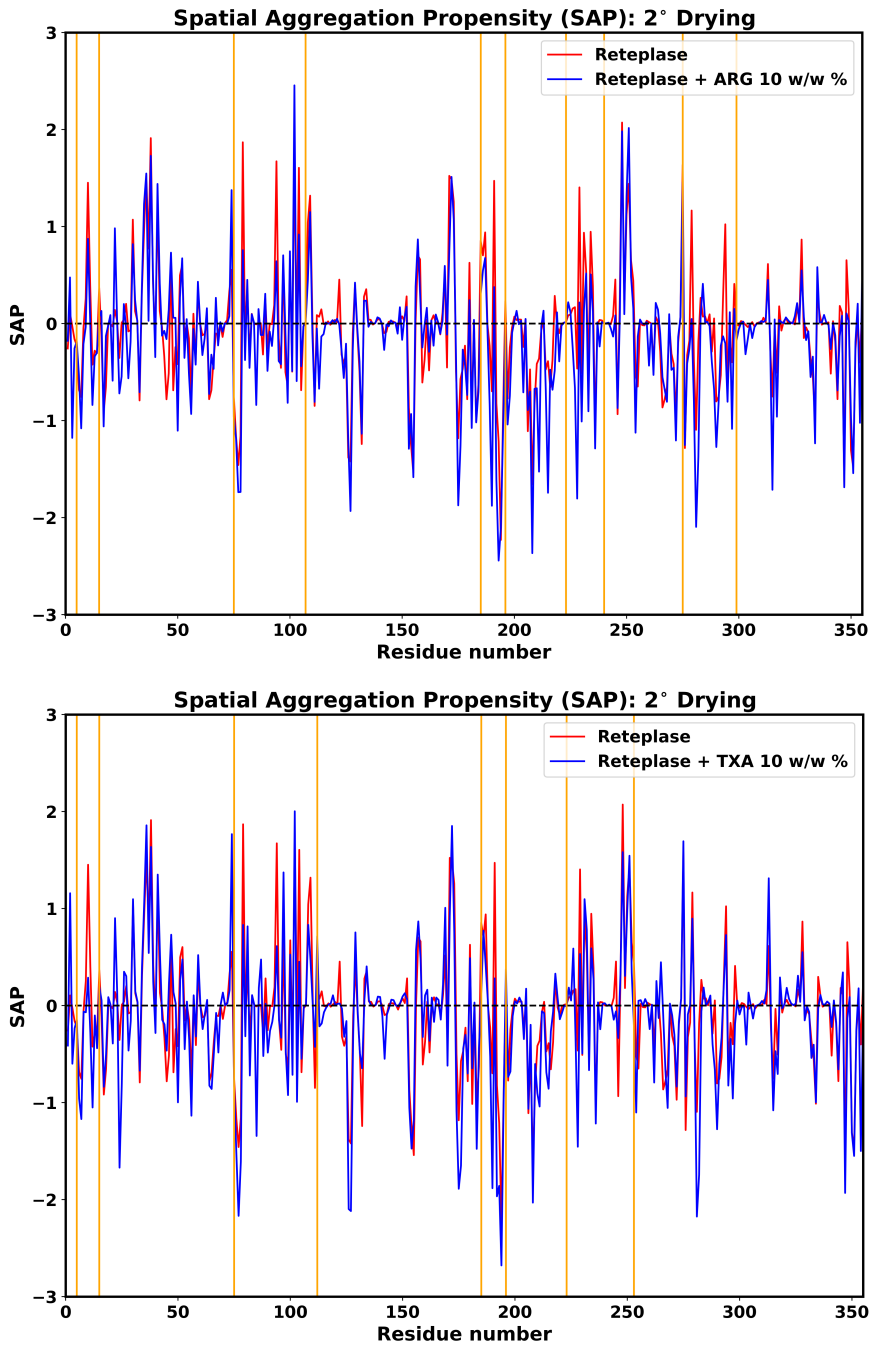


Figure 6.10. Plots displaying the spatial aggregation propensity (SAP) score per residue during the secondary drying stage for Reteplase without excipient (red) and Reteplase in the presence of excipient (blue), referring to either arginine (upper subplot) or TXA (lower subplot). Orange vertical lines mark aggregation prone regions where a large difference in SAP score is observed.

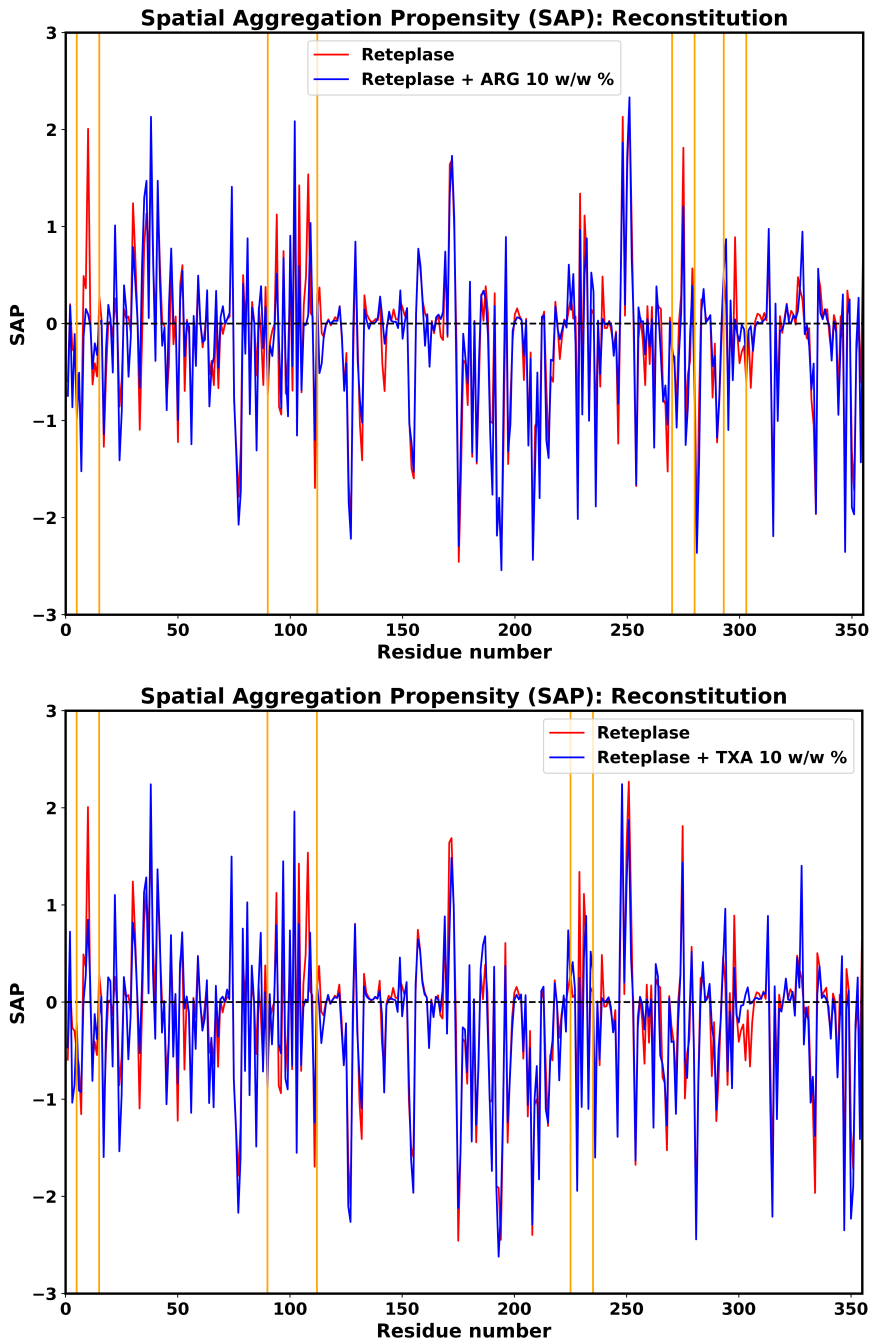


Figure 6.11. Plots displaying the spatial aggregation propensity (SAP) score per residue during the reconstitution stage for Reteplase without excipient (red) and Reteplase in the presence of excipient (blue), referring to either arginine (upper subplot) or TXA (lower subplot). Orange vertical lines mark aggregation prone regions where a large difference in SAP score is observed.

6.2.5 *Excipient Fraction and Preferential Interaction Coefficients*

The average fraction of excipient molecules (x_{ex}) as a function of the distance (r) to the protein surface, as displayed in Figure 6.12, was calculated for simulations at room temperature and freezing according to equation (6.1), where N_{ex} is the number of excipient molecules and N_w is the number of water molecules (91). For simplicity, only the highest concentration (10 w/w %) of arginine, TXA and sucrose are shown. The appearance of the associated graphs indicates the existence of different distributions of excipient molecules in the local and bulk domains around the protein surface respectively, as characteristic of separate stabilization mechanisms.

$$x_{ex}(r) = \frac{\langle N_{ex}(r) \rangle}{\langle N_{ex}(r) \rangle + \langle N_w(r) \rangle} \qquad (6.1)$$

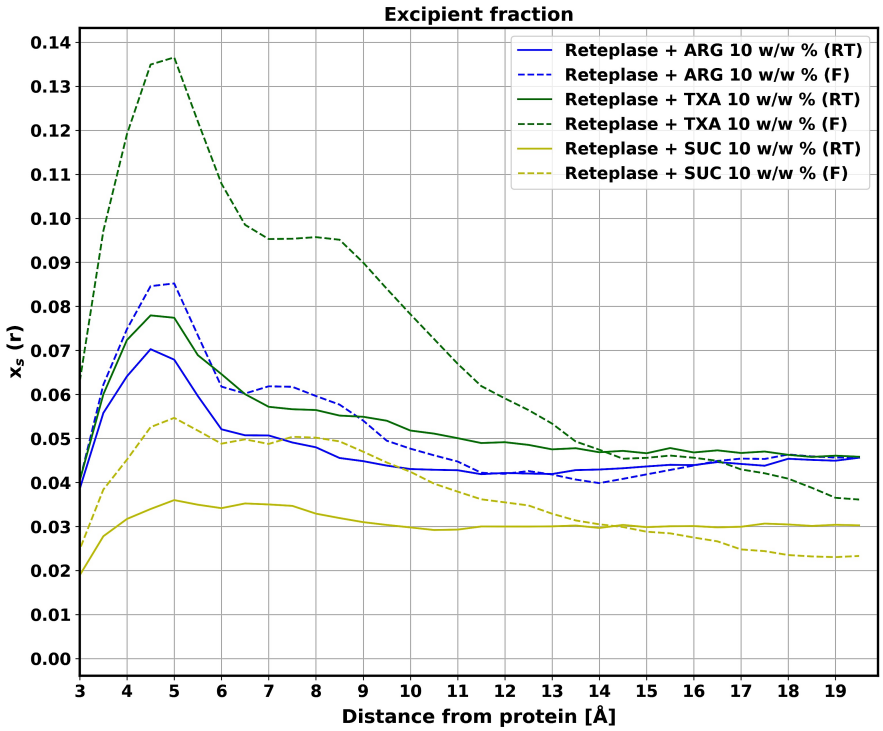


Figure 6.12. Plot of the excipient fraction (x_s) as a function of the distance to the protein surface for simulations of Reteplase in the presence of arginine (blue), tranexamic acid (green) and sucrose (yellow), at room temperature (solid line) and freezing (dashed line) respectively.

