SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

In silico studies in the field of hemostasis and related processes: the reaction mechanism of Factor XIII A subunit and the pentasaccharide binding of antithrombin

by Gábor Balogh

Supervisor: Zsuzsanna Bereczky, MD, PhD; István Komáromi†, PhD

UNIVERSITY OF DEBRECEN

LAKI KÁLMÁN DOCTORAL SCHOOL

Debrecen, 2020

In silico studies in the field of hemostasis and related processes: the reaction mechanism of Factor XIII A subunit and the pentasaccharide binding of antithrombin

By Gábor Balogh MSc

Supervisor: Zsuzsanna Bereczky MD, PhD; István Komáromi† PhD

Kálmán Laki Doctoral School, University of Debrecen

Head of the **Examination Committee**:	György Balla, MD, PhD, DSc, Member of Hungarian Academy of Sciences
Members of the Examination Committee:	János Mótyán, PhD
	Imre Bodó, MD, PhD

The Examination takes place at 11:00 a.m., on August 13, 2020, in online format.

Head of the **Defense Committee**:	György Balla, MD, PhD, DSc, Member of Hungarian Academy of Sciences
Reviewers:	György Ferenczy, PhD, DSc
	Mihály Purgel. PhD
Members of the Defense Committee:	János Mótyán, PhD
	Imre Bodó, MD, PhD

The PhD Defense will be held at 1:00 p.m., on August 13, 2020.

Live online access will be provided. If you wish to participate in the discussion, please send an email to zsberczky@med.unideb.hu, not later than 12:00 the day before (August 12, 2020).

1. Introduction

Hemostasis has two major biological functions: it is responsible for maintaining blood flow, on the other hand, it prevents further blood loss in the case of a vessel injury. It consists of three steps: vasoconstriction at the site of the injury, activation, and aggregation of platelets, and the activation of blood coagulation cascade resulting in the formation of cross-linked fibrin. Several blood coagulation factors (thrombin, VIIa, IXa, Xa, XIa, XIIa) exhibit serine protease activity. In the blood coagulation cascade, the factors bearing enzyme activity activate the next member of the cascade, and so on.

Activation of factor XIII is one of the last events of the coagulation process. Its active A subunit (FXIII-A) has transglutaminase enzyme activity and forms isopeptide cross-links between two substrates when activated. Furthermore, Stieler et al. have proposed that this coagulation factor may constitute a promising target for anticoagulant drug development.

Understanding the reaction mechanism is highly important for drug development. However, in contrast with the related enzyme transglutaminase 2, relatively little amount of information is available on FXIII-A catalytic mechanism. In addition to reaction kinetics measurements, *in silico* methods have been developed for investigating catalytic mechanisms in enzymes. Hybrid quantum mechanics/molecular mechanics (QM/MM) methods represent the most widespread category of techniques for the *in silico* study of enzyme reactions. In such a protocol, computationally expensive quantum chemistry methods are applied for the catalytically most important amino acid residues and atom groups, allowing the exact description of the reaction. The rest of the system is described by more computationally efficient methods, typically molecular mechanics force fields.

The regulators of the blood coagulation cascade play an essential role in hemostasis as they limit the process to the site of injury. The most important regulators of the coagulation cascade are antithrombin, activated protein C and tissue factor pathway inhibitor (TFPI), heparin cofactor II and protein Z dependent protease inhibitor are also involved in this process. The essential anticoagulant role of antithrombin (AT) is underlined by the fact that both the inherited and the acquired deficiency of this protein increase the risk for thrombotic diseases. Additionally, AT is also the main pharmacological target for one of the most important groups of anticoagulant drugs: heparins and heparin derivatives.

Elucidation of the details of AT conformational activation could provide invaluable information for the design of new AT-dependent anticoagulant drugs. Furthermore, several pathogenic AT variants have been described that cause impaired heparin-binding and activation. The study of the activation mechanism could help to understand the consequences of these mutants. In the present work, we have investigated the high-affinity pentasaccharide derivative using *in silico* techniques, especially "advanced sampling" molecular dynamics simulations. Such simulations could provide insights into the mechanisms that cannot be obtained using X-ray diffraction structures, showing a "static" image of the protein. Finally, AT can serve as a model system for studying the phenomenon of allosteric activation in general.

2. Review of literature

2.1 First step of FXIII-A catalytic mechanism

The A subunit of FXIII-A belongs to the family of human transglutaminase enzymes. The family comprises eight members exhibiting catalytic activity: transglutaminases 1-7 (TG 1-7) and Factor XIII. Based on sequence homology, the catalytically inactive erythrocyte band 4.2 protein is also recognized as a member of this family. The active transglutaminase enzymes form covalent, isopeptide cross-links between glutamine and lysine residues in their substrates. The catalytically active members contain a Cys- His-Asp catalytic triad, comparable to the cysteine protease enzymes. Iismaa et al. have proposed a two-step mechanism for transglutaminases that is analogous to a "reversed" cysteine protease reaction.

Furthermore, in the absence of an amide donor substrate (Lys residue in case of proteins and peptides) human transglutaminases can catalyze the hydrolytic reaction of the Gln amide group but this is a significantly slower reaction than the cross-linking. Factor XIII is found in an A_2B_2 heterotetramer form, in which two non-catalytic B subunits are bound to the two A subunits. The intracellular form of Factor XIII consists of two A subunits. The A subunit is comprised of an activation peptide and four domains: beta-sandwich, catalytic core, and two beta-barrels. Based on the results of X-ray diffraction studies, the inactive A subunit is a dimer. The catalytic center in this dimer is made inaccessible by both the Tyr541 residue of the first beta-barrel domain and the activation peptide of the other A subunit (especially the Arg12 sidechain)

In blood plasma, factor XIII is activated by the cleavage of the activation peptides by thrombin, binding of calcium ions and the dissociation of the B subunits. If the calcium concentration is significantly higher than physiological, Factor XIII can also be activated non-proteolytically, without the cleavage of the activation peptides. Factor XIII is activated among the later events of blood coagulation. Its most important role is the cross-linking of fibrinogen alpha chains, but additionally it exhibits several further intra-and extracellular functions.

In 2013, Stieler et al. have published a new X-ray structure of the A subunit, corresponding to a non-proteolytically activated, monomeric state, with an irreversible inhibitor (ZED-1301) bound. In the structure, large displacements of the beta-barrel domains can be observed compared to the native form. It contains three occupied calcium binding sites. However, the existence of further, "weaker" binding sites can be inferred based on ^{43}Ca NMR

measurements and molecular dynamics simulations. These sites are possibly involved in the non-proteolytic activation of the protein. The monomeric nature of activated FXIII-A, as suggested by the structure, has been confirmed recently by analytic ultracentrifugation and atomic force microscopy experiments.

In the human transglutaminase family, the largest amount of information regarding the catalytic mechanism is available for transglutaminase 2. However the amount of data available for FXIII is much more limited. Most papers in the field focus on the catalytic constants of the transglutaminase (or hydrolase) reaction for several substrates, but the mechanism has not been investigated in detail.

Besides enzyme kinetics measurements, hybrid quantum-chemistry/molecular mechanics calculations can reveal important details of enzyme reaction mechanisms. To the best of our knowledge, one of our papers is the first where the catalytic mechanism itself was studied in a human transglutaminase enzyme. In the only previous QM/MM study related to a human transglutaminase, the binding of irreversible inhibitors to transglutaminase 2 was investigated, not the conversion of a model substrate. In this work, QM/MM-based molecular dynamics simulations were performed, using semi-empirical methods. The studies available for the related enzyme family cysteine proteases are much more numerous. Both the protonation states of the catalytic amino acid residues and the acylation/deacylation mechanism have been investigated for several members of this family, including papain, cathepsin B and K, as well as caspase 7. Both energy minimization-based calculations with high level QM methods and semiempirical-based QM/MM molecular dynamics simulations were applied to study cysteine proteases. In a previous work from our group, the catalytic mechanism of papain was investigated using hybrid QM/MM calculations using an ONIOM scheme.

