Short Thesis for the Degree of Doctor of Philosophy (PhD)

IN SILICO STUDIES ON THE INHIBITION MECHANISM OF HUMAN CYCLOOXYGENASE-1 BY ASPIRIN AND ON THE ACTIVATION AND CONFORMATIONAL TRANSITIONS OF ANTITHROMBIN

László Tóth

Supervisor: Dr. István Komáromi, PhD

UNIVERSITY OF DEBRECEN
KÁLMÁN LAKI DOCTORAL SCHOOL

Debrecen, 2015

IN SILICO STUDIES ON THE INHIBITION MECHANISM OF HUMAN CYCLOOXYGENASE-1 BY ASPIRIN AND ON THE ACTIVATION AND CONFORMATIONAL TRANSITIONS OF ANTITHROMBIN

By László Tóth, MSc (chemist)

Supervisor: Dr. István Komáromi, PhD

Kálmán Laki Doctoral School

Head of the Examination Committee:	Prof. Dr. György Balla, MD, PhD, DSc, MHAS
Members of the Examination Committee:	Prof. Dr. Ferenc Erdődi, PhD, DSc Dr. Ferenc Bogár, Phd

The Examination takes place at the Library of Paediatric Institute, Faculty of Medicine, University of Debrecen, 27th May, 2015, at 11 a.m.

Head of the Defense Committee:	Prof. Dr. György Balla, MD, PhD, DSc, MHAS
Reviewers:	Dr. György Ferenczi, PhD, DSc Dr. Attila Bényei, PhD
Members of the Defense Committee:	Prof. Dr. Ferenc Erdődi, PhD, DSc Dr. Ferenc Bogár, PhD

The PhD Defense takes place at the Lecture Hall of Bldg. A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen, 27th May, 2015, at 1 p.m.

1. INTRODUCTION

Hemostasis is a complex multi-component system, whose malfunction cause haemophilia or thrombosis. Owing to the complexity of this system it has several intervention points where one can influence it, e.g. in order to prevent thrombosis. In the following, our attention will be focused on two of the actors playing crucial role at these points.

1.1. Aspirin and COX-1 enzyme

The first commercially available synthetic antipyretic, analgesic and anti inflammatory drug was aspirin (acetylsalicylic acid or ASA), the distribution of which started as early as the end of 19th century by Bayer. It was proposed later that similarly to other non steroid anti inflammatory drugs it suppresses prostaglandin generation [1-3]. Since aspirin suppresses prostaglandin H2 production, a precursor of thromboxane A2, a potent activator of platelets, it leads to the inhibition of thromboxane A2 production in platelets. ASA achieves this effect by acetylating a single serine residue (Ser530) of cyclooxygenase-1 (COX-1). This mechanism is the basis of the protective effect of aspirin against thromboembolic vascular events. However, the exact reaction mechanism of COX-1 Ser530 acetylation has not been revealed by experimental methods.

COXs, also known as prostaglandin H synthases (PGHSs), are bifunctional enzymes with two spatially distinct reaction centers, which are coupled mechanistically. The cyclooxygenase center catalyzes the conversion of arachidonic acid to prostaglandin G2 (PGG2, a hydroperoxide endoperoxide prostaglandin), while the peroxidase reaction center is responsible for PGG2 reduction to prostaglandin H2 (PGH2). PGH2 is then converted to a variety of prostanglandins and thromboxane A2 by specific enzymes. Two main isoforms of COXs, COX-1 and COX-2, are known, although sometimes the splice variant of COX-1 is termed as COX-3. COX-1 known to be responsible for PGH2 production which is the common starting-point of

prostaglandin derivatives involved in basic housekeeping functions. Although COX-2 also plays a role in certain "normal" physiological processes, its expression is dominantly induced by pathophysiological processes, like inflammation and it is a target of non-steroidal anti-inflammatory drugs.

Both COX-1 and COX-2 are membrane-bound homodimers of ~70 kDa subunits which consist of a short EGF domain, a membrane binding domain and a large C-terminal catalytic domain. Although they share only ~60% amino acid identity their structural similarity is remarkable. One of the characteristic features of COX structures is the long hydrophobic channel with an entrance in the membrane-binding domain. The active site sits at the bottom of the channel. Residues R120, Y355 and E524 separate the active site from the front of the channel. Structure based residue numbering scheme applied generally for both isoforms of cyclooxygenases corresponds to sheep COX-1 for which the first COX 3D structure was resolved.

It was pointed out that the reaction of ASA with nucleophilic reagents occurs via direct nucleophilic attack on phenolic ester carbonyl carbon atom instead of via anhydride intermediate. The same mechanism seems to be straightforward for the ASA-COX reaction, as well. Molecular modeling using SCC-DFTB method in QM/MM dynamics supports this assumption.

Several amino acid residues of COX-1 and COX-2 found to be essential in the ASA-COX interactions leading to trans-esterification reaction. It was revealed by experiments using 14C-acetyl labeled aspirin with the wild type and the point mutated COX-2 enzyme, that the role of Y385 is crucial, while R120 and Y348 residues influence the trans-esterification reaction to a somewhat lesser extent. Interestingly, the Y385 also plays essential role in cyclooxygenase reaction while Y348F, Y355F and R120A mutations only partially suppress cyclooxygenase activity. The X-ray structure of the complex of acetylated enzyme and the product, salicylic acid, demonstrated that salicylic acid is ~5Å far from its "ideal" transition-state position. This finding suggests that the role of R120 is to ensure the proper orientation of ASA, rather than direct participation in the reaction.