In particular, the preferential interaction behavior of the excipient depends on the electrostatic surface potential of the protein, and on whether the excipient molecules contain any charged groups, which may promote favourable Coulombic interactions. The protein surface of Reteplase was found to comprise both negatively and positively charged areas, with the former case predominating the catalytic domain and the latter case predominating the kringle-2 domain. By visual inspection in VMD, arginine was found to bind to the catalytic domain to a higher extent compared to the kringle-2 domain, presumably as a consequence of its positive net charge. However, both arginine and TXA contain positively and negatively charged groups, which can interact with exposed residues of opposite charge on the surface of Reteplase. By contrast, sucrose is an uncharged molecule with non-reducing disaccharide properties. As expected from this argument, the excipient fractions of arginine and TXA are significantly higher in the local domain than in the bulk domain, as indicative of stabilization by preferential binding. For sucrose, the excipient fraction in the local domain is considerably lower compared to the cases of either arginine or TXA, and the absence of a peak in the local domain suggests preferential exclusion from the protein surface. Although, a concentration effect related to the peak height is observed, where 10 w/w % corresponds to a lower molar concentration for sucrose than for arginine and TXA in order, the distribution patterns are still clearly different.

In addition, the tendency of the excipient molecules to self-associate in the bulk influences the balance of preferential interaction with the protein and exclusion from the surface. Moreover, the presence of various functional groups, such as the great number of hydroxyl groups in case of sucrose, might promote other types of interaction with exposed protein residues or serve as to substitute for water molecules in the dried state.

For the simulations of the freezing stage, a trend increase of the excipient fraction in the local domain is observed, which might be a result of freeze-concentration, where the growing ice phase pushes the excipient molecules towards the protein surface.

The plot of the excipient fraction was used to define the boundary distance for the cutoff between the local and bulk domains, as graphically represented by the flattening of the curve. Based on this criterion, the boundary distance was set to 12 Å and 15 Å for simulations at room temperature and freezing respectively, where the latter case equals the boundary of the simulation box. These values were employed for the computation of preferential interaction coefficients (Γ_{23}) according to equation (6.2), where n_3 and n_1 are the numbers of excipient and

water molecules respectively, in the local domain. The total numbers of excipient and water molecules are denoted by and n_3^{total} and n_1^{total} respectively (20). The calculation of preferential interaction coefficients relies on a method developed by Baynes and Trout (92), and describes the excess number of excipient molecules in the local domain relative to the bulk domain. In the theoretical definition of the preferential binding mechanism, the coefficient is greater than zero ($\Gamma_{23} > 0$), whereas for the preferential exclusion mechanism, the coefficient is less than zero ($\Gamma_{23} < 0$) (20). The average values of Γ_{23} are displayed in Table (6.1), along with the corresponding local-bulk partition coefficients (K_p), which were calculated according to equation (6.3), describing the ratio of the number of excipient molecules to water molecules in the local and bulk regions (20).

$$\Gamma_{23}(r,t) = n_3(r,t) - n_1(r,t)\left(\frac{n_3^{total} - n_3(r,t)}{n_1^{total} - n_1(r,t)}\right) \quad (6.2)$$

$$K_p = \frac{(n_3/n_1)^{local}}{(n_3/n_1)^{bulk}} \quad (6.3)$$

Table 6.1. The calculated values of the preferential interaction coefficient (Γ_{23}) and the local-bulk partition coefficient (K_p) for simulations of Reteplase in the presence of arginine (ARG), tranexamic acid (TXA) and sucrose (SUC), at room temperature (RT) and freezing (F) respectively.

Simulation	Excipient	Concentration	Γ_{23}	K_p
RT	ARG	10 w/w %	10.8	1.18
	TXA	10 w/w %	79.7	3.06
	SUC	10 w/w %	2.2	1.08
F	ARG	10 w/w %	55.7	2.36
	TXA	10 w/w %	128.8	6.23
	SUC	10 w/w %	38.5	3.13

By the definition of positive average values of Γ_{23}, all of the excipients display behavior of preferential binding to the protein surface. However, sucrose shows signs of preferential exclusion as well, with data of Γ_{23} spanning positive and negative values. Literature suggests that the stabilization mechanism of sucrose is mainly attributed to preferential exclusion (3) (16) (5). The FF parameter set for sucrose was specifically chosen as to reproduce the experimental data of Γ_{23} for sugar alcohols. Nevertheless, it is a non-trivial task to obtain a balance between stickiness and repulsion of small molecules. Furthermore, an accurate description of preferential exclusion behavior would likely require a larger simulation box size than the one being used in this project.

To conclude, a relative comparison of the magnitudes of Γ_{23} and the graph appearances in the excipient fraction plot, indicates that the main stabilization mechanism of arginine and TXA is accredited to preferential binding, while for sucrose preferential exclusion is more relevant. The observed type of preferential interaction is likely a combination of the surface charge of the protein and properties of the small molecules such as polarity and molecular size. For arginine, there is a strong tendency of self-assembly into molecular clusters (15) (20), which was also observed by visual inspections in VMD. Arginine cluster formation effectively reduces the interaction with exposed residues at the protein surface, and might partially explain its less preferential interaction with Reteplase compared to TXA. These conclusions are further confirmed in the data of K_p, as showing an excess of TXA molecules in the local domain compared to the bulk environment, while arginine and sucrose and more equally distributed. As previously discussed, freezing of the system concentrates non-crystallizing small molecules in the remaining liquid phase, which has the effect of increasing the excipient fraction in the local domain, and thus generating higher Γ_{23} values.

6.2.6 *Protein-Excipient Interaction Score*

The protein-excipient interaction (PI) score, as defined by the percentage of frames of the associated trajectory, where a distance less than 3 Å is registered between a protein residue and an excipient molecule, was calculated for the last 200 ns of the simulations at room temperature. The results of the PI score are displayed for four different categorizations of protein residues: all residues (Figure 6.13), negatively and positively charged residues (Figure 6.14) and hydrophobic residues (Figure 6.15). For simplicity, only the PI scores for the highest concentration (10 w/w %) of arginine, TXA and sucrose are shown.

By inspection of the PI score for all protein residues (Figure 6.13), an excipient concentration effect is evident. TXA displays the overall highest PI score, as correlated with the strongest interaction with the protein, followed by arginine and sucrose in decreasing order. These observations are in agreement with the expected results due to the difference in molar concentration. However, the different physical and chemical properties of the individual excipients might have a strong impact as well. Some regions of the protein appear particularly prone to interaction, with some variations in between the different excipients. From the data of R_G and SASA reported previously, it is known that the stabilizing effect of TXA is more concentration-dependent than for arginine.

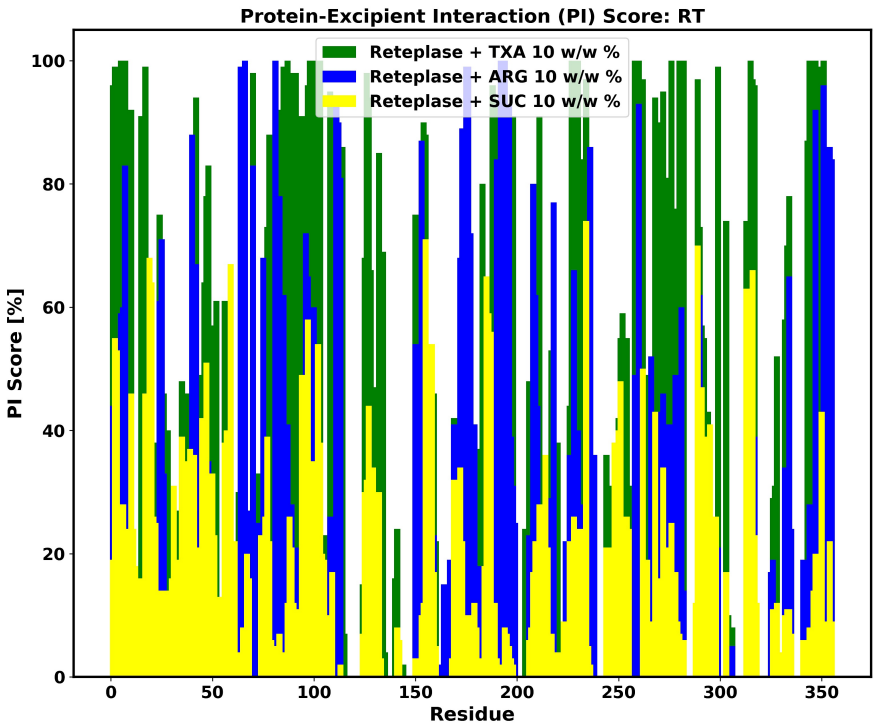


Figure 6.13. Plot of the protein-excipient interaction (PI) score per residue displayed for the highest concentration (10 w/w %) of TXA (green), arginine (blue) and sucrose (yellow) during simulations at room temperature (RT).