2.2 Pentasaccharide binding to antithrombin

AT is a 58 kDa glycoprotein belonging to the serpin superfamily of proteins. The members of this superfamily can be found in animals, plants, bacteria and viruses. As the name suggests, many serpins have a serine protease inhibitory function, but the family also includes inhibitors and cysteine proteases, as well as proteins with various functions (such as hormone transporters). The human genome contains 36 serpin coding genes. Although the sequence homology between serpins is often limited, they share a common tertiary structure, consisting

of 8 or 9 α-helices and three β-sheets. The "native" forms of inhibitory serpins contain a solvent-exposed, flexible loop where the reaction center of the serpin-protease reaction is located. The partial cleavage of the P1-P1' peptide bond, found in this loop, has a crucial importance in the inhibitory mechanism. The proteolytic reaction, however, ends at the thioester or ester intermediate and the intermediate is stabilized through a set of complex conformational changes. The most important of these changes is the insertion of the N-terminal part of the RCL into β-sheet A. The tertiary structure of the protease becomes distorted, this prevents the remaining step of the proteolytic reaction. The pre-requisite of this mechanism is the metastability of the "native" serpins, the insertion of the full RCL or a part of it is accompanied by a large decrease in free enthalpy. RCL insertion takes place during serpin-protease complex formation, the cleavage of the RCL (without protease inhibition) or the transition of the molecule into the inactive, “latent” state.

AT is one of the main regulators of the coagulation cascade, through inhibition of coagulation factors with serine protease activity. Its most important targets are thrombin, factor Xa and IXa, but can also inhibit factor XIa. The circulating native form of AT is an inhibitor of low efficiency, but the reaction rates of the inhibition is increased significantly in the presence of heparin and heparin analogs. Two different mechanisms have been proposed for the increase of inhibitory activity in the literature. First, the binding of a high-affinity, specific pentasaccharide triggers complex conformational changes in the AT molecule. This conformational activation is critical in the case of factors Xa and IXa, but its significance is lower in the case of thrombin. Additionally, longer heparin chains can form bridges between AT and the coagulation factor. This second mechanism is especially important in the case of thrombin. AT has two isoforms in the blood plasma: alpha-AT (90%) is glycosylated on four Asn residues, while the beta (10%) form lacks glycan on Asn135. The beta form has significantly higher heparin affinity than the alpha variant and it is probably a more efficient inhibitor.

AT plays an essential role in the regulation of the coagulation cascade, this is supported by the fact that inherited or acquired AT deficiency predisposes patients to thrombotic diseases. Inherited AT deficiency can be classified based on whether they cause a quantitative (type I) or qualitative (Type II) defect. Type II variants are further classified into three subgroups: mutations affecting the heparin-binding site (II. HBS), the reactive site (II. RS) or exhibiting a pleiotropic effect (II. PE)

To better understand the activation mechanism of AT, several X-ray diffraction studies have been conducted. Beyond the structures of the non-activated and pentasaccharide activated conformations, a structure corresponding to an “intermediate” activated state is also available. In this state, most of the conformational changes accompanying the pentasaccharide binding have already been taken place, except for the expulsion of the "hinge" region (the N-terminal end of RCL) from β-sheet A and the C-terminal extension of helix D. Many “earlier” X-ray structures of AT correspond to a dimer of a native and a latent molecule, where the RCL of native AT forms contacts with the latent form, affecting its conformation significantly. Later, Johnson et. al published a new structure of an AT variant in which the transition into the latent state was prevented using an engineered disulfide bond. This study has revealed a novel conformation of the RCL.

Based on kinetics measurements and X-ray diffraction structures, a three-step mechanism was proposed for the conformational changes caused by pentasaccharide binding. In the first step, the binding of the pentasaccharide is still weak, however, the AT-pentasaccharide interaction induces conformational changes, resulting in a stronger binding. The processes in the second step mainly affect the heparin binding site, as well as the hydrophobic core of the protein. AT reaches its maximum activity in the third step with the elongation of helix D and the expulsion oh “hinge” region.

3. Principles of methods used in the studies

3.1 Quantum chemistry methods

Born-Oppenheimer approximation is used in nearly all practical quantum chemistry calculations. In this approximation, the movements of atoms and electrons are treated separately. The basis of this approximation is the three to four orders of magnitude difference between the masses of electrons and atomic nuclei. It allows calculating the electronic energies while the atomic nuclei are treated as fixed. Hartre-Fock theory is a widely used methodology for the approximate determination of wave functions and electronic energies. The many-electron wave function has the form of a Slater determinant. It is constructed from N one-electron functions where N is the number of electrons in the molecule. (The Slater determinant form ensures the anti-symmetry for the exchange of two electrons.) The Hartree-Fock equations are solved iteratively. As this method uses an "average" electron repulsion potential, it cannot take into account the correlation in the movements between the electrons (electron correlation).

Møller-Plesset perturbation theory is a popular technique for the approximate treatment of electron correlation. The excited states are calculated using perturbatively, based on Rayleigh–Schrödinger perturbation theory. The starting point of this calculation is the Hartre-Fock wave function. A series expansion is applied; depending on the number of terms in the expansion considered, the methods are called MP2, MP3, MP4, and so on. From these methods, MP2 is the most frequently used one. It can take into account most of the electron correlation, however the results are correct only if the ground-state electron configuration is dominant.

3.2 Density functional theory

The theoretical foundations of density functional theory (DFT) are the Hohenberg-Kohn theorems. Based on the theory, Kohn and Sham developed a methodology that can be practically used for approximate calculations. A system consisting of non-interacting electrons is considered. In this system the same electrostatic potential is applied to all electrons and it has the same electron density as the "real" system. The functional has three terms: the first one is the functional of kinetic energy for the hypothetical system with non-interacting electrons, while the second term is the Coulomb energy for electron-electron interactions. The

exact mathematical form of the third term, the exchange-correlation functional is not known, however, several approximate formulas have been developed.

An example of a hybrid functional based technique is the popular B3LYP method. It combines the Becke functional (with three parameters) with the generalized gradient approximation (GGA)-based correlation functional proposed by Lee, Young and Parr. The main disadvantage of this method is the inaccurate treatment of dispersion interactions, "newer" functionals are typically corrected for these interactions. For example the ωB97X-D functional contains empirical atom-atom dispersion correction terms, while the M06 family of methods uses extensive parameterization for the treatment of these interactions.

3.3 Semi-empirical methods

Semi-empirical methods are "approximate" quantum chemistry methods that are much less expensive computationally than *ab initio* techniques, however, they are still significantly slower than "classical" molecular mechanics force fields. Unlike "classical" force fields, semi-empirical methods can be used for modeling chemical reactions and bond breakage. One group of semi-empirical methods is based on Hartree-Fock theory, but contains several approximations, especially for the two-electron integrals. The most commonly used family of such techniques is the NDDO (neglect of diatomic differential overlap) methods. Examples of NDDO methods are AM1, PM3 and PM6.

The DFTB method uses a series expansion of the DFT energy for the systems at a reference electron density. The zero- and first-order terms are computed using a minimal basis set and only the valence shell electrons are considered. The additional second-order terms in the SCC-DFTB (or DFTB2) methods describe the charge redistribution between the atoms. DFTB3 contains further terms for the approximate treatment of the third order terms in the series expansion. All these methods require parameter development, based on DFT calculations.

3.4 Hybrid QM/MM methods

Thee accurate computational treatment of chemical reactions requires high-level quantum chemistry methods. However, the modeling of large systems entirely with a QM method is not feasible in many cases even today. In a hybrid technique, an appropriate QM method (DFT, semi-empirical methods) is applied to the most important atoms and groups involved in the

chemical reaction. The rest of the system is treated using a computationally much less expensive technique, typically a molecular mechanics force field.

The first group of such techniques is called a “classical” (or additive) QM/MM scheme. In such a scheme, several methods exist for the treatment of interaction between the QM and the MM subsystems. The two most important ones are mechanical embedding and electrostatic embedding. In mechanical embedding these interactions are treated using Coulomb interaction between point charges, while in the electrostatic embedding, the partial charges of MM subsystem atoms are also present in the Hamiltonian of the QM systems. Link atoms are the most common way to treat covalent bonds between the two subsystems.

The second group of QM/MM techniques is called a subtractive method. The most important subtractive method is ONIOM which is based on the earlier IMOMM method. The ONIOM method is frequently used as a QM/MM scheme, but it allows more than two subsystems as well as a wide range of computational methods for the “low level” subsystem (or subsystems) (MM force fields, semi-empirical or less expensive *ab initio* methods). In an ONIOM-based QM/MM scheme, the total energy of the systems are calculated in three steps: (1) energy calculation for the full system using the MM method, (2) QM calculation for the “high level” “model” subsystem, and (3) MM energy computation for the same subsystem. One of the main advantages of the ONIOM scheme is that it does not require separate energy terms for the interactions between the QM and the MM systems; it is treated at the MM level. For the treatment of electrostatic interactions, mechanical or electrostatic embedding is used.