1.2. Aims of ASA-COX-1 study

Only one paper was found in the literature in which the mechanism of irreversible inhibition of the COX-1 enzyme by aspirin, i.e. the mechanism of the corresponding transesterification reaction was investigated in details. Even though the reaction mechanism proposed by Tosco et. al. seems reasonable, it requires validation by sophisticated theoretical and/or experimental methods. While QM/MM dynamical reaction path mapping provides, in principle, realistic view on the reaction mechanism, this approach is suffering from the excessively approximate nature of the embedded quantum chemical methods used for such simulations. From this point of view, applying higher levels of quantum chemical theories in static calculations have a great importance as complementary studies to the dynamical reaction modeling. Since our long-term plan was a structure-function studies on the human COX-1 enzyme and 3D structure for this enzyme was not available, at first we had to construct it by means of homology modeling.

Therefore our goal was to elucidate the reaction mechanism of irreversible inhibition of cyclooxygenase-1 by aspirin using static ONIOM type QM/MM methods. While we intended to use "high level" *ab initio* and DFT calculations on a model system we also aimed at taking into account the steric and electrostatic effect of the neighboring amino acids in a relatively large surrounding region. This was achieved by using the electronic embedding method in ONIOM, which allows the polarization of the wave function by the surrounding partial charges.

Since in some cases it can be assumed that water molecules are present at the active site, their role on the reaction mechanism, were investigated as well. We also aimed at finding the "exact" transition state(s) and local minima on the potential energy surface and proving their existence by vibration analyses. In addition, series of calculations at different levels of theories were also intended to be carried out in order to assess the dependency of the results on the method and on the basis sets applied in quantum chemical calculations. We intended to decide whether the above methods predicted a

single elementary step reaction or a multiple steps reaction connecting the ASA-COX-1 complex to the salicylic acid and acetylated COX-1 product.

1.3. Antithrombin

Antithrombin III (antithrombin, AT) is the most abundant serpin circulates in blood vessels and it plays a central role in blood clotting cascade. Serpins (serine protease inhibitors) are the main physiological inhibitors of serine proteases in tightly regulated proteolytic pathways in higher organs. Specific examples are the haemostasis and fibrinolysis where numerous serpins are involved in the regulation of the complex cascading serine protease reactions in order to maintain the proper blood flow and to prevent bleeding at the same time. Antithrombin in its native form only a poor inhibitor of blood coagulation proteases, which is a necessary condition to avoid unwanted bleeding which would be the consequence of the high concentration of antithrombin in blood plasma.

The 3D structure of AT consist of nine helices (A-I) and three β-sheets (A-C), which are highly conserved in this protein family. The typically ~20 residue long "reactive center loop" (RCL) is functioning as bait for proteases. The protease cleaves the P1-P1' peptide bond, which is located in the middle part of the RCL, though closer to its C-terminal end. At first the acyl-enzyme intermediate is forms. However, the hydrolysis of acyl-enzyme intermediate cannot take place because the RCL undergoes large-scale conformational change and incorporates into the β-sheet-A as the 4th strand. This process results in a damaged catalytic center which prevents further hydrolysis. The RCL insertion into β-sheet-A, which is common in all serpin mechanism, shifts the protease position by ~70 Å compare to the AT.

Specific mechanisms need to attain its full (two or three order of magnitude higher) inhibitory potency only when and where it needs, e.g at the injury of blood vessels. These mechanisms are the bridging mechanisms against thrombin and allosteric activation against FXa and FIXa. The bridging mechanism requires longer

chain heparin or heparane sulfates while for allosteric activation a sulfated pentasaccharide fragment is sufficient. It should be noted, that bridging mechanism can play a role in inhibition of FXa and FIXa as well but for this the presence of heparin chain with approximately twice the length of heparin chain capable bridging the thrombin to antithrombin would be required (unfractioned heparin).

Until now several antithrombin X-ray structures with and without sulfated pentasaccharide or heparin fragments and even in ternary complex with serine protease and pentasaccharide have been revealed. From these structures a consensus mechanism for achieving full inhibitory potential of antithrombin against FXa and FIXa which comprises three distinct steps has been proposed. In accordance with this model at first the pentasaccharide ligand loosely attached to the strongly negative surface domain of the antithrombin. At the next step a stronger binding is evolving accompanied by conformational changes mainly at the ligand binding area. It is followed by an additional conformational transition during which the conformational changes from the ligand binding site is spread out to the hinge region part of the antithrombin resulted in the hinge region expulsion from the β-sheet-A and an even more stable antithrombin-pentasaccharide complex.

While these resolved X-ray structures provide us invaluable information on the allosteric activation mechanism, these are basically static snapshot pictures from complex processes. Clearly, the structures obtained this way can reflect not only the "free" protein structure but the protein-protein interaction forces in crystals as well. In other words, the crystal forces select from the many co-existing conformation which is most suitable for crystal packing. A good example for the existence of the just-mentioned effect the co-crystallization of the sulfated glycosaminoglycan free native antithrombin with its latent form which results in substantially different conformation for the reactive center loop than it is observed at the case of the "pure" native antithrombin crystals. It has to be mentioned that neither of these conformation can be applied to explain how the sulfated glycosaminoglycan free antithrombin displays small but non-negligible inhibitory activity against FXa. Namely, in these structures

the Arg393 side chain pointing to the inner side of antithrombin and not accessible by the active site of FXa. Therefore if one would need to explain the basal, however non-negligible, activity of AT against serine proteases, at least existence of one additional conformation has to be supposed.

Despite the several 3D X-ray AT structure available from the Protein Data Bank (PDB), only very limited indirect information can be derived from them for the activation and for the strongly coupled conformational changes of AT. Little is known about the role of monosaccharide units of pentasaccharides (PS), and the propagation of allosteric signal caused by PS binding. On the other hand, the importance of *in silico* methods in such kind of structural studies of proteins is continuously growing.