The PI score displayed for charged residues of the protein (Figure 6.14) states a preference for salt bridge interaction with TXA and arginine, which is expected from the presence of charged groups in either of the two excipients. As a consequence, sucrose displays considerably less interaction with these residues. TXA contains both positively and negatively charged groups and is expected to interact with residues of opposite charge to similar extent, which is in line with observations. By contrast, arginine has a positive net charge and is predicted to interact preferably with negatively charged residues, which is in accordance with the observed results. Sucrose tend to interact more with positively charged residues, which might be related to ion-dipole interaction relying on the presence of partial negative charges located to the oxygen atoms of the multiple hydroxyl groups of sucrose.

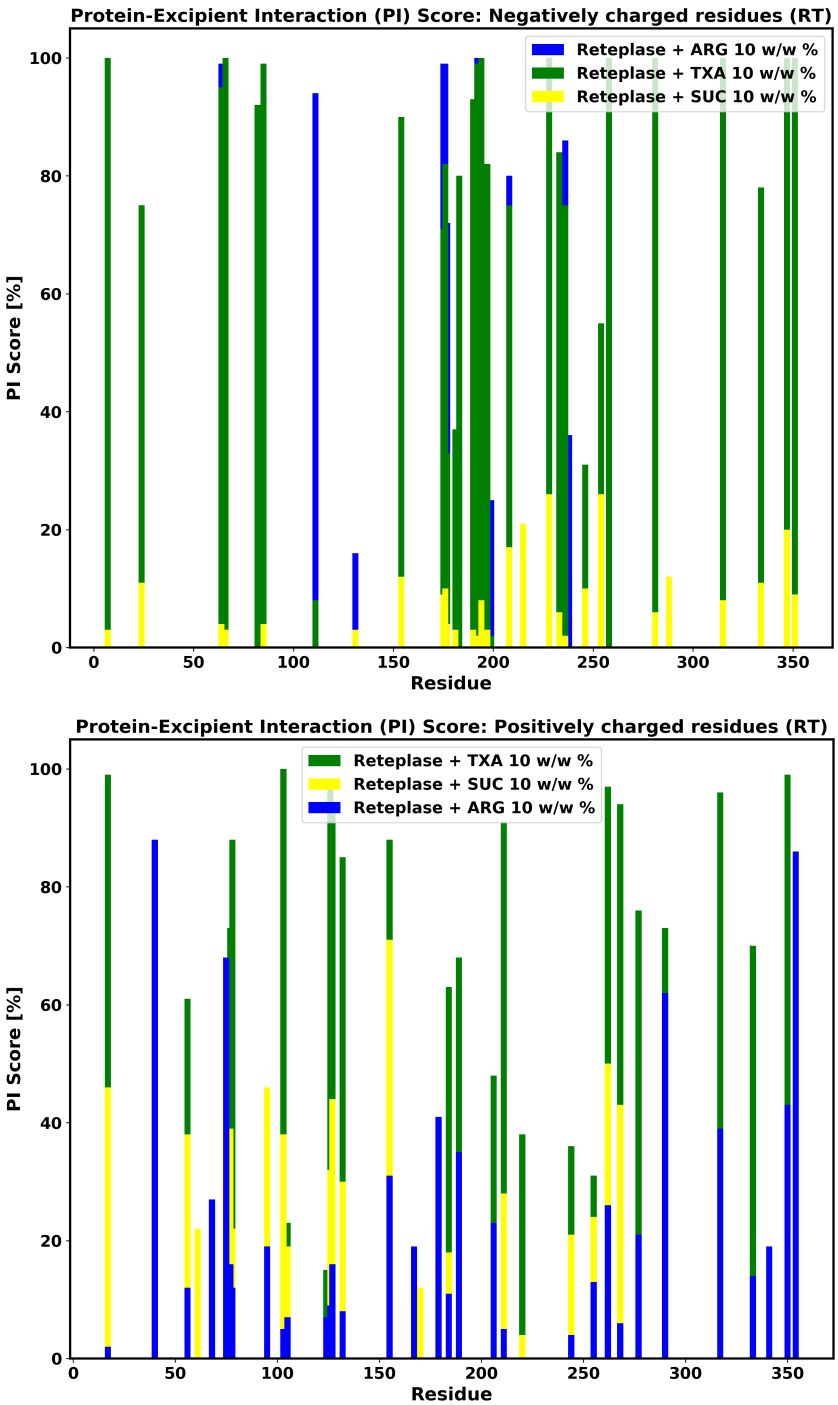


Figure 6.14. Plot of the protein-excipient interaction (PI) score per charged residue displayed for the highest concentration (10 w/w %) of TXA (green), arginine (blue) and sucrose (yellow) during simulations at room temperature (RT). The upper subplot shows the PI score for negatively charged residues and the lower subplot show the PI score for positively charged residues.

According to the results displayed in Figure (6.15), hydrophobic interactions comprises a small contribution to the overall protein-excipient interactions. However, TXA displays a generally high frequency of interaction irrespective of residue character. This property of TXA might be a consequence of its possession of functional groups with varying polarity, its high accessibility to cavities on the protein surface due to its small molecular size and the high concentration applied. Sucrose interacts to similar extent with either of positively charged or hydrophobic residues and offers more selective interactions. As a consequence of its multiple sugar alcohol groups, hydrogen bonding is presumed to be another important type of interaction with the protein. The weak interaction of arginine with hydrophobic residues is reasonable considering its low alkyl chain content and the higher relevance of charged interaction for its stabilization of protein, as supported elsewhere in literature (19).

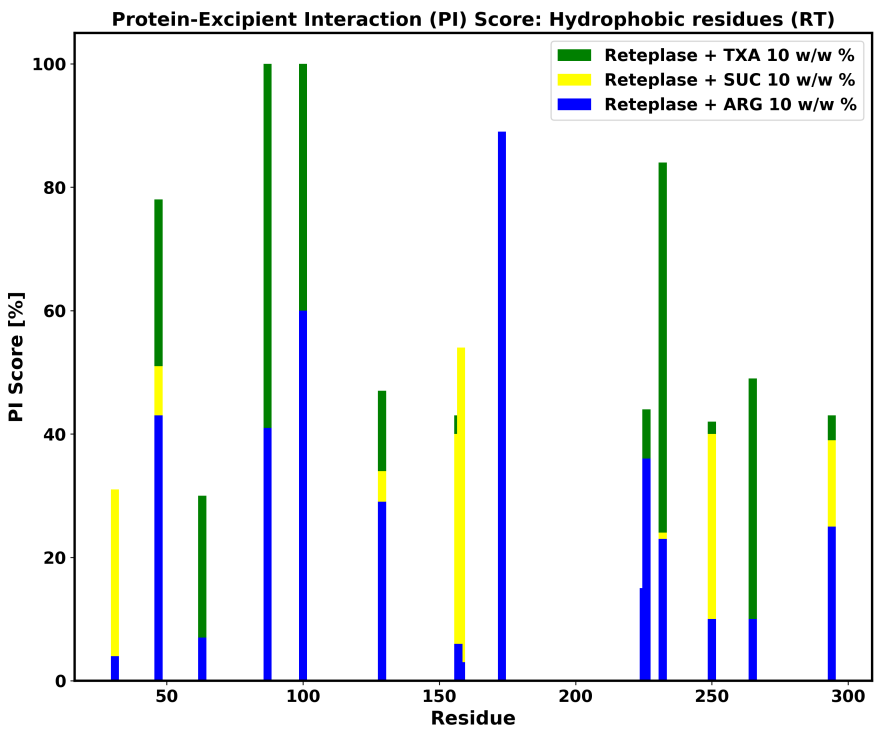


Figure 6.15. Plot of the protein-excipient interaction (PI) score per hydrophobic residue displayed for the highest concentration (10 w/w %) of TXA (green), arginine (blue) and sucrose (yellow) during simulations at room temperature (RT).

It has been proposed elsewhere (5) that selective interactions between excipient molecules and certain exposed residues of the protein, which possess a high aggregation propensity, may act to confer improved stability through the shielding of these critical areas to the surrounding. A comparison of local regions of the protein displaying a significant reduction of SAP score in the presence of excipient, to spans of the protein primary structure where a high PI score is registered, shows an advocate degree of correlation and may reveal specific stabilizing protein-excipient interactions. For arginine, peaks of high PI score are recorded within the intervals Asn5 – Ala15 and Arg170 – Ile200, which are recognized by a considerable reduction of SAP in the presence of 10 w/w % of the excipient at room temperature simulation. For TXA, the first region and the interval Asp85 – Pro100 display the same correlation. A visual representation of protein regions that interact to great extent with either arginine or TXA, and the associated residues experiencing an accompanying sharp reduction in SAP score in the presence of these excipients, is displayed in Figure 6.16.