3.5 Molecular mechanics force fields

In molecular mechanics calculations, empirical potentials are used to study the behavior of molecular systems. The atoms are treated as rigid bodies and the movements of electrons are not taken into account. Thus the energy of the system will only depend on the position of the atoms (atomic nuclei). The bonds between the atoms are pre-defined before the simulations, therefore most molecular mechanics methods cannot be used to study the mechanism of chemical reactions or the breakage or formation of covalent bonds (reactive force fields are a notable exception).

There are three types of bonded interactions: bond stretching, bending and rotation around bonds. The energy terms for bond stretching and bending are proportional to the square of the deviation from an equilibrium value (similar to Hooke's law). The rotation around bonds is

typically described by a formula corresponding to three terms of a Fourier series. The energy depends on four atoms, forming a dihedral angle.

Among the bonded terms, the electrostatic interactions are described by Coulomb forces between point charges of the atoms. The Lennard-Jones potential is used for the *van der Waals* interactions. The Lennard-Jones potential contains two terms, the r^{-6} term corresponds to the attractive, while the r^{-12} term describes the repulsion between the atoms at short distances. “Classical” molecular mechanics force fields cannot take into account the polarizability of the atoms, unlike the more computational expensive polarizable force fields.

3.6 Molecular dynamics simulations

In molecular dynamics (MD) simulations, the behavior of the model system is described as the function of time. In "classical" molecular dynamics, the Newtonian laws of motion are used to describe the motions of the atoms in the model system. The velocities and the movement of atoms are calculated using numerical integration; the integration time-step falls in the range of 0.1-5 fs. Restraining the covalent bonds using e.g. the LINCS or SHAKE algorithms allows using a larger time-step. In the case of larger systems, the energies and forces are usually calculated using a classical molecular mechanics force field. Molecular dynamics simulations using quantum chemistry or QM/MM based methods are also possible.

The input data required for the simulations are the initial positions and velocities of the atoms. Typically, the molecular dynamics software generates the starting velocities randomly after a Maxwell-Boltzmann distribution. In every step, the forces acting on the atoms are calculated first using the applied force field (or QM method). The forces are calculated as the gradient of the potential energy. Using the forces, the velocities and the displacements of the atoms in the next state are computed using the Newtonian equations of motion. The MD software saves the positions and the velocities of the atoms, as well as additional data (temperature, pressure, energies for specific types of interactions) at a given frequency. The trajectory of the simulation consists of the saved structures from the simulation. Molecular dynamics simulations are performed most commonly under constant temperature and volume (NVT) or constant temperature and pressure (NpT). Multiple algorithms are available for temperature and pressure coupling. The most common methods for temperature coupling are the Berendsen, the v-rescale and the Nosé-Hoover thermostats.

3.7 Enhanced sampling molecular dynamics methods

The computational requirements of molecular dynamics simulations (which are often a few microseconds or tens of microseconds long) limit the accessible time scale. These limitations make the simulations of the so-called rare events challenging. The probability of such events is low due to the large energy barriers. (Biologically relevant conformational changes in proteins are often rare events.) For the more efficient simulation of such events, advanced sampling methods have been developed. One class of these methods depends on pre-defined collective variables or reaction coordinates, and the conformational space is explored as the function of these variables. Examples of such methods are umbrella sampling, adaptively biased molecular dynamics and metadynamics. The most important disadvantage of such techniques is that the choice of reaction coordinates often requires extra knowledge about the system and it is often non-trivial.

In a metadynamics simulation, the potential energy function of the system is modified at a given frequency by adding Gaussian functions. The maximum of the added function depends on the current values of the collective variables. Thus the potential energy will depend on the previous states of the system. The free energy or free enthalpy differences between the relevant states can be computed using the added Gaussian function. A popular variant of the technique is well-tempered metadynamics. It allows a more efficient convergence of the simulation. Also, algorithms have been developed in which the widths of the functions are also chosen automatically by the software.

The Gaussian Accelerated Molecular Dynamics (GAMD) technique allows advanced sampling without pre-defined reaction coordinates. The main difference between GAMD and the earlier Accelerated Molecular Dynamics method is that in the former technique the potential follows a Gaussian distribution, allowing exact free energy calculation using reweighting. The method was successfully used for studying the ligand binding mechanisms in various systems, including receptors and the HIV protease.

4. Aims of the study

Several studies were published in which the catalytic mechanism of cysteine protease enzymes was investigated using hybrid QM/MM techniques. However, to the best of our knowledge, no such model is available for any of the members of the human transglutaminases. Enzyme kinetic studies are available for transglutaminase 2, but not for factor XIII. We aimed to study the first acylation step from the reaction mechanism of activated FXIII-A, using two different QM/MM based protocols. Furthermore, we also wanted to investigate the protonation states of the most important catalytic residues, both in the Michaelis complex and the resting state of the enzyme.

The pentasaccharide binding of AT as well as its allosteric activation was the subject of several studies. However, very little information is available for the relatively early steps of activation, when the binding of the high-affinity pentasaccharide is still relatively weak. We aimed to study the pentasaccharide binding using “advanced sampling” molecular mechanics method GAMD to an AT conformation not containing any pentasaccharide or other activator. Such simulations allow us to study the conformational changes of AT from the results of the ligand binding MD simulation, as well as to investigate the allosteric processes using the trajectories.

5. Methods

5.1 First step of FXIII-A catalytic mechanism

5.1.1 Structure of FXIII-A – antiplasmin peptide complex

The conformations of the α2-antiplasmin dodecapeptide (N-terminal cleaved form) for the docking were obtained from a 10 ns molecular dynamics simulation. The starting structure of the peptide was built using UCSF Chimera, with all φ dihedrals set to -139° and all φ dihedrals to 135°. The simulation was performed under NpT conditions, using the AMBER03 force field and the GROMACS software. The 4KTY X-ray structure was chosen as starting conformation for FXIII-A, representing an active monomer form, with a bound irreversible inhibitor (ZED-1301) (the inhibitor was removed from the structure for the simulations).

The HADDOCK web server, version 2.2, was used for the docking. Three conformations obtained from the peptide MD simulations (t = 8, 9 and 10 ns) were docked to FXIII-A. The following amino acids were defined as "active": 214-215, 223, 279-283, 289, 290, 313-315, 317, 339, 342, 350-352, 360, 365-374, 398-400, 441, 456, 459-460, as well as all amino acids of the dodecapeptide. Additional restraints were added between the protein Cys314 amino acid and the Gln2 side chain carbonyl carbon. From the four largest clusters returned by HADDOCK, the representative conformations were analyzed. The most suitable structure was selected for further refinement based on the presence of the Trp279-peptide Gln2 salt bridge and the position of the Cys314 residue compared to the peptide glutamine.

The catalytic domain from the selected docked structure was further refined using a "simulated annealing" simulation, using the GROMACS software and the AMBER 14SB force field. Our simulated annealing protocol consisted of 10 steps, each 10 ns long. In each step, the protein was heated from 5 to 400 K in 2 ns, equilibrated at 400 K for 3 ns and then the temperature was gradually decreased to 5 K. The simulation was run under NpT conditions, using the “v-rescale” thermostat and the Berendsen barostat. The Coulomb cutoff was set to 1 nm, long-range electrostatic interactions were treated using the PME method.

5.1.2 Model systems used in QM/MM simulations

We have investigated the first step of the FXIII reaction mechanism, common in the transglutaminase and glutaminase reactions, using two different QM/MM protocols. Firstly we have carried out ONIOM-type QM/MM calculations, in which the high level system was

treated using high-level MP2 or DFT methods. On the other hand, we performed QM/MM MD simulations, where free energies were calculated using the metadynamics method and the solvation was taken into account using explicit solvent.

A truncated model system, consisting of residues close to the catalytic site, was created for our QM/MM calculations. The QM subsystem consisted of the Cys314 and the His373 residues, involved in the catalysis and the substrate molecule corresponding to the Gln2 residue. In our model system, a propionamide molecule was used as a substitute for the Gln2 sidechain. The reason for the truncation was that the flexibility of the peptide in our early attempts led to convergence problems. The model system included further amino acids (211–216, 219–227, 266–268, 274–279, 283–292, 311–313, 315–322, 334–342, 368–372, 374–376, 395–400, 430–434, 464–466) in the MM subsystem. Energies for these residues were computed using the AMBER force field as implemented in Gaussian 09. To study the protonation states of Cys314 and His373 in the "resting" state, a second model system was built, that did not contain the substrate. Link atoms were defined along the Cα – Cβ bonds for the three amino acids in which the side chain belonged to the QM subsystem.