1.4. Aims of antitrombin study

Owing to the size of antithrombin obtaining detailed experimental information with sufficient time and space resolution on dynamical properties of antithrombin is still beyond the limits of nowadays experimental (e.g. nuclear magnetic resonance) technics. Moreover, the coexistence of different conformational states would make the extraction of atomic resolution structure and its progression as a function of time to be extremely difficult from these experiments.

Therefore we intended to carry out *in silico* molecular dynamical experiments, at first on the antithrombin itself, in order to reveal the dynamical property of AT. First of all, we wanted to reveal the flexibility of the RCL and to see whether RCL conformation which suitable for a Michaelis-type complex formation can exist during the simulation interval. Our goal was also to gain information on the flexibility of pentasaccharide binding region since it is known that remarkable conformational changes should take place at this site for a sufficiently strong PS binding. In connection with this we were wondering whether is there any sign for a new P-helix formation or even for the D-helix extension in the dynamical trajectory of the PS-free AT.

In connection with this we were wondering whether is there any sign for a new P-helix formation in the dynamical trajectory of the PS-free AT. We were wondering also whether in the simulation trajectory can be found any correlation between the movements of the RCL-hinge region and the pentasaccharide binding region.
At an allosteric regulation the information on the local interaction should spread out from the allosteric site to the site of action. The information theory based "mutual information" can be extracted from the dynamical trajectories and it reveals the regions which have information from the other ones. It could help us to understand how the information on the established allosteric interaction can have effect on the other part of the biomacromolecule. We were wondering on the preexistence of such information channels at the free native (i.e. PS is not bound) AT.

2. APPLIED METHODS

2.1. Aspirin and COX-1 enzyme

2.1.1. Homology modeling and docking

As there was no experimental atomic resolution 3D structure available for human COX-1 enzyme it was derived based on the ovine COX-1 (PDB ID# 2ayl) as a template by means of the YASARA software. ASA was then docked to the active site of human COX-1 using the Autodock 4.0 software with its graphical interface. We applied Lamarckian genetic algorithm using rigid receptor-flexible ligand protocol with 0.177Å grid spacing distance for a grid box (98x98x122 grid points) centered at S530 Oγ.

2.1.2. Constructing the model for QM/MM calculations

The surrounding region of the reaction center was selected as a set of residues, which had at least one atom closer than 7Å to either the S530 Oγ or the ASA carbonyl C atoms. The open N- and C ends of chain fragments, arose this way, were closed by

acetyl or N-methyl groups, respectively. The selected region included 81 residues including 61 “real” amino acids and 20 acetyl and N-methyl groups. The protein data bank type files were converted to Gaussian style of input files with AMBER force field atom type and corresponding partial charges by means of a small utility program was written for this purpose.

Two main input groups were created for the 81 residues. In the first one (model A) only the S530 and Y385 side chains and the ASA molecule were allowed to relax during optimization. The inner layer, on which high-level (i.e. quantum mechanical) calculations were carried out, comprised only of ASA and the side chain of S530. In the second series of calculation (model B) the side chains of Arg120, Phe198, Phe205, Phe209, Tyr348, Val349, Leu352, Ser353, Phe381, Leu384, Tyr385, Trp387, Phe518, Ile525, Leu534 and the whole S530 residue as well as the ASA molecule were ‘active’ in geometry optimization, while all the other atoms were clamped. The inner layer in this group comprised the ASA molecule, the side chain of Y385 and the S530 residue completed with the acetyl and N-methyl group to replace the preceding and following residues, respectively.

In spite of its highly hydrophobic character, X-ray crystallography frequently reveals structural water(s) inside the substrate binding channel of COX. Therefore calculations were carried out also on models that included one or two water molecules close to the S530 residue. It was supposed that water molecules contribute either to the stability of the transition state or to the proton transfer from S530 Oγ. The model B was then divided into four subgroups. In model B1 no water was included in the calculations, models B2 and B3 contained one water molecule, while in model B4 both water positions were filled.

2.1.3. QM/MM calculations

A series of ONIOM-type QM/MM calculations applying electronic embedding approximations were carried out at different levels of theories ranging the QM methods from HF/6-31G to MP2/6-31G(d,p). Calculations on model A were carried

out only to map the potential energy surface as a function of the two main internal coordinates, the distance of the breaking S530 Oγ-Hγ bond and the forming Oγ-C(acetyl carbonyl of ASA) bond. The selection of starting geometries for transition state search in subsequent calculations was based on the shape of potential energy surface.

For model B1 calculations were carried out at HF, B3LYP and MP2 levels of theories using different kind of standard basis sets. For the model of ASA-COX-1 complex, the transition state and the "product" the optimized geometries were determined at each level of theories we considered. Zero point vibration energy (ZPVE) corrections to these energy values were calculated at all levels, but MP2. The existence of (first order) saddle points and the local energy minima were proven by the analytically calculated one or zero imaginary frequency. The first local minimum the optimizer found after the breaking of C(carbonyl of ASA)-O(phenolic) bond was regarded as the "product". The calculations were carried out by the Gaussian '03 and '09 suites of softwares. Visualizations were performed by the Molekel, Chimera 1.4, VMD 1.9, DPlot and ESPript software packages. Models B1, B2, B3 and B4 comprise 988, 991, 991 and 994 atoms, while the QM sub-system without water and with one and two water molecules comprises 56, 59 and 62 atoms, respectively.