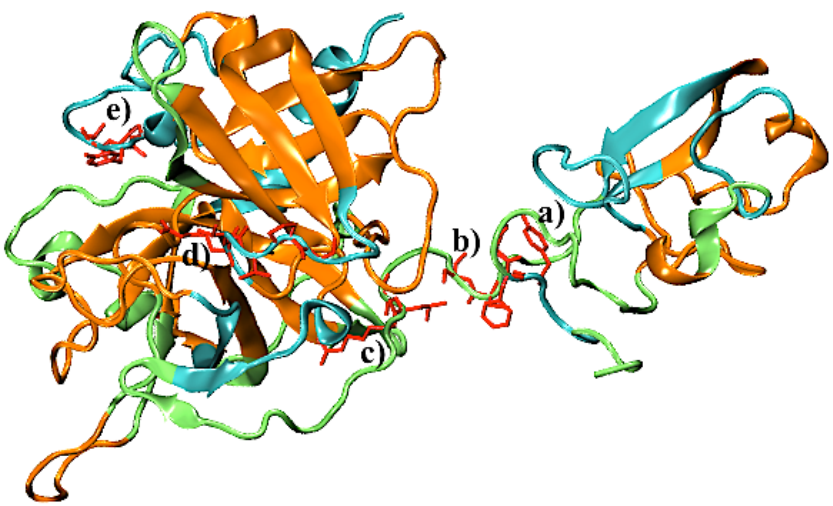


Figure 6.16. Visual representation of Reteplase in cartoon drawing style displaying protein regions with high PI score with respect to arginine (blue) and TXA (green). The residues located in these regions that also experience a significant reduction of SAP score in the presence of excipients, as assigned in Table 6.2, are displayed in red using licorice drawing style. The residue intervals refer to a) Tyr9 – Phe10 – Gly11, b) Ser88 – Cys89 – Ser90, c) Gly93 – Leu94 – Arg95, d) Arg170 – Val171 – Val172 – Pro173 – Gly174 and e) Thr195 – Tyr196 – Asp197. The figure is rendered in VMD.

Through further fine-tuning, the regions that demonstrate a correlation between strong protein-excipient interactions and a reduction of SAP, specifically concern the sequential residues displayed in Table 6.2, which are all located in surface loops. It is logical that these elements are susceptible to conformational change and that specific interactions with excipient molecules might have an effect on the overall tertiary structure of the protein. In this regard, the loop region connecting the two domains of Reteplase, the activation peptide (Ser90 – Arg103), might be of particular importance. From Table 6.2, it can be seen that TXA interacts strongly with sensitive residues within this segment. By visual inspection in VMD presented in association with the data of R_G, it was concluded that TXA had the most pronounced effect on preserving the spatial separation between the kringle-2 domain and the catalytic domain. Thus, it is plausible that a relationship exists between local APRs and tertiary structure instability, which might be reduced by selective stabilizing protein-excipient interactions.

Table 6.2. Residue spans where a correlation is observed between a significant reduction of SAP score and a high PI score (+) with respect to either arginine or TXA.

Residues	ARG	TXA
Tyr9 – Phe10 – Gly11	+	+
Ser88 – Cys89 – Ser90		+
Gly93 – Leu94 – Arg95		+
Arg170 – Val171 – Val172 – Pro173 – Gly174	+	
Thr195 – Tyr196 – Asp197	+	

The residues referred to with interaction with arginine suggest possibilities for hydrogen bonding, cation–π interaction, guanidinium stacking and salt bridges. For TXA, the relevant interactions would be mainly hydrogen bonding and salt bridges. The proposed selective stabilizing interactions of arginine and TXA are displayed in Figure 6.17.

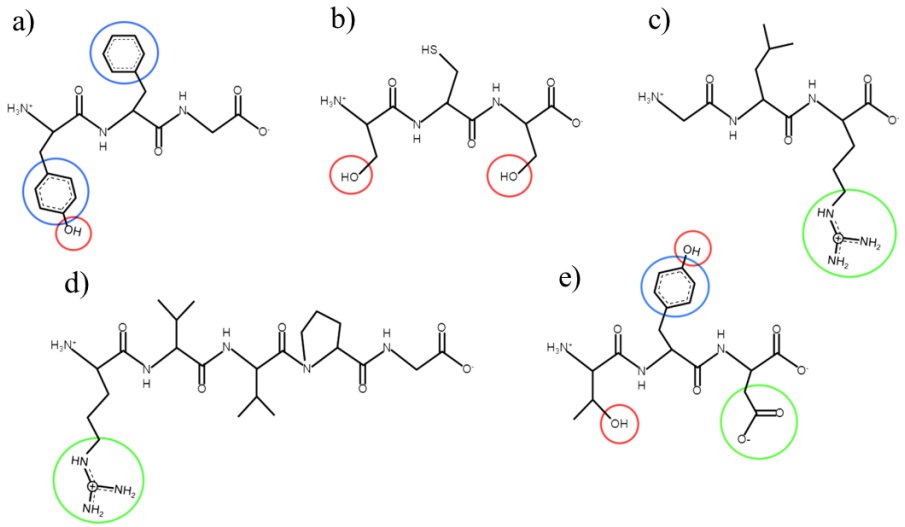


Figure 6.17. The primary structure of the amino acid residues referred to in Table 6.2, showing a) Tyr9 – Phe10 – Gly11, b) Ser88 – Cys89 – Ser90, c) Gly93 – Leu94 – Arg95, d) Arg170 – Val171 – Val172 – Pro173 – Gly174 and e) Thr195 – Tyr196 – Asp197. The functional groups for suggested interactions with arginine (a, c and d) and TXA (a, b and c) are highlighted. The associated protein-excipient interactions are proposed as salt bridges (green), hydrogen bonding (red) and cation–π interaction (blue).

6.3 Comparative Analysis of Granulocyte Colony- stimulating Factor

6.3.1 *Electrostatic Surface Potential Map*

An inspection of the electrostatic potential molecular surface of G-CSF using the Adaptive Possion-Boltzmann Solver (APBS) calculation program of PyMOL 2.0 clearly demonstrated a difference in charge distribution depending on the pH of the solution (Figure 6.18). At pH 4, the surface potential of G-CSF is mainly characterized by positively charged areas, whereas at pH 7.5 there is a dominance by negatively charged sites. This observation is likely to have an effect on the preferential interaction with excipient molecules, and thus on the induced stabilization mechanism.

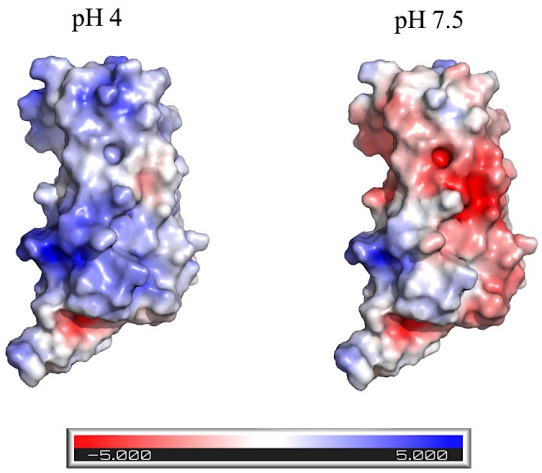


Figure 6.18. The electrostatic surface potential map for the PQR input structure of G-CSF at pH 4 (left) and pH 7.5 (right), as calculated using the Adaptive Poisson-Boltzmann Solver (APBS) software plugin of PyMOL 2.0. Red and blue areas display negatively and positively charged surface patches respectively.

6.3.2 *Protein Conformational Change during the Simulation of Freeze-drying*

A time series analysis of R_G during the freeze-drying simulation of G-CSF (Figure 6.19) reveals a similar trend of increased protein compactness in the dried state, as observed for Reteplase. However, an important difference lies in the observation of structural reversibility during reconstitution independent of the presence of excipient. In this regard, the less complex tertiary structure of G-CSF is likely to have an impact. The features of a compact single-domain protein structure, a high helical content and few loop regions are plausible contributing factors.

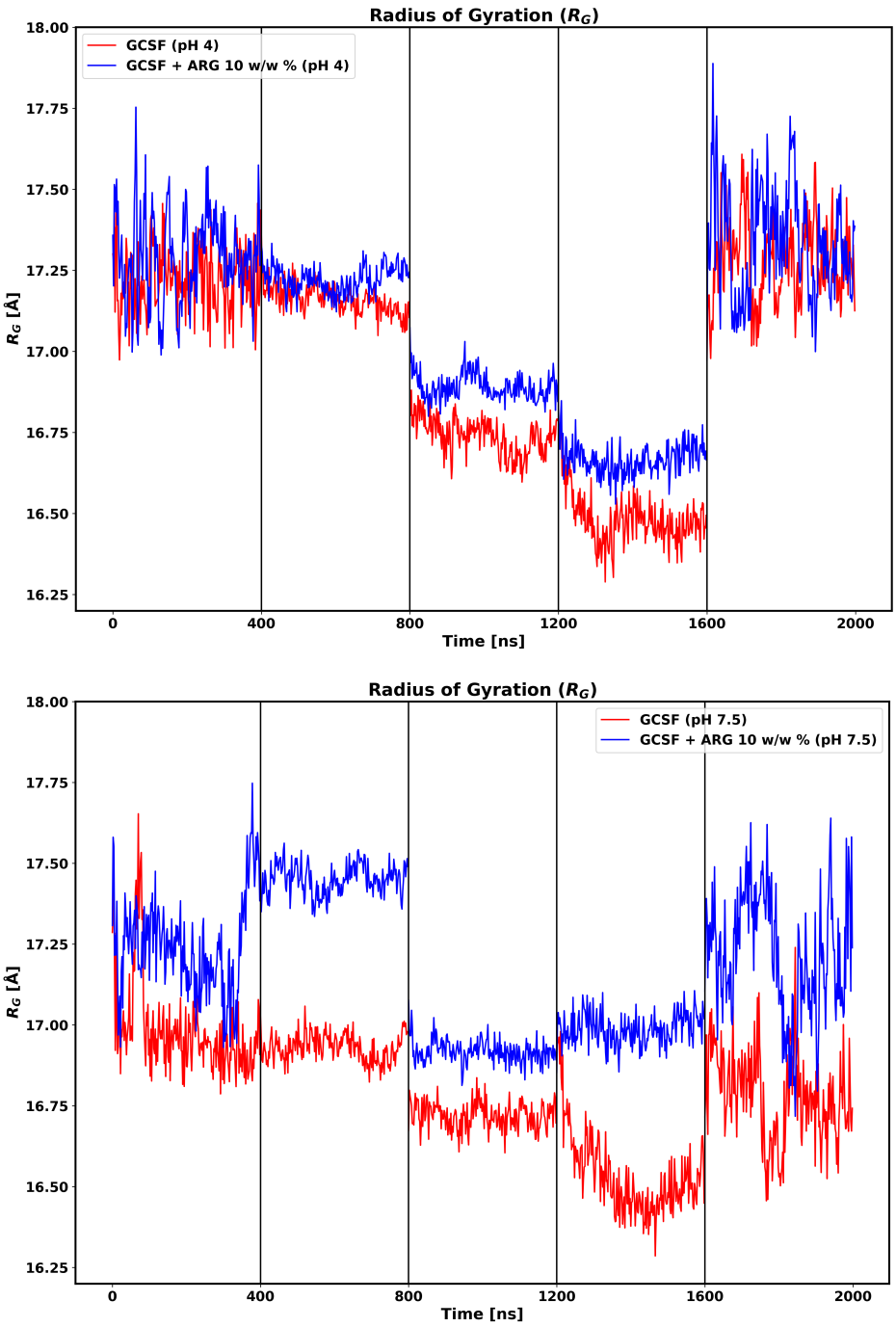


Figure 6.19. Plots displaying the time evolution of the radius of gyration during the different stages of the freeze-drying process of G-CSF in the absence of excipient (red) and in the presence of 10 w/w % arginine (blue). The upper and lower subplots correspond to systems at pH 4 and pH 7.5 respectively.