For comparability with the ONIOM-type simulations, the model system contained the same amino acids as the ONIOM system and the ligand was also identical. The solvation effects were taken into account using explicit TIP3P solvent containing Na^+ and Cl^- ions, corresponding to a 0.15 M salt concentration. As molecular dynamics simulations comprise large number of steps, a DFT-based semi-empirical method, DFTB3 was applied to the QM subsystem, allowing efficient energy evaluation in each step. The 3ob-3-1 parameter set, optimized for organic and biomolecules was used in the DFTB3 calculations. The link atoms were defined similarly as in the ONIOM-type calculations. They were positioned along the Cα – Cβ, bond, approximately 1 Å from the Cβ atom.

5.1.3 ONIOM-type calculations

The Gaussian 03 and 09 software was used for the ONIOM-type calculation. At the beginning, we have performed a geometry optimization on the prepared model system. During the optimization the 211-216, 219-227, 266-268, 274-279, 283-292, 311, 320-322, 334-342, 368-369, 371, 376, 395, 397, 399, 430-434, 464-466 were restrained, all other residues were allowed to move.

To investigate nucleophilic attack step, we have executed two-dimensional potential energy scans along/using two different pairs of reaction coordinates: the Cys314 sulfur – substrate carbonyl carbon distance and the amide carbon-nitrogen distance in the first case and the Cys314 sulfur – substrate carbonyl carbon distance and the His373 ε-proton-amide nitrogen distance for the second scan. Geometry optimizations were run in all grid points with a 0.2 Å density (0.1 Å near the hypothesized position of the transition state). Due to convergence problems, energies were not calculated in some points in the d_1: 2.8-3.2 Å, d_2: 2.0-2.6 Å range, but these are expected to be high energy conformations without biological relevance. Both potential energy surfaces were calculated using two different methods: MP2/6-31G(d) and M06-2X/6-31+G(d,p). We investigated the dissociation of the tetrahedral intermediate (I 1) using a further scan. In these scans, the same two QM methods were applied as previously. The stationary points, including the saddle points were re-optimized with the restraints on the reaction coordinates removed, using multiple methods and basis sets. In the case of the DFT calculations, we performed frequency calculations in order to compute zero-point vibration corrections. We have also checked the presence of imaginary frequencies in transition states.

The protonation states of the Cys314 and His373 amino acids in the Michaelis complex was investigated using a further potential energy scan, with a single reaction coordinate, the distance between the Cys314 sulfur atom and the proton. A similar calculation was also performed on the "resting state" of the enzyme, using a model system that did not contain the substrate.

5.1.4 QM/MM MD simulations

For the QM/MM MD simulations we used a “patched” version of the Gromacs molecular dynamics software, version 5.0. The DFTB3 code for GROMACS was developed by Kubař et al. The software was additionally patched with PLUMED for the metadynamics simulations. An energy minimization, with the steepest descent method was performed first on the model system, this was followed by a 1 ns "equilibration" MD simulation under NVT conditions on 310 K. The integration time step was 0.5 fs. Position restraints were applied to all amino acids except 312-319, 370, 372-375, 396, 398 and 400. During the equilibration phase, a further, weaker restraint was applied to the Trp279 sidechain and the substrate.

The Gromacs 5.0 software was used for the QM/MM MD simulations. The protonation of the Cys314 and the His373 sidechains was investigated using a 2 ns long well-tempered

metadynamics simulation, a bias factor of 20 was chosen. Similarly to the ONIOM simulation of the same reaction step, the reaction coordinate was the distance between the Cys314 sulfur atom and the proton. The free energy surface for the nucleophilic attack step was calculated using a well-tempered metadynamics simulation, with two collective variables: the Cys314 sulfur-carbonyl carbon distance and the amide carbon-nitrogen distance. (the same as in the ONIOM-type calculations). The widths of the Gaussians added during the metadynamics simulations were automatically determined by the software. To limit the phase space accessible to the system, upper and lower "walls" were applied (d_1: 0.17 – 0.37 nm, d_2: 0.13 – 0.33 nm).

5.2 Mechanism of the pentasaccharide binding to antithrombin

5.2.1 Model system preparation

To investigate the early events of pentasaccharide binding, we have created three model systems for molecular dynamics simulations. The first of these was based on the 1T1F X-ray structure of native AT, corresponding to a monomeric form (and not the dimer of a native and a latent serpin molecule.) The first system did not contain a ligand.

The second system was based on the same X-ray structure but a pentasaccharide, idraparinux was added. We have studied the structure of this molecule in solution using nuclear magnetic resonance spectroscopy (NMR) and molecular dynamics simulation. The initial structure of the molecule was obtained from the "NTP" ligand in the X-ray structure 1NQ9 (AT-pentasaccharide complex) that differs slightly from idraparinux. The starting position of the ligand was determined using a two-step procedure. The two structures were aligned first using the root mean square deviations (RMSD). The ligand was then shifted by 12 Å along the vector connecting the centers of masses of the two molecules so that there were no longer any significant contacts between them.

The third model system was derived from the 1NQ9 X-ray structure, representing an "intermediate" activated form of AT, complexed with a pentasaccharide. The ligand was the same as in the second system, idraparinux. The starting structures of missing loops in the X-ray structure were built using MODELLER.

The glycosylation of AT was modeled by disaccharides, consisting of two N-acetyl-D-glucosamine units, bound to three Asn residues of the protein: Asn96, Asn155, and Asn192.

The reason for the truncation of the oligosaccharides was that the exhaustive conformational sampling of the large and highly flexible glycans would be computationally too demanding. The three glycosylation sites corresponded to those found in beta-AT.

The pentasaccharide was modeled using the CHARMM carbohydrate force field. In a previous work from our group, we have found that this carbohydrate force field can best reproduce the available NMR data (hydrogen-hydrogen distances, Nuclear Overhauser Effects) for this molecule. For the protein, the CHARMM36m force filed was applied.

For the building of topology files and the system preparation, the CHARMM-GUI web server was used. The simulations were performed in a cubic box, under periodic boundary conditions. The system was solvated with TIP3P explicit solvent (the CHARMM modified version). The numbers of the added Na+ and Cl- ions were automatically determined by the software so that the ionic strength was approximately 0.15 M.

5.2.2 Molecular dynamics simulations

The AMBER16 pmemd.cuda program was used for all molecular dynamics simulations. The simulations were preceded by energy minimization in two steps. The first minimization consisted of 500 steps with steepest descents and 1500 steps with conjugate gradients method, with position restraints on the protein. A second minimization of 2000 steps was performed later using the conjugate gradients method, with the restraints removed.

Before the advanced sampling simulations, 150 ns equilibrium MD was performed for all model systems. To obtain suitable starting conformations, four independent simulations were run in case of the “non-activated AT and added ligand” systems (labeled A-D). Two of them were used later as starting geometries for further simulations (A, and B), based on the position of the pentasaccharide as measured by RMSD from the complex X-ray structure.

The GAMD simulations were preceded by a 60 ns preparation, equilibration run. The first part of the preparation phase corresponded to a 10 ns equilibrium MD simulation, no data collection occurred in the first 4 ns. In the remaining 50 ns, GAMD acceleration was applied to the system, and the parameters were updated at regular intervals in the last 40 ns part. A dual-boost GAMD scheme was used, acceleration was both used for the both the total potential energy and the dihedral energy. For all starting structures, four GAMD simulations were performed, a total of four simulations in the case of the “AT and added pentasaccharide”

system (A1, A2, B1 and B2). The GAMD simulations were performed under constant temperature (310 K) and volume, using a Langevin thermostat. The integration time step was 2 fs. (A Coulomb cutoff of 10 Å and a Lennard-Jones cutoff of 12 Å was applied, with a force "switch" between 10 and 12 Å). Long-range electrostatic interactions were treated using the PME method.

The production MD simulations were 600 ns long. Three simulations, in which pentasaccharide conformations comparable to the X-ray structures could be observed, were expanded to 1 μs (A1, B1 and one of the simulations based on the 1NQ9 structure).

5.2.3 Trajectory analysis

The CPPTRAJ program was used for trajectory analysis, including the calculation of atom-atom distances, RMSDs and ring conformations for the pentasaccharide subunits. The secondary structure of the protein was analyzed using the DSSP algorithm. The RMSD of the pentasaccharide, compared to the 1 NQ9 X-ray structure, was calculated using a two-step method: amino acids 6-26, 39-133, 137-355, 361-377, 402-431 protein was aligned first on the X-ray structure and then RMSDs were calculated between the ring and interglycosidic atoms of the ligand. Representative conformations from the trajectories were obtained using cluster analysis. The clustering was performed using the K-means algorithm as implemented in CPPTRAJ. In one case (the conformational change of the P helix), however, we have followed a different approach. The trajectory was split into two parts at the point of the change and the frames closest to the average structures were used.