2.2. Antithrombin

2.2.1. Computational details of molecular dynamics simulations

Three systems were considered in our simulations. The first one is based on the 1e04 PDB structure and represented the conformation which is observed when the native form is co-crystallized with the latent one (Simulation 1). The second one is based on the 1t1f PDB structure representing the AT conformation which is crystallized in monomeric form and featured significantly different RCL positions relative to the body of AT. Since, in this case, the engineered Ser137Ala-Val317Cys-Thr401Cys mutations were introduced that ensured a new disulfide bridge formation

in order to prevent the native-latent transition, the Cys317and Cys401 residues were virtually mutated back to the residues present in the wild-type sequence (Simulation2). The third simulation was carried out on the 1t1f Ser137Ala/Val317Cys/Thr401Cys mutant (Simulation 3). In order to carry out molecular dynamics simulation for the whole AT protein chain the missing peptide sections in the AT structures retrieved from the protein data bank had to be completed. These were done either by the automatic loop generation capability of the MODELLER software (1t1f) or by the Schrödinger package (1e04). Structural water molecules were kept according to the original X-ray file. The model was then solvated in a dodecahedral box. The distance between the box wall and the closest protein atom was set to 12 Å at the 1e04 and 15 Å at the 1t1f.The latter value was chosen because of the more compact structure of 1t1f compared to the 1e04. The systems were neutralized and additional Na^+ and Cl^- ions were added to set the ionic strength to 0.15 M. A short simulated annealing period (2 ns) was applied for heating up the system to 310 K, then 2000 ns constant particle number(N), constant pressure ($p=10^5$ Pa), and constant temperature (T=310 K) production dynamics with periodic boundary (PBC) condition took place. For the water, the TIP3P explicit water model was applied. For the short-range electrostatic and van der Waals energy terms, 10 Å cut-off distances were used. The long-range electrostatic energy corrections were calculated by means of the particle mesh Ewald (PME) method. The v-rescale and Berendsen algorithms were used for temperature and pressure couplings, respectively. To have a chance to observe conformational fluctuation, which resembles the large-scale conformational transition that accompanies activation, the simulation time should be sufficiently long. It can be efficiently achieved using longer time steps during simulations. Virtual sites protocol is an alternative to mass repartitioning method which can be used to increase the length of time step in simulations. At this case, the position of H atoms are defined as a function of the neighboring heavy atoms. An increased time step (4 fs) was used in all the simulations presented in this work by means of the virtual site protocol applied for hydrogen atoms. The system

sizes (N), including the virtual sites, were 85,566, 88,291, and 88,165 for Systems 1, 2, and 3, respectively. The GROMACS software suites were used for all the simulations and for the analyses of trajectories. Protein structure visualizations were done by the VMD 1.9.1 software tools.

The "mutual information" or "generalized correlation" proposed by Lange and Grubmüller was calculated by means of the g_correlation utility. For this purpose, the frames from the trajectory were selected at the end of every nanosecond. In order to extract the principal components of the residue movements, principal component analysis (also known as essential dynamics analysis) calculations were carried out on the dynamic trajectories using the ProDy package, which is interfaced as the "normal mode wizard" to the VMD 1.9.1 molecular graphics software. Two thousand frames, saved at the end of every ns, were considered in Principal component analysis (PCA) calculations.

The covariance matrices were calculated only for the Cα atomic displacements. Only the 10 largest amplitude motions were calculated. The default parameters were applied in the normal mode wizard calculations to visualize the eigenvectors. Allosteric pathways were also derived from 2000 trajectory snapshots selected from the end of each nanosecond simulation using the WISP which was interfaced to the VMD 1.9.1 molecular graphics software as well. Only the Cα atoms of residues, as "nodes" were considered during covariance calculations. The allosteric pathways from Lys114 to Ser380 were calculated. The contact map limit of WISP was set to an average of 4.5 Å for the contact node pair generator. Only 20 allosteric pathways were generated. While there are many already-existing methods to map (to find local energy minima and barriers between them) the modestly complex potential energy hypersurface, in this work, we used mainly "brute force" NPT dynamics. The reason was that we were interested in the properties of "equilibrium" dynamics.

3. NEW RESULTS

3.1. Aspirin-COX-1 enzyme

The structure of human COX-1 was constructed from ovine COX-1 by homology modeling and ASA was docked into the substrate-binding cavity of the enzyme. Series of ONIOM-type QM/MM calculations applying electronic embedding approximations were carried out at different QM levels of theories on ASA and on the surrounding amino acids in order to propose a plausible reaction mechanism for the ASA-COX-1 trans-esterification reaction. Within the framework of the models that were constructed for the study, the exact transition state and local minimum geometries as well as the corresponding ZPVE-corrected energy values were determined. We found that, disregarding the ASA-COX-1 complex formation, this reaction is essentially a single elementary step reaction. The ASA COX-1 complex can be considered as the reactants. During the reaction first the ASA comes close to the active site of COX-1enzyme. The Arg120 in the proper orientation of ASA, while the Tyr385 in the ASA-COX-1 complex formation play crucial role.

In this complex the Ser530Oγ atom can be close enough to the carbonyl carbon of the ASA acetyl group for a nucleophilic attack, which results in the transition state structure. In this structure the formation of Oγ-C(ASA-carbonyl) and the O(ASA carboxylate)-Hγ bond are simultaneous processes with the cleavage of phenolic O-C (ASA-carbonyl) bond.

It should be noted that this does not mean that the bond cleavages and the bond formations occur in a fully synchronous process. Based on our calculations, the Hγ migration precedes the C (acetyl carbonyl carbon)-Oγ bond formation. This way the products (acetylated-COX-1: salicylic acid) were obtained without any further transition state, wherein the proton still located at the carboxylate group of salicylic acid.

The existence of transition state was proven by frequency analysis. The imaginary frequency dominated by Ser530Hγ proton transfer and Ser530Oγ-C (ASA

acetyl carbon) “bond stretching” modes, which also proves the one elementary step reaction for proton transfer and Oγ-C bond formation.

It was shown, that water molecules, which can be existed at the active site of COX-1, did not have remarkable effect either on the transition state energies or on the transition state geometries. This finding was essentially independent from the number of water molecules (1 or 2) at the active site and from the level of theory we applied.

Increasing the size of freely moving subsystem during geometry optimization did not influence significantly either the activation energy barrier or the calculated transition state geometries. The transition state energies obtained from density functional calculations using B3LYP method were significantly higher by ~3-4 kcal/mol than the corresponding values from the B97-D calculations in which an empirical dispersion correction term in the functional was introduced. On the other hand the latter two methods predicted very similar geometries for the transition state.

3.2. Antithrombin

In order to reveal the intrinsic dynamic properties and the reason of basal inhibitory activity of antithrombin 2 μs molecular dynamics simulations were carried out on its native free forms. It was found that correlated conformational transitions which can conduct the conformational information from the glycosaminoglycan binding site to the RCL are exists even without the presence of glycosaminoglycans. Computational evidence revealed that specific conformation for Arg393 side chain, suitable for serine protease binding, can exist even at the non-expulsed RCL conformation with non negligible probability. It explains the basal inhibitory potential of AT. From the trajectory analysis of the simulation of 1e04 stucture it is revealed that the peptide sequences correspond to the helix D extension and new helix P formation can be featured with especially large root mean square fluctuations. Mutual information analyses of the trajectory showed remarkable (generalized) correlation between those regions of antithrombin which changed their conformations as the

consequence of AT-PS complex formation. This suggests that allosteric information propagation pathways are present even in the non-activated native form of AT. Both principal component analysis and WISP analysis carried out on the trajectories lead to similar conclusion.

Meta-dynamics calculations were performed using the PLUMED and GROMACS software packages jointly in order to find a realistic minimum energy paths and transition states, which connect the two different types (1e04 and 1t1f) of RCL conformations have been revealed so far. From these simulations the conclusion could be drawn that such path (or paths) are exist between the two class of conformers. The free energy barrier which separates the two conformational classes are low enough to overcome it with considerable probability even by thermal motion.

4. DISCUSSION

4.1. Aspirin-COX-1 enzyme

4.1.1. Homology modeling and docking

The overall residue identity between ovine and human COX-1 is excellent (more than 90%) which very likely means a correct predicted structure. Especially high residue identity can be seen at- and near to the catalytic site of COX-1.

The most stable and most populated pose obtained by docking aspirin to the active center corresponded to a complex suitable for the next step of the reaction. In this pose the Ser530Oγ atom can be close enough to the carbonyl carbon of the ASA acetyl group for a nucleophilic attack, while the Ser530Hγ is close to the carboxylate oxygen, which could help the H atom to leave the Ser530Oγ atom. It should be mentioned that carrying out the same docking but with different docking parameters the most stable pose obtained by Tosco et. al. corresponded to an ion-pair between the ASA and the Arg120 residues. Both poses should have considerable statistical probabilities. The ion-pair formation can play a role in the pre-orientation of the ASA

in the substrate-binding channel while the pose we found is required for the trans-esterification reaction.

4.1.2. Potential energy surfaces

In the case of the simpler model (model A), the potential energy surfaces as a function of Ser530Oγ-C(ASA acetyl carbonyl) and Ser530 Hγ-Ser530Oγ distances suggested only a single saddle point at all ONIOM QM/MM calculations. The shapes of the potential energy surfaces slightly differ from each other. On the surface calculated by B3LYP/6-31G(d):AMBER method a relatively narrow valley lead to the saddle point. Neither HF/6-31G(d):AMBER nor MP2/6-31G(d):AMBER surface shows similar property. There is no sign of a distinct reaction step for the Hγ proton transfer from the Oγ atom to the ASA carboxylate and the nucleophilic attack of Oγ on the ASA carbonyl carbon atom.

4.1.3. Geometry parameters

Regarding the transition state geometry parameters we obtained it is interesting to observe, that the Ser530Hγ proton-ASA carboxylic O distances are close to a normal O-H bond length while the Ser530Hγ-Oγ distances are much longer. It seems as if the proton transfer would precede the O-C bond formation. Concerning product geometries noteworthy variations can be observed in the geometry parameters (Ser530 Hγ-Ser530Oγ and phenolic O-acetyl carbonyl C distances) associated with salicylic acid movement from the transition state. While at MP2 level the salicylic acid moves away moderately from the reaction center, its dislocation is remarkable at HF levels using all but the 6-31G basis. Using 6-31G(d) and 6-31G(d,p) basis sets it is closer to the position observed by X-ray experiments for the product geometry. B3LYP methods resulted in distances, which is in between those obtained by the applied two other methods.

4.1.4. Activation energies

We could not find any accurate experimental value for the activation energy in the literature. Tosco et. al. have demonstrated that QM/MM calculation applying SCC-DFTB QM method probably underestimates the activation energy. Using B3LYP/6-31G(d) correction to the SCC-DFTB potential energy surface they found ~10 kcal/mol activation free energy. Our calculations resulted in comparably higher ZPVE corrected activation energy values (~20-22 kcal/mol at density functional- and MP2 levels of theories). The reason for this discrepancy may be caused by the difference in the models used for computation. We intended to determine the exact transition state geometry at different levels of theories when only a subsystem of COX-1 was allowed to relax during geometry optimization. Tosco et. al. modeled the dynamical reaction path of the transesterification reaction applying a more approximate QM method in QM/MM calculations, which allowed the whole COX-1-ASA system to be included in dynamical simulation.

4.1.5. Imaginary vibrations

In every case only a single imaginary frequency was obtained for the transition state and all frequencies for the geometry of ASA-COX-1 and salicylic acid-acetylated COX-1 complexes were positive. The normal mode corresponding to the imaginary frequency shows that the proton migration between Ser530Oγ and ASA carboxylate O atoms and the bond breaking /formation between Ser530Oγ and carbonyl carbon of the ASA acetyl group occur in the same elementary step even if they are not fully synchronous events. Considering that in the transition state the Hγ proton migration to the ASA carboxyl group is almost completed, this is an interesting result.

4.1.6. Existence of the tetrahedral intermediate

Another question is the existence of the so-called tetrahedral intermediate. In this case a local minimum on the potential energy surface should exist which would correspond to an already formed Oγ-C(ASA-carbonyl) bond with a simultaneously existing phenolic O-C (ASA-carbonyl) bond. Interestingly our effort to find that intermediate failed. Optimizations started from geometries, which were taken from the product side but not too far from the saddle point resulted in finally the product geometry. Although it is not an exact proof for the non-existence of the tetrahedral intermediate, it is a good estimation that even if it would exist it would have a small energy barrier to prevent its bond splitting.

4.1.7. Position of the negative charge on the salicylic acid

We found that even after the completed trans-esterification reaction, the Ser530Hγ proton was bonded to the carboxylic oxygen of newly formed salicylic acid. Since in water at close to neutral pH the proton should be on the phenolic O instead of carboxylic one, the same was expected for the system we studied. Nevertheless, one has to consider, that the electrostatic field of proteins can considerably influence this situation. Therefore a relaxed potential energy scan moving the proton from the carboxylic oxygen to the phenolic oxygen was carried out and the potential energy curve was calculated at B3LYP/6-31G(d):AMBER level. The electrostatic field of protein was taken into consideration by the electronic embedding method, which was applied throughout this study. The proton prefers to stay near to the phenolic oxygen even if the energy gain is not especially large and the energy barrier supposed to prevent the movement of the proton from the carboxylic oxygen is very low, less than 0.5 kcal/mol. This barrier can be easily overcome by the thermal motion.

4.1.8. A possible effect of water molecules on the reaction

We carried out calculations on models in which water molecule(s) at the active site were included as well. In this case only B3LYP method with different basis sets was used as a high level theory in ONIOM computations. The calculations show that water molecule does not influence considerably the geometry parameters of the ASA-COX-1 non-bonding complex state nor the transition state energies and geometries.

4.2. Antithrombin

4.2.1. Molecular dynamics simulations

2μs dynamics simulations intervals are insufficient to observe native → latent transition or even activation of antithrombin since these are rare events on this time-scale. Nevertheless, other molecular dynamics properties e.g. loop conformational flexibilities, rms fluctuations, correlated motions between regions of the protein can be extracted from the trajectories. Based on these properties valuable information on the first phase of conformational transitions and on the pre-existence of allostery pathways can be expected.

4.2.2. RMSD

The root mean square deviations (RMSDs) from the starting geometries as a function of simulation time are calculated. Certain deviations from the X-ray based starting structures are quite usual since simulations are carried out in principle in solution phase. Nevertheless, most of the cases only a modest deviation can be observed. It is in agreement with the fact that the reference AT structures represent kinetically trapped local minima and the simulation do not disturb these structures too much. The first 45 N-terminal residues with the exception of the disulfide-bridged C8-C128 residues and their vicinity are badly- or even not resolved from the diffraction patterns, which clearly indicate that this part is flexible in both of the reference 1e04 and 1t1f X-ray structures. Remarkable flexibility is assumed also for

the RCL. Therefore omitting the highly flexible 45 N-terminal residues as well as the RCL region from the calculation substantially smaller RMSD values can be expected. The equilibrated values are now around ~0.2 nm which fall into the range of resolution of the X-ray structure.

4.2.3. RMSF

The root mean square fluctuation (RMSFs) values give us valuable information on the flexibilities of different regions of proteins. Calculating the RMSF values for the whole proteins it was found that the N-terminal regions of the ATs considered in these simulations can be featured by quite large rms fluctuations. It is not surprising since for two peptide sequences of these regions the X-ray diffraction patterns cannot be resolved and the completed structures expected to reflect this uncertainty. The remarkable fluctuations were found for the RCLs are not surprising as well since the basal inhibitory activity cannot be interpreted without conformational changes in this region.

It is interesting that some peripheral helices and even the strand built from 400-410 residues in structures without engineered V317C-T401Cys mutations also show remarkable fluctuations in the first rigid parts of AT. The considerably large fluctuations at the PS binding regions bordered by helices A, D and the newly formed P as well as the N-terminal region were not surprising. These regions should form optimal PS binding pocket and the reorganization of the otherwise repulsive, positively charged side chains presume remarkable flexibility at this site.

For the sequence between the C-terminal end of helix D and the subsequent 3rd β-strand (which is considered as the extension of helix D) larger rms fluctuation than most of the other loop regions were obtained from the dynamical trajectory. From the X-ray structures of free (i.e. no PS was bound) AT it can be characterized as an remarkably flexible loop (which sometimes cannot be resolved from the diffraction pattern) as well. The flexibility of the C-terminal part of helix F can be understood

considering that this region has to move during pentasaccharide binding and also when the RCL incorporates as 4th β string to the β-sheet-A.

4.2.4. RCL-structures

The dynamical property of the RCL even in its free native conformation has an exceptional importance in the structure-function studies of AT. In all the X-ray structures of free AT have been resolved so far the guanidinium group of Arg393 point to the inner side of AT. Other type of RCL conformation, where the side chain of Arg393 points outwards from the main body of AT (has been found so far only in the AT-proteinase complex) straightforwardly has to exist. This conformation either should appear in non-negligible population in the free AT or can be induced by AT-proteinase interaction.

It is assumed that the RCL conformation adopted in the 1e04 structure in which the guanidinium group is surface bonded is the consequence of crystal forces in the co-crystallized dimer (native+latent) AT structure. In the 1t1f structure the monomeric native AT without PS showed completely different RCL conformation in which the Arg393 side chain was also surface-bonded but by an Arg393-Glu237 salt bridge. Starting from the 1e04 conformation during the 2 μs dynamics simulation a large scale conformational change was observed in the RCL conformation. The distance of the guanidinium Cζ atom from any atom of the main body of AT as a function of time are calculated for all the systems we considered. For the 1e04 system the ~0.3 nm distance clearly correspond to a surface bound guanidinium group. The larger, 0.5-1.4 nm values correspond to conformations which are at least partially solvent exposed, i.e. which potentially can form Michaelis complex with proteases. Simulation carried out for the 1t1f wild type protein even larger flexibility can be seen from the trajectory for the Arg393 side chain position. Interestingly, when the C-terminal end of the RCL is fixed by a disulfide-bond, the flexibility of the RCL is not large enough for escaping the guanidinium group from a salt bridge formed with the Glu237 side chain.

4.2.5. Generalized correlation and allostery

Remarkable correlations can be seen between many residues located in 80-200 residue region in the Simulations 1 we carried out. Interestingly it is in agreement with the proposal that the motions of the whole helix D with the peptide sequence at its N-terminal end as well as the C-terminal ends of helices E and F show remarkable correlation. The considerable correlation of all of these peptide sequences with residues 412-420 as well as with the residues 305-310 peptides can be partially explained by the fact that these are in the proximity of each other. The (generalized) correlations are shown between the corresponding part of helices D-, E- and F with the RCL as well as with residues 340-350 less straightforward. It should be mentioned that there is no considerable correlation between the movement of hinge region and any part of the peptide sequences mentioned above. It also suggests that the hinge region expulsion is occurs in a distinct step. Essentially the same information with some remarkable differences can be seen from the generalized correlation matrices depicted for the Simulations 2 and 3. Both of them show the same regional separation of correlated motions, which was observed at the case of Simulation 1. Noteworthy that the corresponding simulations starting from 1t1f structure show larger weakly correlated regions. It is especially true for the system where there was no extra disulfide bond introduced.

4.2.6. Weighted implementation of suboptimal path (WISP)

Very recently a new method to extract information on allostery from molecular dynamics trajectory has been proposed. Using this for the trajectories we obtained, the allosteric pathways from the AT-PS interaction site to the site of action, in our cases the hinge region, were proposed for the free AT structures. Surprisingly, three different pathways were proposed. While the Simulation 1 showed a signal path which propagated along the D-helix, the Simulation 2 proposed an inner path in which the N-terminal end of A helix is involved. It can be also rationalized since the Trp49

was found to play important role in stabilizing activated state. Simulation 3 predicted a preferable pathway through the P and E helices and the 2nd and 3rd strand of β-sheet-A. While the path obtained from Simulation 1 seems to be the most obvious, even the latter ones cannot be excluded in advance. It should be noted, however, that in AT-proteinase complexes the RCL conformation is more similar to those one which was observed in Simulation 1, it has considerably probability that the 1t1f→salt-bridged 1e04-like conformation transition precedes the activation.

4.2.7. Principal component analysis (PCA)

Global internal motions (removing the translation and rotation degrees of freedom) can be decomposed principal components. These principal components can be regarded as large amplitude (low frequency) modes and can be assumed that conformational transitions begin along these modes. These modes the contribution of those regions are dominant which have the largest fluctuations (N-terminal region, pentasaccharide binding region, E,F, helices as well as the RCL) demonstrating coupled motions to each other. While it is not an exact proof of the direct role of these motions in propagation of allosteric information (linear combinations of these motions can be used as well and even the contribution of these motions to the trajectory snapshots can be different from time to time), it helps to explain allostery.

5. SUMMARY

The normal- and pathological processes of hemostatic system are long-established scientific topics in our department. In this dissertation a report on the scientific achievements I have reached so far concerning on the ASA COX-1 interaction and the dynamical properties of AT will be given.

Acetylsalicylic acid (aspirin) suppresses the generation of prostaglandin H2, which is the precursor of thromboxane A2. Aspirin acts as an acetylating agent. Its acetyl group is covalently attached to a serine residue (Ser530) in the active site of the cyclooxygenase-1 enzyme. The exact reaction mechanism has not been revealed by experimental methods. In this study the putative structure of human cyclooxygenase-1 was constructed from ovine cyclooxygenase-1 by homology modeling, and the acetylsalicylic acid was docked into the arachidonic acid binding cavity of the enzyme.

To characterize the shape of the potential energy surface of the acetylating reaction and to determine the relative energies of the stationary points on the surface, a series of ONIOM-type quantum mechanical/molecular mechanical calculations were carried out at different quantum mechanical levels of theories applying electronic embedding approximations. The acetylsalicylic acid and the surrounding amino acids were included in these calculations. Frequency analyses were performed to prove the existence of first order saddle points (representing transition states) and local minima on the potential energy surface.

It was found that all levels of theories predicted similar transition state geometries. The activation energy values, however, demonstrated significant dependence on the methods that were applied. All the applied “dependable” ab initio and density functional theory methods predicted that the breakage of the Ser530 Oγ-Hγ and formation of the Oγ-C (acetylsalicylic acid carbonyl) bonds occur in a single elementary step.

Antithrombin (AT) is one of the most important physiological inhibitor of the blood coagulation cascade. AT belongs to the SERPIN (SERine Protease INhibitor)

family of proteins with a special folding and inhibits thrombin and other serine protease coagulation factors, e.g. FXa, FXIa and FIXa. The common feature of SERPINs is the relatively long reactive center loop (RCL) which is functioning as "bait" for serine proteases and which is partially incorporated in β-sheet-A at the case of AT. Binding of highly negatively charged glycosaminoglycans expulses the RCL from β-sheet-A by an allosteric mechanism which resulted in dramatically increased inhibitory potential of AT against FXa.

Our goal was to understand how the conformational change at one site of the AT resulted in a specific conformational change at the distant site of the protein. In other words, how the conformational information from one site of the protein flows to the other one. To understand the small basic inhibitory potential of AT, which cannot be explained by the available X-ray structures, was also aimed. To these end, a relatively long molecular dynamics simulation on the AT has been carried out mimicking the real physiological conditions.

It was found that correlated conformational transitions which can conduct the conformational information from the glycosaminoglycan binding site to the RCL are exists even without the presence of glycosaminoglycans. Computational evidence revealed that specific conformation for Arg393 side chain, suitable for serine protease binding, can exist even at the non-expulsed RCL conformation with non negligible probability. It explains the basal inhibitory potential of AT.

6. LIST OF PUBLICATIONS

6.1. Papers

UNIVERSITY OF DEBRECEN
UNIVERSITY AND NATIONAL LIBRARY
PUBLICATIONS

Register number: DEENK/16/2015. PL
Item number:
Subject: Ph.D. List of Publications

Candidate: László Tóth
Neptun ID: MH276T
Doctoral School: Kálmán Laki Doctoral School

List of publications related to the dissertation

1. **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. Epub ahead of print (2014)
   DOI: http://dx.doi.org/10.1080/07391102.2014.986525
   IF:2.983 (2013)

2. **Tóth, L.**, Muszbek, L., Komáromi, I.: Mechanism of the irreversible inhibition of human cyclooxygenase-1 by aspirin as predicted by QM/MM calculations.
   J. Mol. Graph. 40, 99-109, 2013.
   DOI: http://dx.doi.org/10.1016/j.jmgm.2012.12.013
   IF:2.022

H- 4032 Debrecen, Egyetem tér 1. ▫ E-mail publikaciok@lib.unideb.hu

List of other publications

3. Owen, M.C., **Tóth, L.**, Jójárt, B., Komáromi, I., Csizmadia, I.G., Viskolcz, B.: The Effect of Newly Developed OPLS-AA Alanyl Radical Parameters on Peptide Secondary Structure.
J. Chem. Theory Comput. *8* (8), 2569-2580, 2012.
DOI: http://dx.doi.org/10.1021/ct300059f
IF:5.389

Total IF of journals (all publications): 10,394
Total IF of journals (publications related to the dissertation): 5,005

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

26 January, 2015

6.2. Oral presentations (O) and posters (P) related to the dissertation

- L. Tóth, I. Komáromi, L. Muszbek: A humán ciklooxigenáz enzim gátlása Aspirinnel, insilico kísérletek, Laki Kálmán Emlékülés, MTHT, September 26.,2009, Abádszalók, Hungary (O) (In Hungarian)
- L. Tóth, I. Komáromi, L. Muszbek: A humán ciklooxigenáz-1 enzim aspirinnel történő gátlásának mechanizmusa. Insilico kísérletek. QSAR symposium, April 29-30., 2010, Szeged, Hungary (O) (In Hungarian)
- L. Tóth, A. Fekete, I. Komáromi: Az antitrombin aktivációja és konformációs átmenetei molekuladinamikai tanulmányok alapján, Tavaszi szél 2014 konferencia, March 21-23, 2014, Debrecen (O) (In Hungarian)
- L. Tóth, A. Fekete, I. Komáromi: Az antitrombin aktivációja és konformációs átmenetei molekuladinamikai tanulmányok alapján, I. Innováció a természettudományokban 2014 konferencia, May 2-3, 2014, Szeged, Hungary (O) (In Hungarian)
- L. Tóth, A. Fekete, I. Komáromi: A humán ciklooxigenáz-1 enzim aspirinnel történő gátlásának mechanizmusa. Insilico kísérletek. QSAR symposium, May 22-23., 2014, Szeged, Hungary (O) (In Hungarian)
- L. Tóth, I. Komáromi, L. Muszbek: Mechanism of their reversible inhibition of human cyclooxygenase-1 by aspirin as QM/MM predict, 6th International Conference of Structure and Stability of Biomacromolecules, September 9-11., 2009, Kosice, Slovakia (P)
- L. Tóth, L. Muszbek, I. Komáromi: Mechanism of their reversible inhibition of human cyclooxygenase-1 by aspirin as quantum mechanical / molecular mechanical calculations predict, 21st International Thrombosis Congress, July6-9., 2010, Milan, Italy (P)
- L. Tóth, L. Muszbek, I. Komáromi: How Aspirin Inhibits Human Cyclooxygenase-1? Hybrid Quantum Mechanical / Molecular Mechanical (QM/MM) Calculations on the Mechanism, XXIV. Congress of the

international society on thrombosis and haemostasis, June 29-July 3., 2013, Amsterdam, The Netherlands (P)

- L. Tóth, L. Muszbek, I. Komáromi: Mechanism of their reversible inhibition of human cyclooxygenase-1 by aspirin as QM/MM predict, 8th International Conference of Structure and Stability of Biomacromolecules, September 9-11., 2013, Kosice, Slovakia (P)
- L. Tóth, L. Muszbek, I. Komáromi: Mechanism of their reversible inhibition of human cyclooxygenase-1 by aspirin as predicted by QM/MM, I. Innováció a természettudományokban 2014 konferencia, May 2-3., 2014, Szeged, Hungary (P)

This work was supported by the following funds and fellowships: OTKA K-106294, K-105459, PD101120, NIIF-1057, TÁMOP-4.2.2.A-11/1/KONV-2012-0045, TÁMOP 4.2.4.A/2-11-1-2012-0001, TAMOP-4.2.2.C-11/1/KONV-2012-0010.1, NKFP-07-A1-2008-0127), NIIF 1057 and MTA11003.