6.3.3 *Excipient Fraction*

The distribution of arginine (10 w/w %) in the local and bulk domains of G-CSF is displayed graphically in the excipient fraction plot in Figure 6.20. Evidently, the fraction of arginine molecules within the local domain is higher at pH 7.5 compared to at pH 4. This observation is likely a consequence of the nature of electrostatic interactions between net positively charged arginine molecules and residues located at the protein surface. At pH 4, arginine is repelled by positively charged amino acid side chains, and thus preferentially excluded from the surface. By contrast, at pH 7.5 the protonation state is switched and arginine is attracted by the negative charge of exposed residues, which produces the excipient distribution associated with preferential binding. These observations further emphasize relevance of charge specificity in terms of preferential interaction mode.

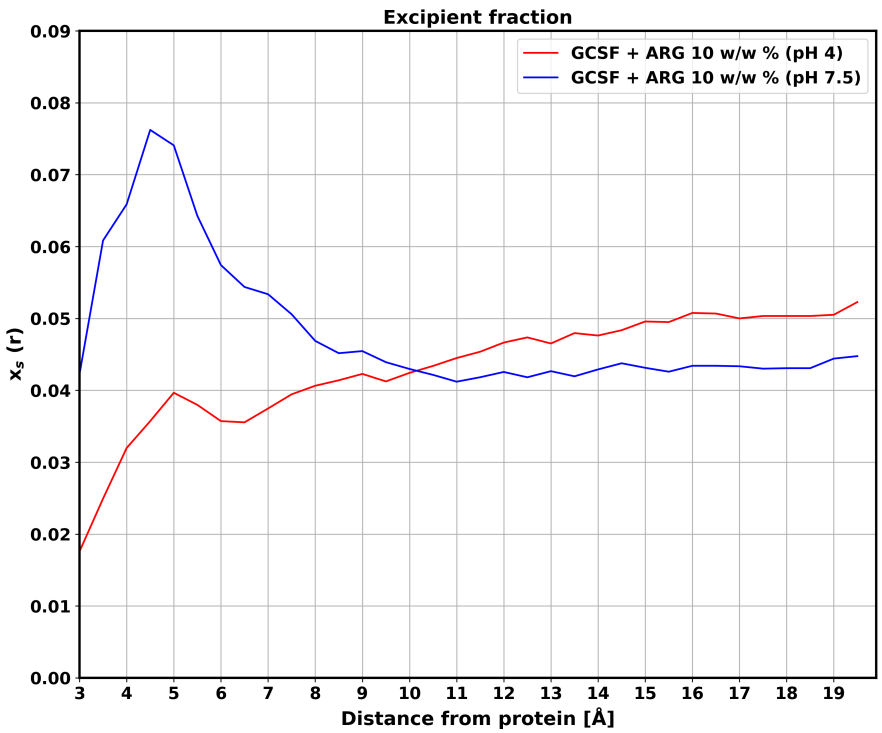


Figure 6.20. Plot of the excipient fraction (x_s) as a function of the distance to the protein surface for simulations at room temperature of G-CSF in the presence of 10 w/w % arginine, at pH 4 (red) and pH 7.5 (blue) respectively.

6.4 Cluster Analysis of Accelerated Molecular Dynamics Simulations

Normal boost and high-boost aMD simulations of Reteplase in the absence or in the presence of arginine (10 w/w %) were analyzed by means of cluster analysis, where the observed protein conformations in each simulation were divided into three different clusters based on structural similarity, as displayed in Figure 6.21. The three clusters within each total ensemble were aligned to the respective input structure, which was used as the reference for computation of RMSD. The aMD simulations performed in this context served mainly as a screening tool for providing a basic idea of the conformational space of the protein.

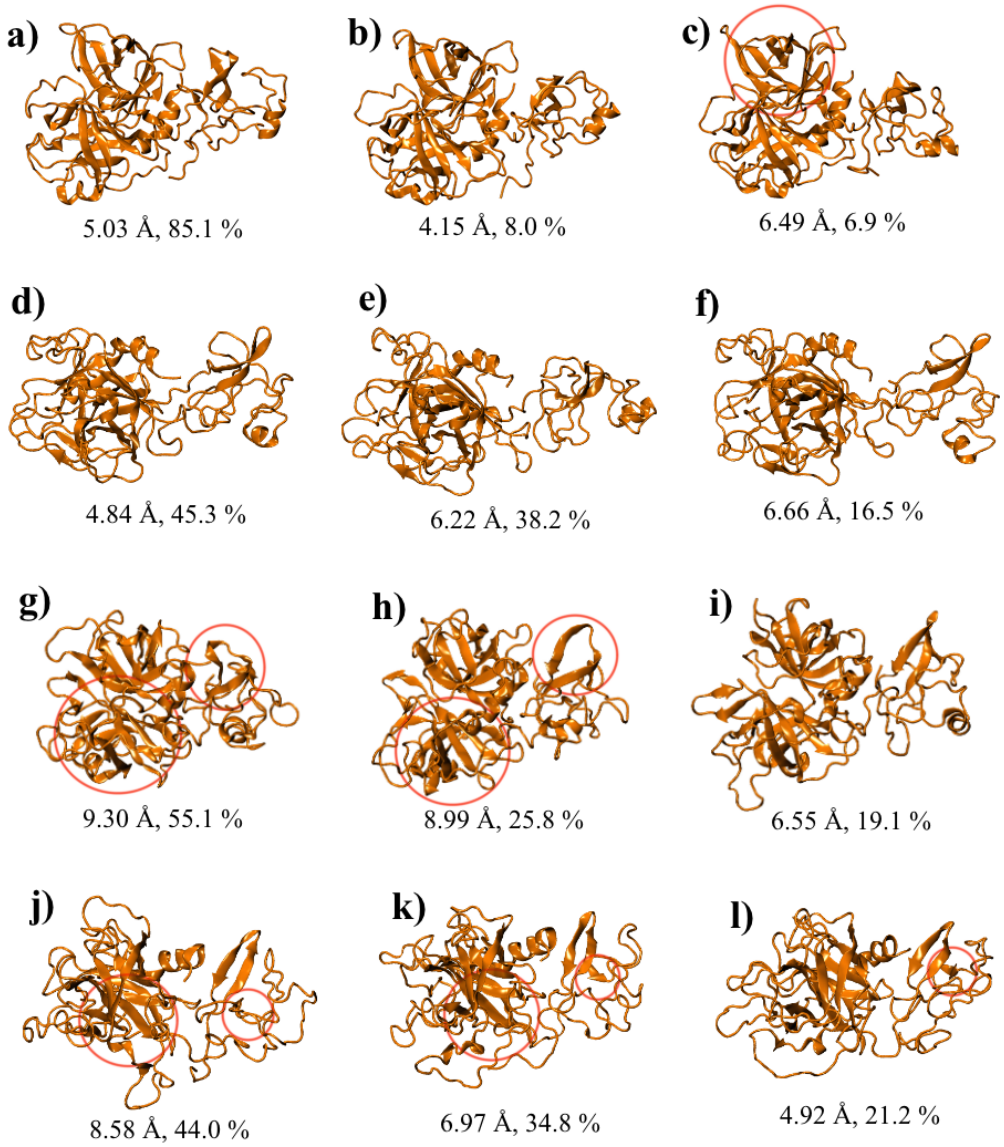


Figure 6.21. Visual representation of the three different protein conformations within each total ensemble as obtained from cluster analysis of the aMD simulation trajectories. Subfigures a) – c) normal boost aMD simulation of Reteplase without excipient, d) – f) normal boost aMD simulation of Reteplase in the presence of arginine (10 w/w %), g) – i) high-boost aMD simulation of Reteplase without excipient, j) – l) high-boost aMD simulation of Reteplase in the presence of arginine (10 w/w %). RMSD was calculated using the respective input structure as a reference. The data of RMSD of the associated cluster conformation is shown below each subfigure, along with the fraction of the cluster relative to the total ensemble. Red circles mark secondary structure elements that were destroyed during the simulations. Figures rendered in VMD.

A low to modest degree of conformational change was observed among the clusters obtained from normal boost aMD simulations, mainly affecting the tertiary structure of the protein. For Reteplase without excipient, a partial destruction of one of the two β-barrels in the catalytic domain was observed for the least prevalent conformational cluster (Figure 6.21 c)), as associated with an energetically unfavourable conformation. For Reteplase in the presence of arginine, the conformational change was attributed to an extension of the activation loop connecting the two domains, which acts to increase their separation. This effect is most notable in Figure 6.21 e) and f), which explain the relatively high values of RMSD associated with these conformations.

The high-boost aMD mode was specifically designed with the purpose of crossing the energetic barriers associated with partial unfolding of the protein. For Reteplase both in the absence and in the presence of arginine, one of the β-barrels was structurally impaired in the clusters associated with the two highest fractions within each of the total ensembles (Figures 6.21 g), h), j) and k)). In addition, the β-sheet in the kringle-2 domain was practically destroyed for the case of Reteplase without excipient, (Figures 6.21 g) and h)). This observation was not monitored when arginine was added to the system. However, in return the short α-helix in the kringe-2 domain was unfolded (Figures 6.21 j) – l)).