Correlated motions between the amino acids of protein were analyzed with the “generalized correlation” algorithm developed by Lange and Grubmüller, as implemented in the “g_correlation” program. The free energy calculations (“reweighting”) from the GAMD simulations were performed using the PyReweighting program.

6. Results and discussion

6.1 First step of FXIII-A catalytic mechanism

6.1.1 Structure of FXIII – α2-antiplasmin peptide complex

The last frame (t = 100ns) from the simulated annealing simulation of the peptide complex was used as the starting structure for the QM/MM calculations. The structure obtained is in good agreement with the available NMR and enzyme kinetics data. The positively charged N-terminal of Asn1 forms an intermolecular salt bridge with Glu3 thus affecting the conformation of the peptide significantly. The importance of Asn1 is supported by kinetic measurements as reaction rates are much slower for related peptides not containing this amino acid. The same amino acid also forms a hydrogen bond with the Tyr372 side chain in the protein. The Pro7 residue is located in a small hydrophobic pocket, formed by FXIII-A amino acids Val369, Phe339, Leu439 and Tyr441. The Leu8 and Leu10 amino acids of the peptide are mainly involved in hydrophobic interactions with the regions of the protein homologous with the "hinge" regions of TG2. The interaction of these two residues is supported by experimental data. The Lys12, known to be involved in the binding, forms a salt bridge with the Asp456 residue of the protein. The minor differences between the refined structure and the NMR data can be explained by the fact that NMR provides information about an ensemble of structures, but we have only obtained a single conformation for further QM/MM study.

6.1.2 The protonation state of the Cy314 and His373 residues in ONIOM-type calculations

The role of Proton transfer is essential in the catalytic mechanism of transglutaminases, as well as cysteine and serine proteases. Both the protonation states of the catalytic His and Cys residues (the Cys314 must be deprotonated before the nucleophilic attack) and the proton transfer from the protonated His residue to the substrate amide group play an important role in the process. Considerable amount of data is available on the protonation of these residues in cysteine proteases, an enzyme family related to human transglutaminases. These studies support the existence of a Cys-His ion pair in these enzymes in a wide pH range. In contrast, Case and Stein has proposed a neutral Cys-His dyad in guinea liver pig transglutaminase and human transglutaminase 2. Regarding activated FXIII-A, no such study was published to the best of our knowledge.

We have investigated the protonation of the two residues in both the Michaelis complex and the resting state of FXIII using ONIOM-type QM/MM simulations. We performed two relaxed potential energy scans on the two systems, as described in the Methods section. The energy minima are located at 2.4 and 2.3 Å in the two cases, respectively, corresponding to a more stable ion pair form than the neutral state. The observed higher stability of the His-protonated form was also confirmed by zero-point vibration calculations. Due to the large energy difference, the probability of the latter form is likely very small. Thus the proton transfer and the nucleophilic attack can be regarded as two separate reaction steps. Thus, a mechanism in which the Cys-His proton transfer would occur in the same elementary step as the nucleophilic attack seems unlikely. This is different from the proposal of Case and Stein but is in good agreement with the findings for cysteine proteases.

6.1.3 Mechanism of the acylation step based on ONIOM-type calculations

In the family of cysteine proteases, different mechanisms were proposed for different enzymes, thus a single model for the entire family seems unlikely. The first scheme was proposed by Lowe, by analogy with the related serine proteases. Based on enzyme kinetics and theoretical chemistry-based studies, multiple different reaction mechanisms can be hypothesized: the nucleophilic attack of the Cys residue may either occur in the same step as the proton transfer from the His, or the proton transfer could even precede the attack. A tetrahedral intermediate (thio-hemiacetal) is present in some, but not all mechanisms. Regarding the human transglutaminase family, Iismaa et al. have proposed, based on mutagenesis studies that the catalytic efficiency decreases significantly if Trp249 (corresponding to FXIII-A Trp279) is mutated to other amino acids. They suggested that this residue may be involved in the stabilization of a tetrahedral (oxyanionic) intermediate.

To study the acylation step of the reaction, we performed potential energy scans using two different reaction coordinate pairs (d_1-d_2 and d_1-d_3; d_1: Cys314 sulfur-substrate amide carbon distance, d_2: substrate amide C-N distance, d_3: substrate amide N-proton distance) and two different methods (MP2/6-31G(d) and M06-2X/6-31+G(d,p)).) Regarding the two potential energy surfaces obtained using the d_1 (Cys314 sulfur-amide carbon distance) and d_2 (amide carbon-amide nitrogen distance) variables, the lowest energy state was determined to be the Michaelis complex (at approximately d_1 = 1.8 Å, d_2 = 1.7 Å). A transition state (TS 1) is probably present near d_1 = 2.0 Å, d_2 = 1.5 Å. A second local minimum can be observed around d_1 = 1.8 Å, d_2 = 1.7 Å, this corresponds to a zwitterionic, tetrahedral intermediate (I

1) with a protonated $–NH_3^+$ group. In the $d_1 = 1.8$ Å, $d_2 = 2.4$-2.6 Å region, the surface is remarkably flat. A second local minimum can be found in this region, especially in the energy surface based on M06-2X/6-31+G(d,p) calculations. This can be considered a second reaction step, the mechanism of which was investigated separately. In summary, the nucleophilic attack and the proton transfer from protonated histidine to the substrate amide were predicted to take place in the same reaction step, and the release of the ammonia from the intermediate (I 1) occurs in a separate reaction step. The shape of the two potential energy surfaces is very similar, but the energies relative to the Michaelis complex are consistently higher in the case of the M06-2X/6-31+G(d,p) calculations.

The same reaction step was further investigated with two additional potential energy scans, using the same two methods. The potential surfaces are compatible with a one-step, concerted reaction mechanism, in good agreement with our assumption based on our previous results. The end state is, however, the tetrahedral intermediate in both cases. Therefore the breakage of the C-N bond does not occur spontaneously during energy minimalization (unless it is defined as a reaction coordinate)

The breakage of the C-N bond, which is probably a second separate reaction step according to our previous potential energy scans, was investigated using two scans with only one reaction coordinate, the C-N distance. The difference between the two calculations was the QM method applied (see above). Based on the energies calculated, a second transition state (TS 2) assumed at d_2 = 2.3-2.4 Å. The energy barriers were calculated to be 23 and 27 kJ/mol, respectively. This corresponds to a rapid reaction step which is not rate limiting. Furthermore, the energy barrier explains why no spontaneous bond breakage was observed in the potential surface scans using the d_1-d_3 coordinate pair. The energy of the “thioester intermediate” state (I 2) is still higher than that of the tetrahedral intermediate (I 1), however, the dissociation of ammonia did not fully take place in the “final” state in our study. However, the entire process of ammonia dissociation into the solvent phase probably cannot be described using the simple reaction coordinates available in Gaussian.

In summary, the nucleophilic attack and the proton transfer from the protonated histidine was predicted to take place in a single reaction step, which is also the rate limiting step. The breakage of the C-N bond and the release of the ammonia is a separate, but very fast reaction step.

The geometries of stationary points, obtained from the scans were re-optimized using various methods and basis sets, with the restraints removed. The energy of the first transition state, relative to the Michaelis complex, was significantly lower in the MP2 than in the DFT calculations, the differences were between 23-33 kJ/mol in the wB97XD and 10-20 kJ/mol in the M06-2X optimized states, depending on the basis set. The energies of the tetrahedral intermediate were -29-35 and 20-25 kJ/mol higher, respectively. The zero-point vibration corrected energies state were in the 80.8-91.5 kJ/mol range for the transition state and 80.8-93.3 for the intermediate (compared to the Michaelis complex). The corrected energies for the two states are remarkably similar for nearly all methods and basis sets. Therefore we can assume that the half-life of this state may be very short if it exists at all. The uncorrected energies of the second transition state, relative to the intermediate were predicted to be 21-27 kJ/mol, the corrected energies are 4-6 kJ/mol lower. The energies of the "final" state were 3.0-10.6 kJ/mol higher than the intermediate.