To summarize, the observed level of partial protein unfolding, as defined by the number of destroyed secondary structure elements, was increased for high-boost aMD compared to normal boost aMD, as expected from the greater amount of added potential energy. In this regard, the β-barrel motifs in the catalytic domain, and the β-sheet or the α-helix in the kringle-2 domain appear to be the most sensitive structural elements. By comparison, the overall protein structure of Reteplase is virtually conserved during the dehydration step of the freeze-drying process. Consequently, the protein structure of Reteplase displays a high extent of conformational stability during the simulated time span of cMD simulation, and any conformational change is more related to an increased compactness of the protein. Thus, the observed aggregation propensity of Reteplase is more likely a result of colloidal instability due to electrostatic forces, rather than exposure of hydrophobic regions by means of protein unfolding.

Given the conditions applied and the time of observation, any structural deformation requires a bias potential energy boost being induced to the system by aMD simulation. However, the uncertainty to whether the energy landscape has been fully explored should be taken into account in this context.

6.5 Protein-protein Interaction Heatmap Analysis of Coarse-grained Molecular Dynamics Simulations

Preliminary results from CG simulations using SIRAH 2.0 force field indicate a tendency of arginine to improve the colloidal stability of Reteplase by reducing protein-protein interactions. The hypothesis relies on a relative comparison from analysis of interaction heatmaps associated with Reteplase in the absence of excipient (Figure 6.22) or in the presence of arginine (Figure 6.23). An interaction is defined by a distance less than 4 Å between individual residues in two separate protein monomers. The frequency of unique protein-protein interactions is normalized to the strongest interaction, which displays the maximum number of interaction events registered throughout the simulation trajectory. However, a complete analysis was not feasible within the frames of this project due to time constraints. Moreover, evidence by statistical significance would require a greater number of repetitions to be conducted for each system. Nevertheless, the initial study has shown that the addition of arginine reduces the interactions that are involved in the aggregation process. Thus, a similar system setup could be used to run CG simulations for a longer time interval in order to thoroughly study the molecular mechanisms of colloidal stabilization by the relevant formulation.

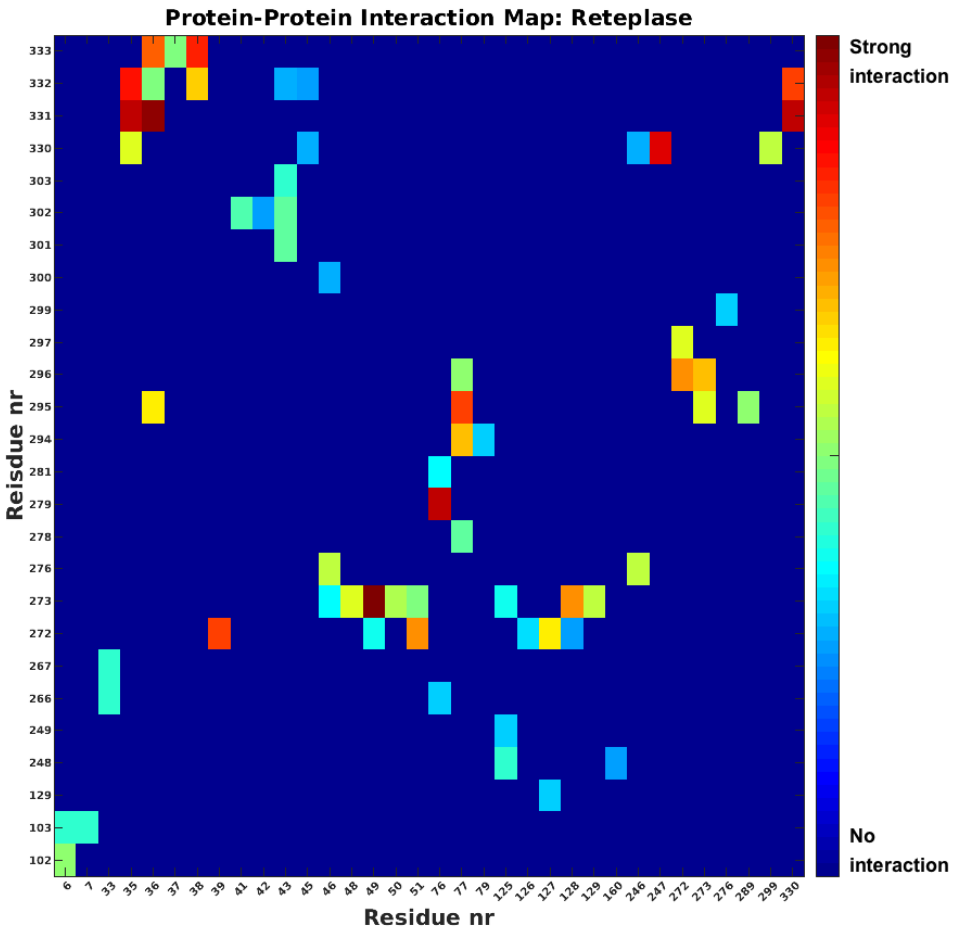


Figure 6.22. Protein-protein interaction heatmap for Reteplase in the absence of excipient. The residue-residue pairs associated with the registered interaction sites in separate protein monomers are determined from the (x, y) coordinates. The colour bar displays the relative registered frequency of the interactions, where red is the strongest interaction and blue is the weakest interaction.

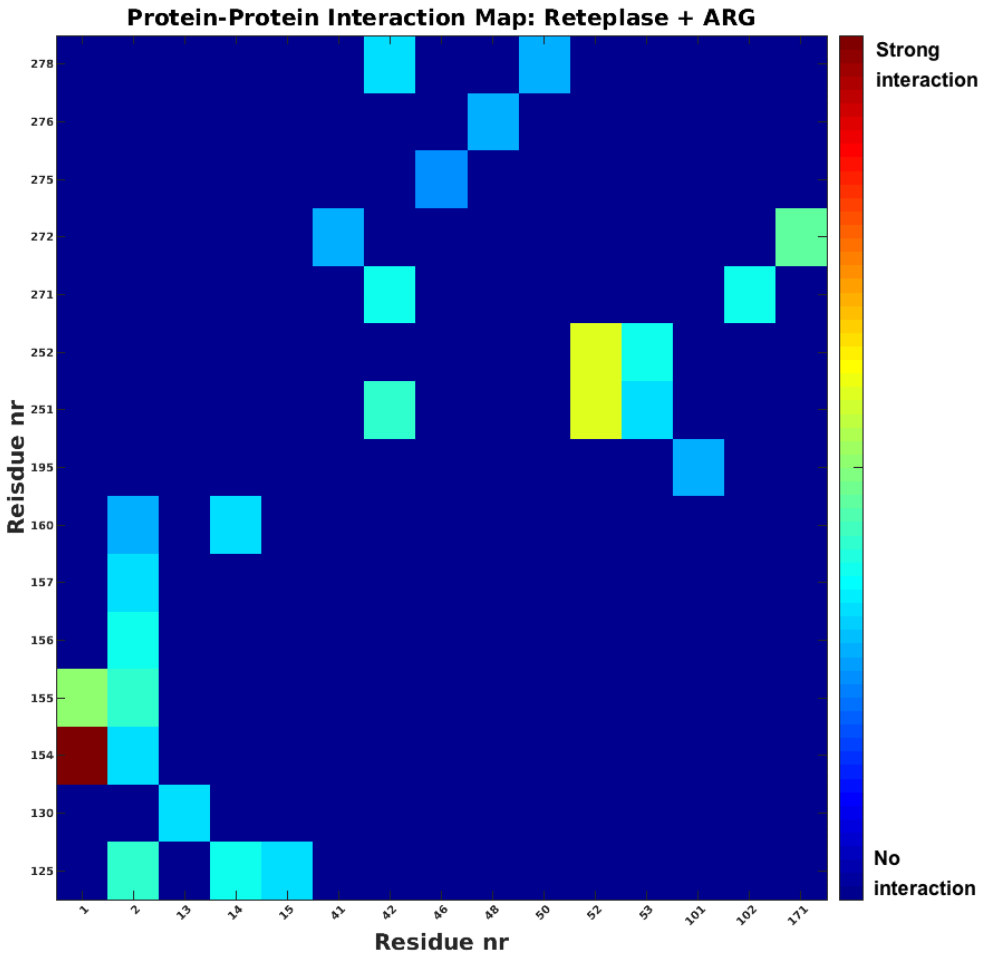


Figure 6.23. Protein-protein interaction heatmap for Reteplase in the presence of arginine (10 w/w %). The residue-residue pairs associated with the registered interaction sites in separate protein monomers are determined from the (x, y) coordinates. The colour bar displays the relative registered frequency of the interactions, where red is the strongest interaction and blue is the weakest interaction.

Chapter 7. CONCLUSION

The process of freeze-drying introduces various stresses to the structure of Reteplase and the addition of formulation excipients is essential to ensure protein stability. By experimental observation, Reteplase is characterized by poor colloidal stability and commercially available formulations include either arginine or TXA as stabilizers. The molecular level study of the protein stability of Reteplase during freeze-drying, as performed in this project by means of molecular dynamics simulation, reveals a concentration-dependent effect of excipient in preservation of the native structure.

Investigations of protein structural descriptors during the different cycles of the lyophilization process demonstrate an increased level of compactness of the tertiary structure during the critical step of dehydration, which is effectively reduced in the presence of excipient molecules, whereas secondary structure elements are virtually conserved. The here mentioned observations have also been seen in studies of Tenecteplase (16). Furthermore, the addition of arginine or TXA induces structural reversibility during reconstitution in terms of recovering the original protein dimensions, as associated with the active conformation of Reteplase. A comparative study of the performance of arginine and TXA suggests that arginine is effective at a lower concentration than TXA. However, the interaction with the Reteplase is stronger for TXA than for arginine, and TXA demonstrates an improved ability of preserving the inter-domain separation of the protein in the dried state.

Studies of the aggregation propensity of Reteplase, as related to the conformational stability, display a reduction of SAP score at certain APRs in the presence of either arginine or TXA, where the effect is more pronounced in the dried state of the protein. A correlation is shown between the reduction of SAP and selective interactions between excipient molecules and exposed protein residues, which are located to surface loop regions. The trend observed hereof is proposed in literature (5). Characterization of the relevant amino acid residues suggests that the stabilizing interactions with arginine are related mainly to salt bridges, hydrogen bonding and cation–π interaction, where the two former cases are also likely for interaction with TXA. In general, salt bridges tend to be the dominating type of interaction between Reteplase and arginine or TXA. In this regard, arginine has a preference for interaction with negatively charged residues due to its positive net charge. TXA displays a higher frequency of interaction, which

could be explained by its diversity of different functional groups and its small molecular size, which enables access to cavities located on the protein surface. In addition, the higher molar concentration of TXA has a logical impact on the number of registered protein-excipient interactions. Sucrose displays more global interactions, which is expected from its absence of charged groups. It is likely that hydrogen bonding plays a crucial role in its interactions with protein due to its multitude of hydroxyl groups.

The excipient fraction distribution of arginine and TXA suggests protein stabilization by preferential binding to exposed residues, while preferential exclusion from the protein surface is proposed as the main stabilization mechanism for sucrose. The hypothesis is further confirmed by the associated preferential binding coefficients. The high relative fraction of TXA in the local domain might be explained by its chemical and physical properties, as described above. The strong tendency of arginine for self-interaction to form molecular clusters in the bulk domain contribute to a reduction of its protein affinity. Arginine oligomerization is mainly attributed to interactions via the guanidinium group, either via hydrogen bonding or by a stacking mechanism (21) (22). The latter type would be less relevant in the dried state of the protein due to the low content of water molecules, which are required for an energetically favoured stacked complex (22).

The charge specific, rather than protein specific, basis for the mode of stabilization is assessed by comparative analysis of G-CSF. Arginine binds preferentially to G-CSF at pH 7.5, when the protein surface is mainly negatively charged, while repulsion occurs at pH 4, as associated with an excess distribution of positive surface potential. The impact of the relative number and the protonation state of charged residues on the strength and mode of preferential interaction with arginine is confirmed in studies of other proteins (20). These observations might also explain the mainly preferential exclusion of sucrose, as a consequence of its absence of charged groups. The structural reversibility, as observed for G-CSF but not for Reteplase during reconstitution in the absence of excipients, is likely affected by the level of protein complexity. G-CSF comprises a single domain, whereas Reteplase is a structurally intricate two-domain protein. However, in this study the addition of arginine or TXA has proven effective in refolding and improving the preservation of an adequate separation between the kringle-2 and the catalytic domains of Reteplase. The observed stabilizing interactions between excipient molecules and

APRs located in loop regions, in particular within the activation peptide segment, might have a crucial impact in this sense.

The conformational space of Reteplase was explored by aMD simulation, which further demonstrated the high stability of secondary structure elements, while conformational changes are mainly restrained to the tertiary structure level. When allowing for potential energy barriers to be crossed at high-boost mode, partial unfolding is observed with respect to the β-barrels in the catalytic domain and the α-helix or β-sheet in the kringle-2 domain. However, it is arguable that the addition of arginine to the system reduces the efffect of protein structure destabilization.

The initial results from CG simulation suggest that arginine may improve the colloidal stability of Reteplase by reducing protein-protein interactions. The strengths of the unique interactions are normalized, and thus a relative comparison is amenable. Further work within this subject should be initiated for statistical purpose, and might unravel other important aspects involved in the protein aggregation process of Reteplase.

The present study in this project demonstrates some of the challenges in formulation design for freeze-drying of proteins. The addition of excipients as stabilizers has an experimetally and theoretically proven effect on preservation of the native fold and protection against protein aggregation. In this regard, MD simulation serves as an important tool to investigate the stabilizing protein-excipient interactions at the molecular level with a potential to optimize the efficency of biopharmaceutical formulation screening. Further work might be aimed at the identification of a suitable molecular indicator, as to assess the performance of the excipient.

BIBLIOGRAPHY

1. Liu B, Zhou X. Freeze-Drying of Proteins. In: Wolkers WF, Oldenhof H, editors. *Cryopreservation and Freeze-Drying Protocols*. 4th ed. New York: Springer Science + Business Media, LLC; 2021. p. 683–702.
2. Wang W. Lyophilization and development of solid protein pharmaceuticals. *Int J Pharm*. 2000;203:1–60.
3. Arsiccio A, Pisano R. Clarifying the role of cryo- and lyo-protectants in the biopreservation of proteins. *Phys Chem Chem Phys*. 2018;20(12):8267–8277.
4. Arsiccio A, Paladini A, Pattarino F, Pisano R. Designing the Optimal Formulation for Biopharmaceuticals: A New Approach Combining Molecular Dynamics and Experiments. *J Pharm Sci*. 2019;108(1):431–438.
5. Cloutier T, Sudrik C, Mody N, Sathish HA, Trout BL. Molecular Computations of Preferential Interaction Coefficients of IgG1 Monoclonal Antibodies with Sorbitol, Sucrose, and Trehalose and the Impact of These Excipients on Aggregation and Viscosity. *Mol Pharm*. 2019;16(8):3657–3664.
6. Kasper JC, Winter G, Friess W. Recent advances and further challenges in lyophilization. *Eur J Pharm Biopharm*. 2013;85(2):162–169.
7. Kawasaki H, Shimanouchi T, Kimura Y. Recent Development of Optimization of Lyophilization Process. *J Chem*. 2019;1–14.
8. Monte SPD and PRD. Effect of glass transition temperature on the stability of lyophilized formulations. *Pharm Res*. 1997;14(5):591–595.
9. Morais ARDV, Alencar ÉDN, Xavier Júnior FH, Oliveira CM De, Marcelino HR, Barratt G, et al. Freeze-drying of emulsified systems: A review. *Int J Pharm*. 2016;503:102–114.
10. Chang LL, Pikal MJ. Mechanisms of protein stabilization in the solid state. *J Pharm Sci.* 2009;98(9):2886–2908.
11. Arsiccio A, McCarty J, Pisano R, Shea JE. Heightened Cold-Denaturation of Proteins at the Ice-Water Interface. *J Am Chem Soc*. 2020;142(12):5722–5730.

12. Tsai CJ, Maizel JV, Nussinov R. The hydrophobic effect: A new insight from cold denaturation and a two-state water structure. *Crit Rev Biochem Mol Biol*. 2002;37(2):55–69.

13. Bhatnagar BS, Bogner RH, Pikal MJ. Protein stability during freezing: Separation of stresses and mechanisms of protein stabilization. *Pharm Dev Technol.* 2007;12(5):505–523.

14. Stärtzel P. Arginine as an Excipient for Protein Freeze-Drying: A Mini Review. *J Pharm Sci.* 2018;107(4):960–967.

15. Ohtake S, Kita Y, Arakawa T. Interactions of formulation excipients with proteins in solution and in the dried state. *Adv Drug Deliv Rev*. 2011;63(13):1053–1073.

16. Bayat M, Gourabi H, Khammari A, Ahmad F, Saboury AA. A comparative study of structure, stability and function of sc-tenecteplase in the presence of stabilizing osmolytes. *J Biotechnol.* 2018;280:1–10.

17. Centocor Inc. *Retavase Product monograph*. Vidyya Medical News Service. USA; 1999.

18. European Medicines Agency. *Summary of Product Characteristics* - *Rapilysin* [Internet]. 2006. Available from: https://www.ema.europa.eu/en/documents/product-information/rapilysin-epar-product-information_en.pdf [Accessed 8th March 2021]

19. Schneider CP, Trout BL. Investigation of cosolute-protein preferential interaction coefficients: New insight into the mechanism by which arginine inhibits aggregation. *J Phys Chem B.* 2009;113(7):2050–2058.

20. Shukla D, Trout BL. Preferential interaction coefficients of proteins in aqueous arginine solutions and their molecular origins. *J Phys Chem B*. 2011;115(5):1243–1253.

21. Shukla D, Trout BL. Interaction of arginine with proteins and the mechanism by which it inhibits aggregation. *J Phys Chem B*. 2010;114(42):13426–13438.

22. Vazdar M, Vymětal J, Heyda J, Vondrášek J, Jungwirth P. Like-charge guanidinium pairing from molecular dynamics and ab initio calculations. *J Phys Chem A.* 2011;115(41):11193–11201.

23. Vondrášek J, Mason PE, Heyda J, Collins KD, Jungwirth P. The molecular origin of like-

charge arginine - Arginine pairing in water. *J Phys Chem B*. 2009;113(27):9041–9045.

24. Mousavi SB, Fazeli A, Shojaosadati SA, Fazeli MR, Hashemi-Najafabadi S. Purification and efficient refolding process for recombinant tissue-type plasminogen activator derivative (reteplase) using glycerol and Tranexamic acid. *Process Biochem*. 2017;53:135–144.

25. Ghaheh HS, Ganjalikhany MR, Yaghmaei P, Pourfarzam M, Mir Mohammad Sadeghi H. Improving the solubility, activity, and stability of reteplase using in silico design of new variants. *Res Pharm Sci.* 2019;14(4):359–368.

26. Lamba D, Bauer M, Huber R, Fischer S, Rudolph R, Kohnert U, et al. The 2.3 Å Crystal Structure of the Catalytic Domain of Recombinant Two-chain Human Tissue-type Plasminogen Activator. *J Mol Biol.* 1996;258(1):117–135.

27. Mohammadi E, Mahnam K, Jahanian A. Reteplase : Structure , Function , and Production. *Adv Biomed Res.* 2019;8(19):1–6.