The calculated activation energy values can be compared with the Eyring-Polányi equation with the experimental catalytic efficiency values (kcat). It must be noted, however that all kinetic data available for FXIII is based on kinetic measurements in which the two steps have not yet been investigated separately. Assuming the slowest reaction step, the Cys nucleophilic attack as rate limiting, the activation energy would be about 85 kJ/mol using the high-level MP2 calculations. Unfortunately the zero-point vibration corrections could not be computed at this level, but we can assume a correction of -15 - -10 kJ/mol based on the DFT data. A "corrected" energy barrier of 70-75 kJ/mol, corresponds to a kinetic constant 0.5-4 s-1. In contrast, the experimental values are in the range of 0.5 – 4 s-1 for various substrates. Probably more than one peptide residue is required for more efficient catalysis.

Furthermore, we analyzed the d_1-d_4 characteristic distances in the stationary points. The geometry parameters of the Michaelis complex are relatively constant independently from the method or basis set used. The geometries of the first transition state and the tetrahedral intermediate are very similar, the difference in d_1 and d_2 distances are small. In these structures the presence of a hydrogen bond Trp279 between and the substrate carbonyl oxygen was evident (corresponding to TG2 Trp241, known to be involved in transition state stabilization). In contrast, the amide nitrogen-proton distance (d_3) is significantly higher in the transition state than in the intermediate. In the zwitterionic intermediate the C-S distance (2.00 – 2.02 Å) is still significantly larger than a typical C-S bond (~1.82 Å). The d_1 and d_2 distances are generally smaller in structures optimized using basis sets containing diffuse

functions. Regarding the second transition state (TS 2), the C-N distance (d_2) falls in the range of 2.12 - 2.23 Å. The C-S distance is minimally higher than in the thioester form. Among all geometries, the d_1 distance in the Michaelis complex and the d_2 and d_4 distance in the thioester generally show larger variability than other parameters. The reason for this is that the substrate and the ammonia, respectively are non-covalently bonded in these two states.

6.1.4 QM/MM MD simulations

Additionally to the ONIOM-type calculations, the proton transfer between the Cys314 and His373 amino acids was also investigated using a QM/MM MD simulation, where the QM subsystem was treated using the semiempirical method DFTB3. In the semiempirical-based QM/MM MD simulations, the solvation was taken into account using explicit TIP3P solvent. In the trajectory of the 2 ns metadynamics simulation of Cys-His protonation, multiple proton transfer events have taken place in both directions. In most cases, the time intervals where the Cys residue was protonated were much shorter than the typical duration His-protonated states. The relative free energy of the two states was calculated from the output of the simulation, the energy difference was calculated to be 38 kJ/mol. In agreement with the results obtained using the ONIOM-type calculations, the zwitterionic form was much more stable than the neutral form. The conformations of the non-restrained amino acids remained relatively stable during the simulations.

Regarding the QM/MM MD simulation of the nucleophilic attack-ammonia-release step, a two-dimensional free energy surface was calculated based on the results of metadynamics simulation. The shape of the potential energy surface corresponds to a single-step, concerted reaction mechanism. The only transition state was located at approximately $d_1 = 0.22$ nm, $d_2 = 0.18$ nm and had 98 kJ/mol higher free energy than the Michaelis complex. The free energy difference is in the same range as the vibration-corrected energy differences in the ONIOM-type calculations. The thioester intermediate had an approximately 40 kJ/mol higher energy than the Michaelis complex. However, it must be noted that the ammonia release has not fully taken place, thus the actual energy of the intermediate is probably lower. Compared to the ONIOM-type calculations, the tetrahedral intermediate was extremely labile during the simulation, and no corresponding energy minimum could be observed on the potential energy surface. The nucleophilic attack, the proton transfer from the protonated histidine and the ammonia release took place in a single, concerted reaction step.

In the trajectory, the first nucleophilic attack event could be observed at 1.54 ns, followed by the formation of the thioester intermediate immediately. At around 3.05 ns, the reverse reaction took place (formation of the Michaelis complex). A further nucleophilic attack event occurred at 4.10 ns. The positions of non-restrained residues were stable until 4.3 ns, however, at this point the hydrogen bond between the Trp279 residue and the amide carbonyl group broke. Therefore the first 4.3 ns part of the simulation was taken into account in free energy calculations.

6.2 Mechanism of the pentasaccharide binding to antithrombin, based on MD simulations

6.2.1 Mechanism of pentasaccharide binding

For the pentasaccharide binding mechanism of AT, a three step-mechanism was proposed, based on X-ray diffraction structures and binding kinetics studies with various (altered or truncated) oligosaccharide derivatives. However, no X-ray structure is available that would correspond to the early events of binding or a relatively weak complex in this early stage. All such structures represent either a fully or “intermediate” activated state.

To investigate the binding mechanism, we performed GAMD simulations, according to the protocol described in the “5.2 Mechanism of the pentasaccharide binding to antithrombin” section. Among the four “non-activated antithrombin with added pentasaccharide” systems (A1, A2, B1 and B2), pentasaccharide binding with an RMSD for ring and interglycosidic atoms lower than 2.5 Å in a longer time interval was observed in two cases (A1 and B1). In these simulations, conformations with an RMSD lower than 1 Å occurred with a relatively high frequency. As for the A1 simulation, "strong" binding was observed especially in the last 300 ns of the trajectory, while in the B1 simulation, conformations closest to the X-ray structure were found mostly in the 450-800 ns time interval. From 1000 snapshots of the A1 and B1 trajectories, the conformation with the lowest RMSD compared to the 1NQ9 structure was chosen, and the selected snapshots were fitted to the X-ray structure. The same fitting was also performed for the lowest-RMSD conformation from the 1 μs 1NQ9 simulation. In these lowest-RMSD conformations, the binding was very similar to that in X-ray structure, the differences were slightly higher for rings G and H than rings D-F.

To study the binding mechanism of the pentasaccharide, the RMSDs for rings D, F and H were determined separately. These corresponded to the nonreducing end, the center and the reducing end of the molecule, respectively. We could observe significant similarities in the

binding mechanism in the case of the A1 and B1 simulation. Ring H could only reach a position comparable to that found in the X-ray structure when rings D and F already reached their final positions. However, in the case of the 1 μs 1NQ9 simulation, a different mechanism was observed: the "strong" binding of ring H did not depend on the position of ring F. The likely explanation is that the 1NQ9 structure contained a pentasaccharide and therefore the binding site was already in a more “favorable” conformation for binding. The mechanism observed in the case of the A1 and B1 systems shows significant similarity to a proposal by Desai et al., based on kinetic measurements with various penta- tetra- and trisaccharide derivatives. Based on their experiments, they proposed a mechanism in which the DEF subunits bind first, inducing a conformation change of the binding site for the GH part.

The positively charged amino acids in the AT binding site play an essential role in the interaction with the highly negatively charged pentasaccharide. The three most important residues are Lys114, Lys125, and Arg129, but Lys11, Arg13, Arg46 and Arg47 also contribute significantly to the binding energy. We analyzed the presence or absence of interactions between the positively charged AT residues and the negatively charged groups of the pentasaccharide, as a function of time. Two groups were considered to interact with each other if the distance between them was less than 5 Å. According to the analysis, many salt bridge interactions are present even if the pentasaccharide is not exactly in the position found in the X-ray structure. A possible explanation for this observation is the high flexibility of the Arg and Lys residues. However, the presence of the electrostatic interactions in conformations with a relatively high RMSD does not mean that the energy is comparable to the energy minimum. The AT-pentasaccharide interaction has a significant hydrophobic component, and the hydrophobic interaction depends strongly on the correct positioning of AT residues.

6.2.2 Conformational changes of antithrombin

The details of the AT activation mechanism were mostly inferred from X-ray diffraction studies. The steps of the activation process can be classified into two groups, based on whether these changes are present in the “intermediate” activated or only in the “fully” activated state. In the first step of AT conformational activation, conformational changes occur near the A and D helices, including the pentasaccharide binding site. The formation of the helix P, near the N-terminal end of helix D, is considered to occur in this step. However, there exists an X-ray diffraction structure of non-activated AT in which this helix is present.

The most important changes in the second step are the hinge region expulsion, the elongation of helix D, these changes lead to stronger binding of the pentasaccharide to AT.

Previously, we have attempted the modeling of helix P formation, starting from the 1E04 X-ray structure that does not contain this structural element. Despite the use of an advanced sampling technique, we were not able to observe the formation of the helix. As an alternative solution, we have chosen a non-activated AT structure for the ligand binding simulations that contained a pre-formed helix. The DSSP analysis of this part of the molecule also supports the high stability of this structural element. In one of the simulations without a pentasaccharide, however, a small change in the conformation was detected at the C-terminal end of the helix. This small change does not contradict the hypothesized high stability of this structural element but may provide some limited information on its formation mechanism. We have confirmed our previous results obtained by "conventional" MD simulation with an advanced sampling MD technique. We can conclude that such conformations are probably present under physiologic conditions and they can are probably involved in the binding,

Recently, it has been suggested the expulsion of the hinge region takes place in the last step of the conformational activation of AT. In contrast, Langdown et al. have proposed a rapid equilibrium between the “closed” and the “open” conformations of this region. They also suggested that the closure of the β-sheet A may lock the protein in the "open" conformations.