28. Cai Y, Bao J, Lao X, Zheng H, Chen J, Yu R. Computational design, functional analysis and antigenic epitope estimation of a novel hybrid of 12 peptides of hirudin and reteplase. *J Mol Model.* 2015;21(229):1–9.

29. Mican J, Toul M, Bednar D, Damborsky J. Structural Biology and Protein Engineering of Thrombolytics. *Comput Struct Biotechnol J.* 2019;17:917–938.

30. Renatus M, Stubbs MT, Huber R, Bringmann P, Donner P, Schleuning WD, et al. Catalytic domain structure of vampire bat plasminogen activator: A molecular paradigm for proteolysis without activation cleavage. *Biochemistry.* 1997;36(44):13483–13493.

31. Layton JE, Nicholson S, Basser R, Cebon J. Cytokines: Granulocyte Colony-Stimulating Factor. In: Leroith D, Bondy C, editors. *Growth Factors and Cytokines in Health and Disease*. Amsterdam: Elsevier Science; 1997. p. 557–612.

32. Wood VE, Groves K, Cryar A, Quaglia M, Matejtschuk P, Dalby PA. HDX and in Silico Docking Reveal that Excipients Stabilize G-CSF via a Combination of Preferential Exclusion and Specific Hotspot Interactions. *Mol Pharm*. 2020;17(12):4637–4651.

33. Hill CP, Osslund TD, Eisenberg D. The structure of granulocyte-colony-stimulating factor

and its relationship to other growth factors. *Proc Natl Acad Sci U S A.* 1993;90(11):5167–5171.

34. Krishnan S, Chi EY, Webb JN, Chang BS, Shan D, Goldenberg M, et al. Aggregation of granulocyte colony stimulating factor under physiological conditions: Characterization and thermodynamic inhibition. *Biochemistry*. 2002;41(20):6422–6431.

35. Lindahl E. Molecular Dynamics Simulations. In: Kukol A, editor. *Molecular Modeling of Proteins*. 2nd ed. New York: Springer Science + Business Media; 2015. p. 3–26.

36. Hug S. Classical Molecular Dynamics in a Nutshell. In: Monticelli L, Salonen E, editors. *Biomolecular Simulation: Methods and Protocols.* New York: Springer Science + Business Media; 2013. p. 127–152.

37. Lewars EG. Molecular Mechanics. In: Roberts E, editor. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics.* 3rd ed. New York: Springer Science + Business Media; 2016. p. 51–99.

38. Deeth RJ. Molecular Mechanics. In: McCleverty JA, Meyer TJ, editors. *Comprehensive Coordination Chemistry II.* London: Elsevier Ltd; 2003. p. 457–465.

39. Schneider R, Sharma AR, Rai A. Introduction to Molecular Dynamics. In: Fehske H, Schneider R, Weiße A, editors. *Computational Many-Particle Physics.* Berlin: Springer Science + Business Media; 2008. p. 3–37.

40. Hünenberger PH. Thermostat algorithms for molecular dynamics simulations. *Adv Polym Sci.* 2005;173:105–147.

41. Cerutti DS, Duke R, Freddolino PL, Fan H, Lybrand TP. Vulnerability in Popular Molecular Dynamics Packages Concerning Langevin and Andersen Dynamics. *J Chem Theory Comput.* 2008;4(10):1669–1680.

42. Farago O, Grønbech-Jensen N. The fluctuation–dissipation theorem. *J Stat Phys.* 2014;156(6):1093–1110.

43. Åqvist J, Wennerström P, Nervall M, Bjelic S, Brandsdal BO. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem Phys Lett.* 2004;384:288–294.

44. Chow KH, Ferguson DM. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Comput Phys Commun.* 1995;91:283–289.

45. Arsiccio A, McCarty J, Pisano R, Shea JE. Effect of Surfactants on Surface-Induced Denaturation of Proteins: Evidence of an Orientation-Dependent Mechanism. *J Phys Chem B.* 2018;122(49):11390–11399.

46. Arsiccio A, Pisano R. Stability of Proteins in Carbohydrates and Other Additives during Freezing: The Human Growth Hormone as a Case Study. *J Phys Chem B.* 2017;121(37):8652–8660.

47. Arsiccio A, Pisano R. The Preservation of Lyophilized Human Growth Hormone Activity: how Do Buffers and Sugars Interact? *Pharm Res.* 2018;35(131):1–13.

48. Abascal JLF, Sanz E, Fernández RG, Vega C. A potential model for the study of ices and amorphous water: TIP4P/Ice. *J Chem Phys.* 2005;122(23):1–10.

49. Vega C, Abascal JLF. Simulating water with rigid non-polarizable models: A general perspective. *Phys Chem Chem Phys.* 2011;13(44):19663–19688.

50. Abascal JLF, Vega C. A general purpose model for the condensed phases of water: TIP4P/2005. *J Chem Phys.* 2005;123:1–13.

51. Matsumoto M, Saito S, Ohmine I. Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature.* 2002;416(6879):409–413.

52. Matsumoto M, Yagasaki T, Tanaka H. GenIce: Hydrogen-Disordered Ice Generator. *J Comput Chem.* 2018;39(1):61–64.

53. Steinberg MZ, Breuker K, Elber R, Gerber RB. The dynamics of water evaporation from partially solvated cytochrome c in the gas phase. *Phys Chem Chem Phys.* 2007;9(33):4690–4697.

54. Urbassek HM, Thirumuruganandham SP. Water evaporation and conformational changes from partially solvated ubiquitin. *Biochem Res Int.* 2010;1–6.

55. Gedeon PC, Thomas JR, Madura JD. Accelerated Molecular Dynamics and Protein Conformational Change: A Theoretical and Practical Guide Using a Membrane Embedded Model Neurotransmitter Transporter. In: Kukol A, editor. *Molecular Modeling of*

Proteins. 2nd ed. New York: Springer Science + Business Media; 2015. p. 253–288.

56. Hamelberg D, Mongan J, Mccammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys.* 2004;120:11919–11929.

57. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, et al. *Amber 2020 Reference Manual.* Vol. 53. 2019.

58. Pierce LCT, Salomon-Ferrer R, Augusto F. De Oliveira C, McCammon JA, Walker RC. Routine access to millisecond time scale events with accelerated molecular dynamics. *J Chem Theory Comput.* 2012;8(9):2997–3002.

59. Ingólfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, et al. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip Rev Comput Mol Sci.* 2014;4(3):225–248.

60. Markvoort AJ. Coarse-Grained Molecular Dynamics. In: van Santen RA, Saute P, editors. *Computational Methods in Catalysis and Materials Science*. Weinheim: Wiley - VCH; 2009. p. 151–166.

61. Darré L, Machado MR, Brandner AF, González HC, Ferreira S, Pantano S. SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J Chem Theory Comput.* 2015;11(2):723–739.

62. Darré L, MacHado MR, Dans PD, Herrera FE, Pantano S. Another coarse grain model for aqueous solvation: WAT FOUR? *J Chem Theory Comput.* 2010;6(12):3793–3807.

63. Machado MR, Barrera EE, Klein F, Sónora M, Silva S, Pantano S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J Chem Theory Comput.* 2019;15(4):2719–2733.

64. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):27-28,33-38.

65. PyMOL Molecular Graphics System. Version 2.0. New York: Schrödinger LLC; 2015.

66. MATLAB. Version 9.6.0 (R2019a). Natick, Massachusetts: The Math Works Inc; 2019.

67. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006;34:668–672.

68. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–3402.

69. RAMPAGE: Ramachandran Plot Analysis [Internet]. Available from: http://mordred.bioc.cam.ac.uk/~rapper/rampage.php [Accessed 7th February 2020].

70. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, III, et al. *AmberTools19 and Amber18.* AMBER 2019. San Francisco: University of California; 2019.

71. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general Amber force field. *J Comput Chem.* 2004;25(9):1157–1174.

72. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J Comput Chem.* 2008;29:1859–1865.

73. Brooks BR, III CLB, MacKerell AD, Jr. LN, Petrella RJ, Roux B, et al. CHARMM: The Biomolecular Simulation Program. *J Comput Chem.* 2009;30:1545–1614.

74. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput.* 2016;12(1):405–413.

75. Lee J, Hitzenberger M, Rieger M, Kern NR, Zacharias M, Im W. CHARMM-GUI supports the Amber force fields. *J Chem Phys.* 2020;153:1–10.

76. Kim S, Lee J, Jo S, Brooks CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem.* 2017;38(21):1879–1886.

77. Chéron N, Naepels M, Pluhařová E, Laage D. On protein preferential solvation in water: Glycerol mixtures. *ChemRxiv.* 2019;33:1–38.

78. Stark AC, Andrews CT, Elcock AH. Toward optimized potential functions for protein-protein interactions in aqueous solutions: Osmotic second virial coefficient calculations using the MARTINI coarse-grained force field. *J Chem Theory Comput.* 2013;9(9):4176–4185.

79. Schmalhorst PS, Deluweit F, Scherrers R, Heisenberg CP, Sikora M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J Chem Theory Comput.* 2017;13(10):5039–5053.

80. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys.* 2004;120(20):9665–9678.

81. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;19–25.

82. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS BT - Solving Software Challenges for Exascale. In: Markidis S, Laure E, editors. *EASC 2014 Conference Proceeding*. Cham: Springer International Publishing; 2015. p. 3–27.

83. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013;29(7):845–854.

84. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput.* 2008;4(3):435–447.

85. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26(16):1701–1718.

86. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model.* 2001;7(8):306–317.

87. Berendsen HJC, Spoel D Van Der, Drunen R Van. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun.* 1995;95:43–56,.

88. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys.* 1977;23(3):327–341.

89. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci.* 2001;98(18):10037–10041.

90. Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* 2015;43:306–313.

91. Wedberg R, Abildskov J, Peters GH. Protein dynamics in organic media at varying water activity studied by molecular dynamics simulation. *J Phys Chem B.* 2012;116(8):2575–2585.

92. Baynes BM, Trout BL. Proteins in Mixed Solvents: A Molecular-level Perspective. *J Phys Chem B.* 2003;107(50):14058–14067.