The distance between the α-carbon atoms of Val375 and Ser380 was used to describe the conformation of the hinge region in our molecular dynamics trajectories. This parameter can be used to distinguish between the two forms, values below approximately 6 Å correspond to the "closed" state region expulsion was observed in all but two trajectories. Re-insertion of this region has taken place in three simulations, including the A1 and the 1 μs 1NQ9 system.

The relationship between the hinge region expulsion and the pentasaccharide binding was investigated using free energy calculations (reweighting) of the GAMD data. The two reaction coordinates were the Val375-Ser380 distance and the RMSD pentasaccharide. This calculation was performed on all GAMD simulations containing a pentasaccharide. Two further free energy calculations were performed, using data from all four “non-activated AT with added pentasaccharide” and the two “1NQ9” simulations. Based on the reweighting data, the hinge region expulsion can take place in all activation states of AT, not just in the last step of the activation process. It is unlikely that the A1 and the B1 systems have reached a full

activated state due to the lack of helix D extension. This latter conformation change was, however, observed in the final part 1 μs 1NQ9 simulation, this may represent a fully activated state. The conformational change required for the fully activated state may be the closure of the β-sheet A.

A significant expansion of helix D was observed in only one simulation that was based on the AT-pentasaccharide complex structure 1NQ9. The conformational change started at approximately 720 ns. After 850 ns, most analyzed conformations contained an “elongated” helix. In this state, the helix is approximately one amino acid shorter than that found in the fully activated AT structures, such as 1E03. Since this conformational transition has occurred at the same time as the hinge region expulsion, we can conclude that AT has reached a "fully activated" state in this simulation.

The reactive center loop (RCL) plays an essential role in the inhibition mechanism of the serpins. However, the X-ray diffraction structures can only provide a limited amount of information, because these structures can be considered a" static" conformation. Furthermore, crystal contacts, not found in the solution phase affect the conformation of this highly flexible loop drastically. Several X-ray structures of AT actually correspond to the dimer of a “native” and a “latent” AT molecule. The RCL of "native" AT forms contacts with the “latent” molecule, altering its conformation significantly. Johnson et al. have published an AT structure in which the transition into the latent state was prevented by an engineered disulfide bridge. However, in this structure a new conformation of the RCL was observed, significantly different from that found in previous structures. In a previous study from our group, we have investigated the conformation of this loop using long molecular dynamics simulations, starting from both conformational types. In these simulations, we could not observe a transition from either state into the other. In contrast, we were able to simulate this process in one direction using a metadynamics simulation.

The distance between the α-carbon atoms of residues Arg236 and Ile390 was used for describing the conformational behavior of the RCL, as this parameter can be used to distinguish between the two conformational types. At the beginning of the A1 and B1 simulations, distance values were well below 15 Å. This corresponds to the position of the loop found in the 1T1F structure. At 270 ns in simulation A1 and at 140 ns in simulation B1, a significant conformational change could be observed. From this point, the Arg236.CA – Ile390.CA distance was above 15 Å in most snapshots analyzed. In the A1 trajectory, the

parameter fluctuated significantly, but it was in the same range as the distances from the 1 μs 1NQ9 simulation. In the B1 simulation, however, the conformation of this loop remained relatively stable. A large number of different conformations were present in the 1 μs simulation of the 1NQ9 system, but the distance parameter fell in the 15 – 18 Å range only a small fraction of them. None of the snapshots analyzed had an RCL conformation comparable to the 1T1F X-ray structure by Langdown et al.

However, despite the advanced sampling technique, either no or only a single conformational transition could be observed in the trajectories. A likely explanation of this is that some energy barrier slows down the "mixing" of the two conformational types, even in advanced sampling simulations. To investigate the RCL conformational changes further, we calculated free energy surfaces using GAMD reweighting, with two reaction coordinates (the position of RCL and the "hinge" region). The presence of such an energy barrier is evident from the energy surfaces. As an alternative explanation, the sampling of the conformational space was insufficient to observe such a transition.

6.2.3 Allosteric processes

The binding of the pentasaccharide to AT triggers allosteric processes that result in more efficient inhibition of factors Xa and IXa. The study of such allosteric pathways provides invaluable information for understanding the allosteric activation. In the present work, these pathways were investigated based on our GAMD trajectories, using a method proposed by Lange and Grubmüller. In the case of the A1 simulation, correlated motions involving the 230-310 amino acids of the protein as well as helix D were detected. Several amino acids found in this region (Asn233, Arg235, Glu237, Tyr253, and Glu255) are known to be involved in the interaction with FXa and FIXa. A further region that may be involved in allostery belongs to the hydrophobic core of the protein (amino acids 75-78). The 233-253 region, known to be involved in factor Xa binding, shows correlated motions with the C-terminal of AT (amino acids 400-432), but these regions are located close to each other in the three-dimensional structure. Regarding the 1 μs 1NQ9 simulation, already containing a pentasaccharide, the matrices of correlated motions are significantly different from the two systems discussed previously. Here, correlated motions between the previously mentioned exosite, the RCL, the C-terminal part of the protein and the majority of amino acids in the region 110-230. The latter region includes the D, the E and the F helices as well as two

strands from β-sheet A. Similar correlated motions were observed with the region consisting of amino acids 75-95, including helix B.

To study the effects of pentasaccharide binding on the flexibility of various amino acids in AT, we have performed RMSF calculations. Based on the RMSF data, the most flexible parts of the molecule are those not having a well-defined secondary structure. Greater than average flexibility was detected in the case of the C-terminal end of helices D and F (amino acids 190-210), as well as in some residues close to the secondary binding site FXa and FIXa (235-240). A further RMSF calculation was performed on those parts of the three 1 μs simulations, where most low-RMSD conformations were detected (A1: 640-1000ns, B1: 440-800ns, 1NQ9: 720-1000 ns.) The fluctuation was decreased near the ligand binding site (amino acids 110-140), in both the A1 and the B1 simulations. This may indicate the stabilization of conformations interacting stronger with AT. Furthermore, changes in RMSD were observed in region 220-240. This is more apparent in simulation A1 than B1. Several residues are found in this region (Asn233, Arg235, Glu237) known to interact with FXa. The decrease of the RCL flexibility may be surprising as increased flexibility of this loop has been reported in a previous paper when a ligand was bound. However, in our study, the exclusion of many conformations from the analysis could explain this decrease in RMSF. The RCL remains one of the most flexible parts of the molecule even in this case.

7. Conclusion

In the present work, we have investigated the first, acylation step of the catalytic mechanism of factor XIII-A, using hybrid QM/MM techniques. To our knowledge, no reaction mechanism supported by such calculations is available for the human transglutaminase family (Unlike the cysteine proteases, an enzyme family with a similar catalytic triad). We aimed to determine the protonation state of the catalytic Cys314 and His373 residues and to investigate the mechanism of the nucleophilic attack and ammonia release steps.

The cysteine thiolate-protonated histidine form was calculated to be significantly more stable than the neutral form using both QM/MM based methodologies. Due to the large energy difference, the probability of the Cys-protonated form is probably very low. Therefore, a mechanism in which the Cys-His proton transfer would occur in the same elementary step as the nucleophilic attack could be considered unlikely. As for the acylation step, we proposed a two-step mechanism using the ONIOM-based protocol. The nucleophilic attack of the Cys314 sulfur atom and the proton transfer to the amide group occurred in the first step while the dissociation of the ammonia molecule has taken place in a second, relatively fast reaction step. In contrast, all three steps have taken place in a highly concerted manner in our QM/MM based metadynamics simulation. The differences in the results obtained by the two different protocols can be explained by the usage of a semiempirical method in the case of the QM/MM MD simulation. Alternatively, the usage of TIP3P explicit solvent in the QM/MM MD simulation could also be a contributing factor. Our reaction mechanism also differs from the proposal by Iismaa et al. In their proposal, the intermediate was oxyanionic with a neutral $–NH_2$ group rather than zwitterionic (in the ONIOM-type calculations). In our QM/MM MD simulations, we could not even detect such an intermediate.

We investigated the binding mechanism of the pentasaccharide to AT using both “conventional” and advanced sampling MD (GAMD) simulations. For much of the simulations, we have chosen an AT conformation that does not contain a pentasaccharide or any other activator as the starting structure. In two independent GAMD simulations, we were able to observe bound states with “pentasaccharide RMSD” values below 1 Å compared to the 1NQ9 X-ray structure of the complex. Based on these two trajectories, as well as further simulations of the “intermediate” activated AT-pentasaccharide complex (1NQ9) we could gain insight into several conformational changes triggered by the pentasaccharide binding, including the expulsion of the “hinge” region and the elongation of helix D. The P helix was

very stable in all simulations, including those not containing a pentasaccharide, even if an advanced sampling method was used. This confirms our previous results from “conventional” MD simulations. From these results, we can conclude that such states might exist at physiological conditions and might play a role in the binding of heparins and heparinoids. The expulsion of the hinge region, as well as its re-insertion into the β-sheet A was observed in multiple trajectories. In contrast to several proposals in the literature, the “fully activated” state of AT was not a requirement of this conformational change. This conclusion was drawn from our observations that other conformational changes described in the same activation step (e.g. D helix expansion) did not take place. But our results are in agreement with a different model in which the two conformations of the hinge region was proposed to be in a rapid equilibrium, and the equilibrium is influenced by the activation state of AT The C-terminal expansion of the helix D was observed in only one simulation in which the partially activated AT structure (1NQ9) was used as the starting conformation. The GAMD technique also allowed better conformational sampling of the reactive center loop (RCL).

The applicability of the GAMD method for ligand binding studies has already been demonstrated in the literature. In the present work we have demonstrated the usability of the GAMD technique for studying the binding of a heparin pentasaccharide analog. This molecule has several characteristics that differ significantly from the ligands used previously in such studies (large flexibility due to carbohydrate structure, high negative charge). The method might be useful for comparing the binding of different pentasaccharides to AT, and it could provide valuable information for the development of new AT-dependent anticoagulants. Furthermore, several mutations have been described in the human AT gene that cause impaired heparin binding. Simulations with this technique may provide insights into the mechanisms how these mutations affect heparin binding.

8. New results of the PhD thesis

First step of FXIII-A catalytic mechanism

- In the present work, we proposed catalytic mechanisms, supported by two different QM/MM methodologies the first time for a human transglutaminase enzyme: activated Factor XIII A subunit.
- Regarding the Cys314 and the His373 amino acids of the catalytic center, the ion pair form was predicted to be significantly more stable than the neutral form, using two different QM/MM protocols. This differs from the proposal by Case and Stein for transglutaminase 2, but is in good agreement with the results available for various cysteine proteases, enzymes with a related catalytic triad.
- In our ONIOM-based calculations, where the MP2 method or DFT methods were applied to the QM subsystem, the acylation step was found to be a two-step process. In contrast, in our QM/MM MD calculations, all three events (nucleophilic attack, proton transfer, ammonia release) occurred in a single elementary step. The different mechanisms obtained are presumably explained by methodological differences.

Mechanism of the pentasaccharide binding to antithrombin

- We have investigated the binding of pentasaccharide derivative idraparinux to a non-activated AT conformation determined by X-ray crystallography.
- We observed a significant number of conformations in two independent trajectories with low RMSD for the ligand compared to X-ray diffraction structure of the complex.
- Using GAMD simulations for two types of model systems (“non-activated AT with added pentasaccharide”, “intermediate activated”), we investigated the mechanism of the „hinge” region expulsion, the conformational changes of helix D as well as the conformational behavior of the RCL.
- We could confirm the high stability of the P helix in non-activated AT conformations. This strongly suggests that such conformations might play an important role under physiological conditions.
- Using “generalized correlation” matrices, regions of the AT molecule have been identified that play a role in the transmission of the allosteric signal form the pentasaccharide binding site to the binding exosite for FXa and FIXa.

9. Summary

The A subunit of factor XIII (FXIII-A) is a multifunctional enzyme belonging to the human transglutaminase family. Although a proposed reaction mechanism has been published for these enzymes, no QM-based calculations have been performed to confirm this. Additionally, the mechanism was studied in details in case of transglutaminase 2, but not in the case of FXIII-A. Our aim was to propose a mechanism for the first step of the reaction (*i.e.* formation of a thioester intermediate), supported by QM/MM based calculations. The reaction mechanisms were investigted using two different QM/MM based protocols: an ONIOM-based model where high level MP2 or DFT methods were applied to the QM subsystem, and QM/MM MD simulations, where the QM region was treated using the semiempirical method DFTB3. In our calculations using both protocols, the ion pair form of Cys314 and His373 was predicted to be significantly more stable than neutral form. The formation of the thioester intermediate was calculated to be a two-step process using the ONIOM protocol, the nucleophilic attack and the ammonia release took place in the first, the ammonia release in the second step. However, in our QM/MM MD based metadynamics simulations, all three events occurred in the same reaction step.

Significant amount of information is available on the allosteric activation mechanism of the serpin antithrombin (AT), which is one of the main regulators of the coagulation cascade. However, the „static" structures obtained by X-ray crystallography cannot provide the full picture on the complex allosteric process. Furthermore, only kinetic measurements are available for the early events of binding. In the present work, the pentasaccharide-binding and conformational behavior of AT was studied by the means of an „advanced sampling" MD technique, GAMD. Using a non-activated AT conformation as starting strucutre, we could observe AT conformations with low RMSD compared to the X-ray structure of the complex in two independent MD trajectories. From these trajectories, as well as additional ones based on the AT-pentasaccharide complex structure, we could gain insight into the conformation changes discussed in the literature, including the expulsion of the hinge region and the expansion of helix D. We have also confirmed the high stability of the P helix in non-activated AT conformations, that was previously proposed based on „equilibrium" MD simulations. Using „generalized" correlation calculations from the trajectories, we could gain insight into the allosteric pathways connecting the pentasaccharide binding site with the FIXa and FXa binding exosites.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/10/2020.PL
Subject: PhD Publikációs Lista

Candidate: Gábor Balogh
Neptun ID: FCP6BD
Doctoral School: Kálmán Laki Doctoral School

List of publications related to the dissertation

1. **Balogh, G.**, Muszbek, L., Komáromi, I.: First Step of the Transglutaminase Reaction Catalyzed by Activated Factor XIII Subunit A, Hybrid Quantum Chemistry/Molecular Mechanics Calculations.
   J. Phys. Chem. B. 123 (18), 3887-3897, 2019.
   DOI: http://dx.doi.org/10.1021/acs.jpcb.9b00542
   IF: 2.923 (2018)
2. **Balogh, G.**, Komáromi, I., Bereczky, Z.: The mechanism of high affinity pentasaccharide binding to antithrombin, insights from Gaussian accelerated molecular dynamics simulations.
   J. Biomol. Struct. Dyn. [Epub ahead of print], 2019.
   DOI: http://dx.doi.org/10.1080/07391102.2019.1688194
   IF: 3.31 (2018)

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

List of other publications

3. **Balogh, G.**, Gyöngyösi, T., Timári, I., Herczeg, M., Borbás, A., Fehér, K., Kövér, K. E.: Comparison of Carbohydrate Force Fields Using Gaussian Accelerated Molecular Dynamics Simulations and Development of Force Field Parameters for Heparin-Analogue Pentasaccharides.
J. Chem Inf. Model. 59 (11), 4855-4867, 2019.
DOI: http://dx.doi.org/10.1021/acs.jcim.9b00666
IF: 3.966 (2018)

4. Gindele, R., Selmeczi, A., Oláh, Z., Ilonczai, P., Pfliegler, G., Marján, E., Nemes, L., Nagy, Á., Losonczy, H., Mitic, G., Kovac, M., **Balogh, G.**, Komáromi, I., Schlammadinger, Á., Molnárné Rázsó, K., Boda, Z., Muszbek, L., Bereczky, Z.: Clinical and laboratory characteristics of antithrombin deficiencies: a large cohort study from a single diagnostic center.
Thromb. Res. 160, 119-128, 2017.
IF: 2.779

5. Tóth, L., Fekete, A., **Balogh, G.**, Bereczky, Z., Komáromi, I.: Dynamic properties of the native free antithrombin from molecular dynamics simulations: computational evidence for solvent-exposed Arg393 side chain.
J. Biomol. Struct. Dyn. 33 (9), 2023-2036, 2015.
DOI: http://dx.doi.org/10.1080/07391102.2014.986525
IF: 2.3

Total IF of journals (all publications): 15,278
Total IF of journals (publications related to the dissertation): 6,233

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

10 January, 2020

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár