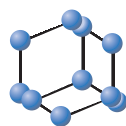


RESEARCH ARTICLE

BENTHAM
SCIENCE

A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin



Xiaohua Zhang¹, Horacio Perez-Sanchez^{2,*} and Felice C. Lightstone^{1,*}

¹Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, USA; ²Bioinformatics and High Performance Computing Research Group, Department of Computer Science, Universidad Católica San Antonio de Murcia (UCAM), Murcia Spain

Abstract: Background: A high-throughput virtual screening pipeline has been extended from single energetically minimized structure Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) rescoring to ensemble-average MM/GBSA rescoring. The correlation coefficient (R^2) of calculated and experimental binding free energies for a series of antithrombin ligands has been improved from 0.36 to 0.69 when switching from the single-structure MM/GBSA rescoring to ensemble-average one. The electrostatic interactions in both solute and solvent are identified to play an important role in determining the binding free energy after the decomposition of the calculated binding free energy. The increasing negative charge of the compounds provides a more favorable electrostatic energy change but creates a higher penalty for the solvation free energy. Such a penalty is compensated by the electrostatic energy change, which results in a better binding affinity. A highly hydrophobic ligand is determined by the docking program to be a non-specific binder.

Results: Our results have demonstrated that it is very important to keep a few top poses for rescoring, if the binding is non-specific or the binding mode is not well determined by the docking calculation.

Keywords: Docking, MM/GBSA, Rescoring, VinaLC, BINDSURF, Binding Affinity, Molecular Dynamics, Antithrombin.

1. INTRODUCTION

High-throughput virtual screening is one of the most commonly used techniques in the drug discovery procedure. A high throughput virtual screening pipeline has been built to leverage the high performance computing (HPC) resource at Lawrence Livermore National Laboratory for *in-silico* screening of large virtual compound databases [1]. The pipeline consists of five modules: receptor/target preparation, ligand preparation, VinaLC docking calculation [2], single-structure MM/GBSA rescoring, and ensemble-average MM/GBSA rescoring [3]. The pipeline uses a down-selecting scheme to filter large numbers of compounds through the docking, single-structure rescoring, and ensemble-average rescoring. One of the notable features of this pipeline is an automated receptor preparation scheme with unsupervised binding site identification, which enables automatically running the whole pipeline with little or no human intervention. Perez-Sanchez and co-workers have developed a similar approach to improve drug discovery using massively parallel GPU hardware instead of

supercomputers [4]. Their GPU-based program, BINDSURF [5], takes advantage of massively parallel and high arithmetic intensity of GPUs to speed-up the calculation in low cost desktop machine.

In molecular docking, scoring functions are employed to evaluate the docking poses generated by conformational searching applications. The scoring functions simplify the calculation with many approximations to achieve high throughput screening. In the previous study, we have demonstrated the parallel VinaLC docking program can screen 1 million compounds in less than 1.5 hours by using 15K CPUs [2]. Docking programs sacrifice accuracy for speed to make the assessments of a large number of compounds feasible. Even though the ligand binding conformation can be identified in most cases by using various existing docking programs, there are no universal scoring functions for all types of molecules and protein families [6]. Rescoring steps are necessary for the accurate prediction of binding energies.

By using approximated scoring functions, molecular docking usually fails to estimate binding energies when compared to experimental values [7]. The MM/GBSA method is selected for rescoring because it is the fastest force-field based method that computes the free energy of binding, as compared to the other computational free energy methods, such as free energy perturbation (FEP) or thermodynamic integration (TI) methods [8]. Comparison studies have also shown that MM/GBSA outperforms Molecular

ARTICLE HISTORY

Received: May 05, 2016
Revised: June 12, 2016
Accepted: June 19, 2016

DOI: 10.2174/15680266166661611171
12604

*Address correspondence to these authors at the Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory (LLNL), Livermore 94550, CA, USA; Tel/Fax: +1-925-423-8657, +1-925-423-0785; E-mail: lightstone1@llnl.gov

Bioinformatics and High Performance Computing Research Group, Department of Computer Science, Universidad Católica San Antonio de Murcia (UCAM), Spain; E-mail: hperez@ucam.edu

Mechanics/Poisson Boltzmann Surface Area MM/PBSA [9]. The MM/GBSA method has been widely exploited in free energy calculations [10, 11], and MM/GBSA rescoring overly yields better results than docking for the DUD-E (Directory of Useful Decoys, Enhanced) data set [1, 12].

There is evidence showing that molecular dynamics (MD) simulations are necessary for some systems to identify the correct binding conformations [9]. In this study, we have extended our pipeline from single-structure to ensemble-average MM/GBSA rescoring by carrying out long-time MD simulations. To validate the new approach, we have gathered a panel of 8 antithrombin ligands (Fig. 1), including heparin and non-polysaccharide scaffold compounds. For the purpose of comparison, both single-structure and ensemble-average MM/GBSA rescoring are employed in the binding affinity calculations of antithrombin ligands. We must point out that estimation/calculations of the entropy term are tricky. In most scenarios, the entropy term is neglected in the calculation for relative free binding energies. Quite a few researchers dispute the benefits of including the entropy term, which can be a major source of error due to the drawback of the entropy calculation method [13, 14], despite others who advocate its usage [15]. We choose to neglect the partial entropy effects estimated by the normal mode analysis in our calculations.

Antithrombin is a glycoprotein that plays a crucial role in the regulation of blood coagulation by inactivating several enzymes of the coagulation system and, thus, is an important drug target for the anticoagulant treatment. Antithrombin has two major isoforms, α and β , in the blood circulation [16]. α -Antithrombin is the dominant form of antithrombin and consists of 432 amino acids with 4 glycosylation sites, where an oligosaccharide occupies each glycosylation site [17]. Heparin is the first compound that was identified and used as an anticoagulant and antithrombotic agent. It is a sulfated polysaccharide containing a specific pentasaccharide fragment (Fig. 1, NTP) that binds and activates the antithrombin [18]. This binding localizes the function of antithrombin to inhibition of serine proteases in the coagulation cascade in the bloodstream, which allows coagulant activity in damaged tissue outside the vascular system [17].

Due to increasing interests in clinical applications, computational studies have been carried out to investigate the structure and behavior of antithrombin. Verli and co-workers performed molecular dynamics simulations to study the induced-fit mechanism of the antithrombin-heparin interaction and effects of glycosylation on heparin binding [19, 20]. Several detailed conformational changes associated with heparin binding to antithrombin were revealed. They also confirmed an intermediate state between the native and activated forms of antithrombin. Because of the weak surface complementarity and the high charge density of the sulfated sugar chain, the docking of heparin to its protein partners presents a challenging task for computational docking. Wade and Bitomsky developed a protocol that can predict the heparin binding site correctly [21]. Navarro-Fernandez and colleagues screened a large database *in silico* by using FlexScreen docking program [22] and identified a new, non-polysaccharide scaffold able to interact with the heparin binding domain of antithrombin [23]. They predicted D-

myo-inositol 3,4,5,6-tetrakisphosphate (Fig. 1, L1C4) to strongly interact with antithrombin, which was confirmed by experimental binding affinity study.

Here, we carried out molecular dynamics (MD) simulations of ligand recognition upon binding antithrombin and calculate the ligand binding affinity. According to the characteristics of the compounds, we divide the study into two categories. The first category is devoted to the ligands that specifically binding in the active sites of the antithrombin. Compound L1C1, L1C2, L1C3, L1C4, L1C5, L1C6, and NTP belong to this category. The binding affinity of the top pose of each ligand-complex is computed by using the ensemble-average MM/GBSA rescoring method, as a complementary study to the previous docking work [21, 23]. The advantage of long-time MD simulations is that more configurational space can be explored and dynamical properties of systems can be revealed. Seven long-time MD simulations are performed for this category. The second category contains only one ligand, named Penfluridol. As shown in the following section, Penfluridol is a non-specific binder as determined by the docking program. We selected top 10 poses and three additional poses for ensemble-average MM/GBSA rescoring. Thus, thirteen long-time MD simulations are performed for Penfluridol.

2. METHODS

In this study, a total of 20 MD simulations are performed. The initial structure for MD simulation of the antithrombin complex with Compound NTP is obtained from the PDB bank (PDB ID: 1AZX). The initial structures of the antithrombin complex with Compound L1C1, L1C2, L1C3, L1C4, L1C5, and L1C6 (Fig. 1) are obtained from the FlexScreen docking program [22]. The initial structures of the antithrombin complex with Penfluridol are the top 10 poses from the docking calculation and three additional poses, where they are close to the active site of antithrombin.

The MM/GBSA calculations are applied to these initial structures by using our in-house developed pipeline [1, 2] and Amber molecular simulation package [24]. The Amber forcefield f99SB [24] is employed in the calculation for the antithrombin receptor. Ligands use the Amber GAFF forcefield [25] as determined by the Antechamber program [26] in the Amber package. Partial charges of ligands are calculated using the AM1-BCC method [27]. The fourth module of the pipeline is employed for the single-structure MM/GBSA calculation, where the receptor-ligand complexes are energetically minimized by the MM/GBSA method as implemented in the Sander program of the Amber package [28]. The atomic radii developed by Onufriev and coworkers (Amber input parameter $igb=5$) are chosen for all GB calculations [29]. For the ensemble-average MM/GBSA rescoring, a MD trajectory is obtained for each ligand-receptor complex. The snapshots of complex are extracted from the MD trajectory at every 10 ps, which yields 10,000 snapshots. The various free energy terms of complexes, receptors, and ligands derived from snapshots are calculated by the MM/GBSA methods. The final binding free energy is an average of the binding free energies of snapshot. The MD simulation for each ligand-receptor complex starts with the energetically minimized structure from single-structure

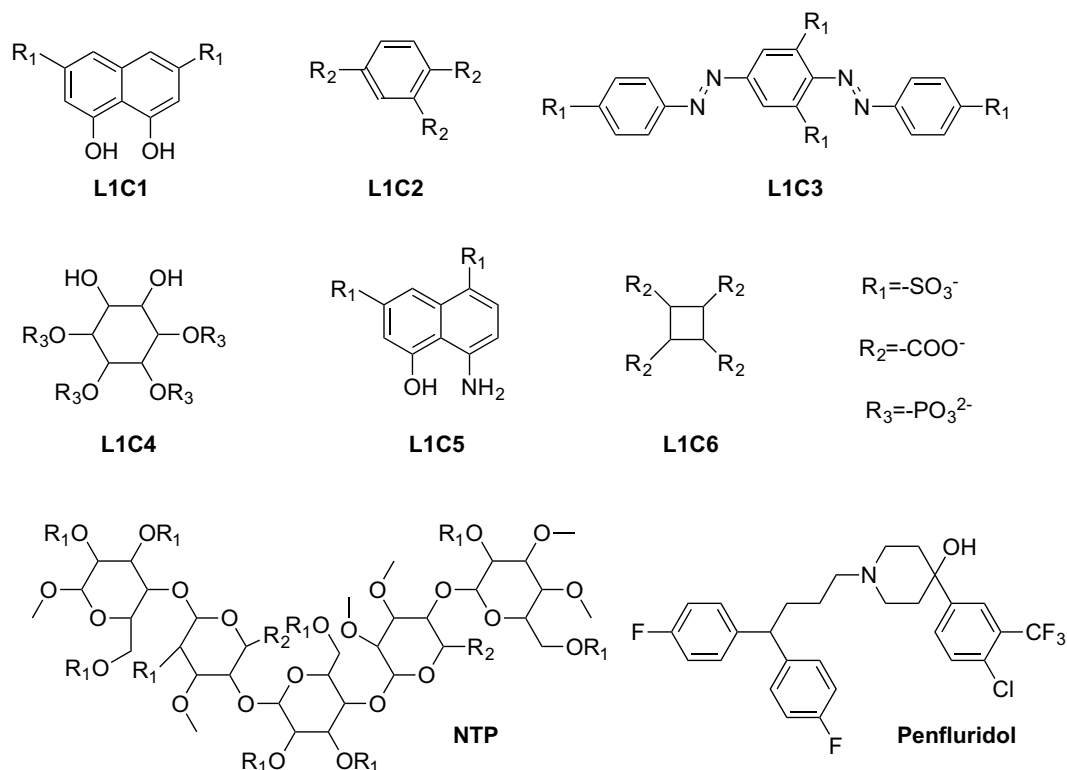


Fig. (1). Compounds targeting antithrombin. Compound NTP is a synthetic pentasaccharide compound from the crystal structure (PDB ID: 1AZX).

MM/GBSA rescoring. The systems are heated from 0 K to room temperature, 300 K. The MD simulations with a time step of 2 fs for the integration of the equations of motion are carried out at room temperature. The systems are equilibrated at room temperature for 500 ps. Each MD trajectory is followed to 100 ns after equilibrium. Binding affinities of antithrombin and its 8 ligands are calculated by post-processing the ensembles of structures extracted from MD trajectories using MM/GBSA calculations. In the MM/GBSA calculation, the binding free energy between a receptor and a ligand is calculated using the following equations:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \approx \Delta E_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S \quad (2)$$

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{ELE}} + \Delta E_{\text{VDW}} \quad (3)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{GB}} + \Delta G_{\text{Surf}} \quad (4)$$

The binding free energy (ΔG_{bind}) is decomposed into different energy terms. Because the structures of complex, receptor, and ligand are extracted from the same trajectory, the internal energy change (ΔE_{int}) is canceled. Thus, the gas-phase interaction energy (ΔE_{gas}) between the receptor and the ligand is the sum of electrostatic (ΔE_{ELE}) and van der Waals (ΔE_{VDW}) interaction energies. The solvation free energy (ΔG_{sol}) is divided into the polar and non-polar energy terms. The polar solvation energy (ΔG_{GB}) is calculated by using the GB model. The non-polar contribution is calculated based on the solvent-accessible surface area (ΔG_{Surf}). A value of 80 is used for the solvent dielectric constant, and the

solute dielectric constant is set to 1. The calculated binding free energy (ΔG_{bind}) is the sum of the gas-phase interaction energy and solvation free energy because we neglect the entropy term. The experimental binding free energy is estimated from the experimental dissociation constant (Kd) by the equation:

$$\Delta G_{\text{Exp}} = -RT \cdot \ln(Kd) \quad (5)$$

where R is the gas constant, and T is temperature.

To study the docking and rescoring, we choose a hydrophobic compound (Penfluridol) to carry out the docking and the ensemble-average rescoring calculations. To compare docking and rescoring, we utilize the BINDSURF program to explore all possible binding sites on the surface of antithrombin and dock Penfluridol to the predicted binding sites by using the FlexScreen docking program. We use BINDSURF instead of FlexScreen because FlexScreen is designed to take into account the structure of the ligand, and the results obtained are similar (data not shown). The docking results have not shown much difference in docking scores for the top 50 poses (see Results and Discussion). In the previous study, the top 20 docking poses were kept for the MM/GBSA single-structure rescoring [1]. Due to the intensive computational resource requirement in the ensemble-average MM/GBSA rescoring calculations, here we keep only the top 10 docking poses for ensemble-average MM/GBSA rescoring of one ligand binding study. In the top 10 docking poses, there is only one pose that is docked into the antithrombin active site. Because the poses near the active site of antithrombin are our primary interests, we include

Table 1. Calculated and experimental binding free energies (kcal/mol) of antithrombin ligands.

Cmpd	ΔE_{ELE}	ΔE_{VDW}	ΔE_{gas}	ΔG_{Surf}	ΔG_{GB}	ΔG_{GB-ELE}	ΔG_{Sol}	ΔG_{bind}	Kd(μ M)	ΔG_{Exp}
L1C1	-552.67 \pm 33.60	-23.68 \pm 5.10	-576.35 \pm 32.43	-2.75 \pm 0.09	480.12 \pm 20.19	-72.55 \pm 18.72	477.37 \pm 20.22	-98.97 \pm 17.90	-	-
L1C2	-442.99 \pm 39.57	0.47 \pm 3.73	-442.52 \pm 38.98	-1.20 \pm 0.17	417.60 \pm 38.11	-25.39 \pm 6.40	416.41 \pm 38.09	-26.11 \pm 4.28	13700	-2.54
L1C3	-836.77 \pm 43.49	-39.96 \pm 5.12	-876.73 \pm 43.67	-4.06 \pm 0.26	781.94 \pm 38.79	-54.83 \pm 11.15	777.88 \pm 38.68	-98.85 \pm 11.91	10.02	-6.81
L1C4	-1599.09 \pm 166.95	33.02 \pm 14.06	-1566.07 \pm 156.65	-2.98 \pm 0.15	1261.05 \pm 93.52	-338.04 \pm 78.16	1258.07 \pm 93.43	-308.01 \pm 66.94	0.088	-9.62
L1C5	-613.30 \pm 36.20	-19.00 \pm 4.52	-632.31 \pm 36.87	-2.57 \pm 0.15	525.82 \pm 19.51	-87.48 \pm 21.16	523.25 \pm 19.46	-109.06 \pm 22.04	0.69	-8.40
L1C6	-818.73 \pm 121.90	8.21 \pm 5.04	-810.52 \pm 123.16	-1.54 \pm 0.15	752.94 \pm 127.91	-65.79 \pm 11.97	751.41 \pm 127.85	-59.11 \pm 9.01	17.52	-6.48
NTP	-2598.87 \pm 351.39	-60.89 \pm 7.55	-2659.76 \pm 351.76	-7.58 \pm 0.61	2387.77 \pm 301.40	-211.09 \pm 52.56	2380.20 \pm 300.83	-279.57 \pm 53.29	0.104	-9.52

three additional poses near to the antithrombin active site based on the docking scores. Although these three poses are not in the top 10 list, the docking scores for them are very similar to those in the top 10 list.

3. RESULTS AND DISCUSSION

The calculated binding free energies of Compound L1C1, L1C2, L1C3, L1C4, L1C5, L1C6, and NTP using the ensemble-average MM/GBSA rescoring are shown in Table 1 together with their corresponding experimental values. Each calculated binding free energy is averaged from snapshots extracted from 100 ns MD trajectories. Except for Compound L1C1, all the antithrombin ligands have experimental binding free energies. As determined experimentally, Compound L1C4 is the best binder with a Kd value of 0.088 μ M [23]. As predicted by the MM/GBSA method, Compound L1C4 has the most negative binding free energy (-308.01 kcal/mol), which is in agreement with the experimental results. The second best binder as predicted by the MM/GBSA calculation is Compound NTP with a calculated binding free energy of -279.57 kcal/mol, confirming the experimental ranking relative to Compound L1C4. Compound L1C2 is predicted to have the worst binding free energy of the six ligands, which is also in agreement with its experimental ranking value. In summary, the MM/GBSA calculations rank the binding affinities of all six antithrombin ligands in the same exact order as that of experimental binding free energy rankings. The calculated binding free energies of six antithrombin ligands using the ensemble-average MM/GBSA rescoring have been plotted against the free energies derived from experimental dissociation constants. The correlation coefficient (R^2) is 0.69, which indicates good correlation between the calculated and experimental values (Fig. 2). Obviously, L1C2 is an outlier. The value of R^2 increases to 0.85 after removing the outlier. By comparison, the correlation coefficient calculated by single-structure MM/GBSA is only 0.36, and Compound NTP is predicted to be the best binder, instead of Compound L1C4. Thus, using the ensemble-average MM/GBSA rescoring method significantly improves the accuracy of the prediction over the single-structure MM/GBSA rescoring.

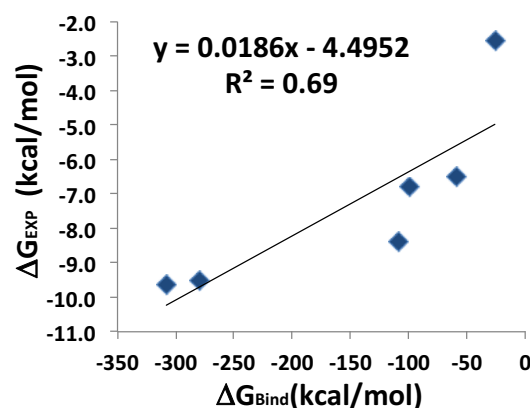


Fig. (2). The scatter plot of calculated MM/GBSA binding free energy versus experimental binding affinity estimated from dissociation constant.

As shown in Fig. (1), Compound L1C1, L1C2, L1C3, L1C4, L1C5, L1C6, and NTP contain negatively charged groups, suggesting electrostatic interactions should be a key factor in the binding affinity. Compound NTP has a total charge of -11, and Compound L1C4 has a total charge of -8. By decomposing the binding free energy, Compound NTP and L1C4 have the largest electrostatic energy changes upon binding in both gas phase (ΔE_{ELE}) and GB solvent (ΔG_{GB-ELE}). The energy change upon binding in gas phase is equivalent to the energy change upon binding for the solute. Thus, in other words, Compound NTP and L1C4 have the largest electrostatic energy changes upon binding in solute and solvent. In contrast, Compound L1C2 has the smallest electrostatic energy changes in solute and solvent. Although Compound L1C4 has the least favorable van der Waals energy change upon binding, the electrostatic energy change compensates significantly. For all ligands, the van der Waals energy changes (ΔE_{VDW}) upon binding are less than the electrostatic energy changes (ΔE_{ELE}) by 1-2 orders of magnitude. The contribution of the van der Waals energy change has been overpowered by the electrostatic energy change. Non-polar contribution of solvation free energy of Compound NTP and L1C3 are more negative than that of the other compounds because the sizes of Compound NTP and L1C3 are larger than the other compounds. Nevertheless, non-polar

contributions for all compounds are small. The non-polar contribution is overwhelmed by the polar contribution of solvation free energy. Thus, the two major factors to determine the binding affinity are the electrostatic energy change and solvation free energy change. The larger the total charge of the compound, the larger the penalty cost is for solvation free energy. However, the high penalty for large total charge of compound has already been paid by the favorable large changes in electrostatic energy. Although, the electrostatic energy change of Compound L1C4 is less than that of Compound NTP, Compound L1C4 needs less compensation for the solvation free energy. Thus, Compound L1C4 is a better binder than Compound NTP.

Hydrogen bonding analysis determines the numbers of hydrogen bonds to antithrombin that are persistent at >20% of the time. Compound NTP has 40 hydrogen bonds to antithrombin, while L1C4 has 25 hydrogen bonds. For Compounds L1C1, L1C2, L1C3, L1C5, and L1C6, the numbers of hydrogen bonds are 5, 10, 12, 12, and 12, respectively. Taking the molecular weight into account and using a similar approach as Reynolds' ligand efficiency method [30], Compound L1C4 has the highest ligand efficiency.

Compound L1C4 forms double hydrogen bonds with Arg47 (Fig. 3A). One hydrogen bond (O6-HH21-NH2) is 94.81% persistent, and the other one (O6-HE-NE) is 89.55% persistent. The average hydrogen bond distances between the heavy atoms are 2.74 Å and 2.70 Å, respectively. Compound L1C4 has strong hydrogen bonds with Arg47, and one of the four phosphate groups from Compound L1C4 is locked to the Arg47. According to the hydrogen bonding analysis, Compound L1C4 is also hydrogen bonded to Arg46, Arg13, Lys114, Lys11, Lys125, and Asn45, which are key residues to the binding process. Except for the phosphate group locked to Arg47, the other three phosphate groups of Compound L1C4 can rotate so that key residues can form hydrogen bonds to different oxygen atoms of phosphate at different times during the MD trajectory. Notably, Arg13 starts far away from Compound L1C4 in the initial conformation. After 8 ns of MD simulation, Arg13 begins to make hydrogen bonds with the phosphate group of Compound L1C4, which suggests that long-time MD simulations are essential to obtaining accurate binding affinities. As shown in Fig. (3B), Compound NTP makes hydrogen bonds to antithrombin mainly via its negatively charged sulfate groups. Compound NTP forms high persistent hydrogen bonds with Arg13, Arg129, Arg47, and Asn45 (70-88%) and forms medium persistent hydrogen bonds with Arg132, Lys125, and Thr44 (43-66%). Only relatively weak hydrogen bonds are observed with Arg46, Lys114 and Lys11.

To compare the docking and rescoring procedures of previous and current studies, Penfluridol has been docked to antithrombin using BINDSURF. The distribution of docking scores shows little energetic difference between the docking poses (Fig. 4). The top 10 Penfluridol docking poses and 3 additional poses near the active site of antithrombin are selected for the MM/GBSA rescoring (Fig. 5). The Penfluridol (Fig. 1) compound consists of aromatic rings and alkane groups, which suggests that it is hydrophobic. Thus, the docking sites, as determined by the BINDSURF program, are close to hydrophobic surfaces of antithrombin, as shown

in the Fig. (5). The active site of antithrombin is highly charged with many positively charged residues, such as arginine and lysine. In the end, most poses have been docked outside of the binding active site of antithrombin. The binding of Penfluridol to antithrombin is non-specific.

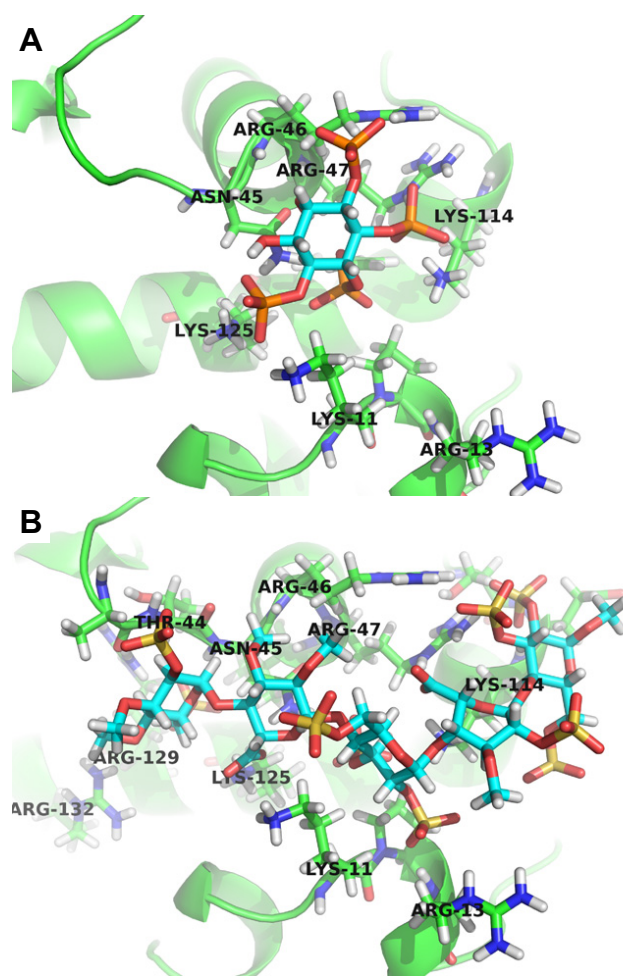


Fig. (3). Initial structures of Compounds L1C4 (A) and NTP (B) complexed with antithrombin.

The ensemble-average MM/GBSA rescoring has been performed for the selected 13 docking poses. The calculated binding free energies are shown in Table 2. The binding free energies for these selected poses range from -28.75 to -59.68 kcal/mol, which are much smaller than those of good antithrombin binders, e.g. Compound L1C4 and NTP. As expected, MM/GBSA rescoring ranks the poses differently than the docking method because the docking program uses an empirical scoring function as more like "machine learning" than direct physics-based in nature. The scoring function does not contain any exact terms for hydrophobic, VDW, or solvation components. In contrast, MM/GBSA is a physics-based method, which contains explicit terms for hydrophobic, VDW, or solvation components. Thus, the ranks of the poses from two methods could differ significantly. The MM/GBSA method is considered to be more accurate at the cost of computer time. The top docking ranked poses are the lowest ranked poses using MM/GBSA rescoring, indicating the very importance of keeping a few

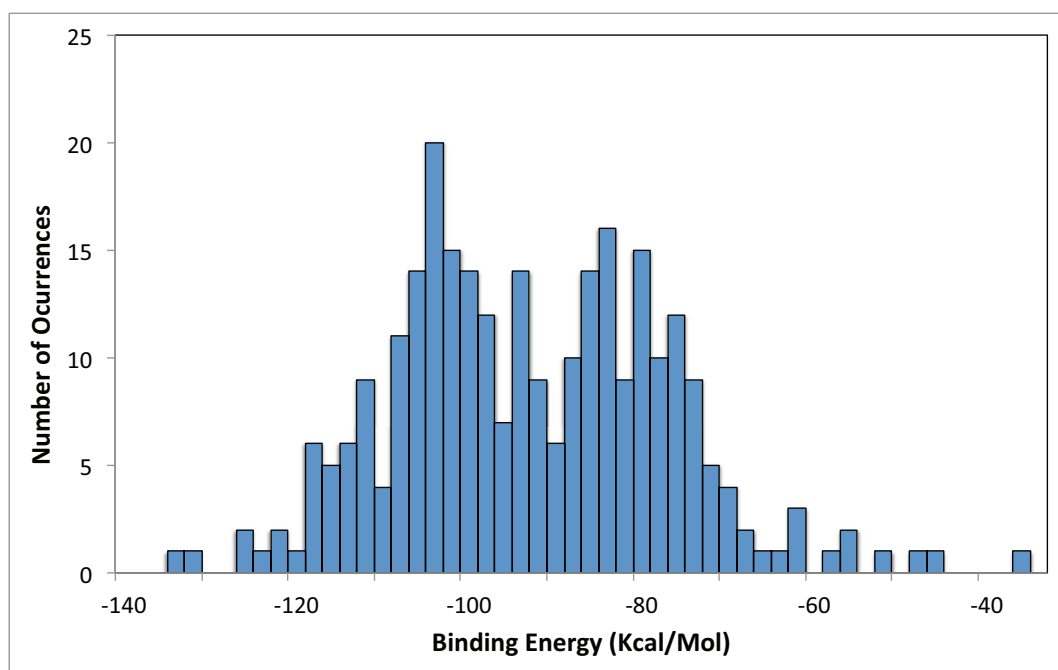


Fig. (4). Distribution of the docking scores for Penfluridol.

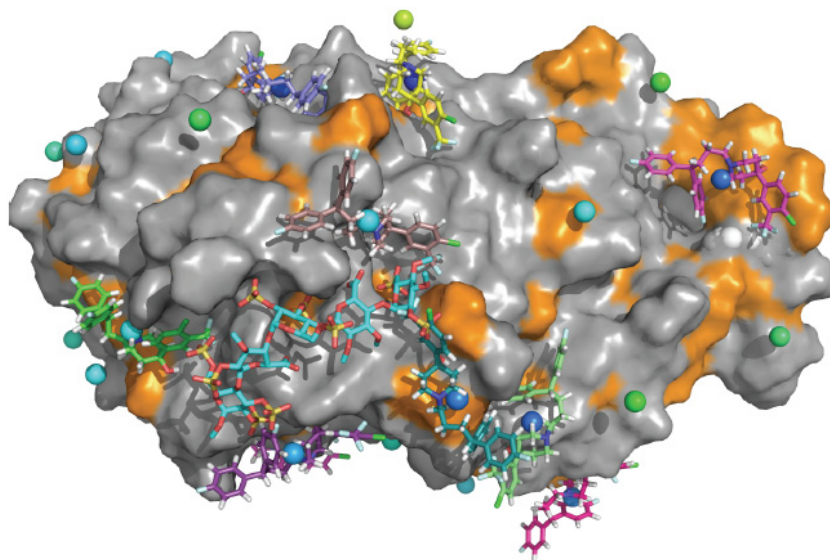


Fig. (5). Top 10 Penfluridol docking poses and three additional poses near the active site of antithrombin are selected for rescoring. The hydrophobic surfaces of antithrombin are colored in orange. The NTP compound is shown as a reference (carbon atoms shown in cyan). The color dots are the docking sites determined by BINDSURF. (*The color version of the figure is available in the electronic copy of the article.*)

top poses for rescoring, if the binding is non-specific or the binding mode is not well determined by the docking calculation.

The decomposition of the calculated binding free energy has demonstrated that the electrostatic energy changes upon binding in both gas phase (ΔE_{ELE}) and GB solvent ($\Delta G_{\text{GB-ELE}}$) are much less than those of the compounds in Table 1. On average the van der Waals energy changes (ΔE_{VDW}) upon binding are more than those of the compounds in Table 1. For Penfluridol, the ΔE_{VDW} is the main factor to determine the binding affinity. Pose Top6 has the largest ΔE_{VDW} , which results in the largest binding affinity for all thirteen docking

poses. As we pointed out in the previous section, the electrostatic interaction plays a crucial role in the ligand binding of the antithrombin active site due to the large number of positively charged residues present in the active site. Compound L1C4 and NTP possess strong electrostatic interactions with antithrombin and can bind much tighter than Penfluridol.

Penfluridol has only two hydrogen bond donors, nitrogen and oxygen (N1 and O1). Hydrogen bonding analysis of Penfluridol and antithrombin complex reveals that it has fewer hydrogen bond pairs and are less persistent during the MD simulations. For the pose Top1, there are only two hydrogen bonds (O1-HD22-ND2 and N1-HD22-ND2) formed

Table 2. Calculated binding free energies (kcal/mol) of the selected 13 Penfluridol docking poses.

Pose	ΔE_{ELE}	ΔE_{VDW}	ΔE_{gas}	ΔG_{Surf}	ΔG_{GB}	ΔG_{GB-ELE}	ΔG_{Sol}	ΔG_{bind}
Top1	-30.48±7.18	-30.07±3.14	-60.59±9.69	-2.02±0.30	33.85±8.92	3.37±1.51	31.84±8.72	-28.75±3.62
Top2	-25.67±6.60	-45.08±4.19	-70.40±7.83	-5.09±0.30	24.50±5.11	-1.17±3.02	19.41±5.03	-50.99±6.87
Top3	-19.57±4.02	-47.70±4.10	-66.84±5.53	-4.73±0.23	18.73±3.98	-0.85±2.48	14.00±3.94	-52.84±4.87
Top4	-24.64±7.57	-43.84±4.25	-68.63±7.74	-5.54±0.27	22.08±8.13	-2.56±3.92	16.54±8.22	-52.09±5.46
Top5	-32.59±6.24	-34.51±5.67	-67.00±9.62	-3.23±0.39	30.85±5.15	-1.74±2.98	27.62±5.01	-39.38±6.47
Top6	-33.11±4.60	-51.91±3.53	-84.45±5.18	-5.42±0.18	30.19±3.70	-2.92±3.49	24.77±3.70	-59.68±4.61
Top7	-21.95±5.59	-44.37±3.96	-65.13±8.06	-4.48±0.27	21.31±6.01	-0.65±3.10	16.82±5.86	-48.31±4.75
Top8	-32.17±5.25	-31.24±3.06	-62.71±6.16	-4.52±0.24	30.61±5.14	-1.56±2.32	26.09±5.14	-36.63±3.20
Top9	-13.82±4.07	-31.10±4.06	-44.17±5.91	-2.63±0.30	9.49±4.10	-4.34±2.65	6.86±4.02	-37.31±3.93
Top10	-35.34±9.07	-30.27±4.26	-66.02±10.92	-4.86±0.40	37.56±8.41	2.22±3.37	32.70±8.29	-33.32±5.40
Near1	-26.13±4.93	-26.60±2.73	-52.66±5.31	-2.90±0.25	19.13±4.26	-6.99±3.76	16.23±4.17	-36.42±3.66
Near2	-31.59±6.05	-41.50±4.57	-72.36±7.68	-4.73±0.37	25.85±4.85	-5.74±3.61	21.12±4.73	-51.24±5.24
Near3	-30.45±7.44	-40.63±5.82	-69.95±10.60	-4.53±0.49	30.23±7.31	-0.22±4.45	25.70±7.01	-44.24±5.65

during the MD simulation between Penfluridol and two asparagine residues (Asn61, Asn414). The hydrogen bond distances for O1-HD22-ND2 and N1-HD22-ND2 are 2.86 and 2.92 Å, respectively. Both hydrogen bond angles are ~20 degrees. The persistence of hydrogen bond O1-HD22-ND2 is 16.62% and that of hydrogen bond N1-HD22-ND2 is 36.11%. Only one hydrogen bond exists in the initial docking pose, where hydrogen bond O1-HD22-ND2 is formed between Penfluridol and Asn61. After ~18 ns of simulation, hydrogen bond O1-HD22-ND2 is split while hydrogen bond N1-HD22-ND2 starts to form. Pose Top2 has four hydrogen bonds between Penfluridol and two arginine residues. The persistence of these hydrogen bonds ranges from 15.25% to 33.35%. There are no hydrogen bonds present in the initial docking pose. The docking score is relatively high in this pose because of a hydrophobic interaction between a phenyl ring and Val375 and a π - π interaction between a phenyl ring and Arg221. The π - π interaction is lost and two hydrogen bonds form between Penfluridol and Arg221 after 36 ns. Pose Top3 has only one hydrogen bond between Penfluridol and Asn121. However, the persistence is 55.15%, which is the largest among the thirteen docking poses. The hydrogen bond is absent from the initial docking pose but appears quick after ~4ns. Pose Top4 has four hydrogen bonds form between Penfluridol and Lys283/Ser277 during the course of the MD simulation. It is very interesting that the four hydrogen bonds appear and disappear one by one sequentially. So, any time point during the MD simulation, there is only one hydrogen bond present between Penfluridol and antithrombin. Pose Top5 has three hydrogen bonds between Penfluridol and Lys93. All of them are very weak hydrogen bonds with angles greater than 40 degrees. The persistence of them is about 10%. Pose Top6 has one hydrogen bond between Penfluridol and Trp175. The persistence is 46.85%. The hydrogen bond forms after less than 2 ns of MD simulation. Pose Top6 has the largest rescoring binding affinity amongst

the thirteen docking poses, which is largely due to its binding in the hydrophobic groove of antithrombin. Pose Top7, Top8, Top9, Top10, Near1, Near2, and Near3 have 0, 1, 3, 3, 4, 5, and 5, respectively. The poses nearest to the active site (Near1, Near2, and Near3) tend to have more hydrogen bonds. However, none of them are strong hydrogen bonds. The persistence is usually around 15% for them. The number of hydrogen bonds for the Penfluridol docking poses is no more than 5. The good binders (*e.g.* L1C4 and NTP) each have more than 25 hydrogen bonds.

Navarro-Fernandez and colleagues have also predicted the interacting residues by using the FlexScreen scoring function [22]. For compound L1C4, they identified Lys11, Asn45, Arg46, Arg47, Lys114, and Lys125 as key residues, but Arg13 is missing from their list. For the Compound L1C4 docking calculation, they used the receptor structure from the X-ray crystal structure of antithrombin complexed with Compound NTP. Thus, the initial receptor structure for Compound L1C4 docking is biased. Docking calculations usually hold the receptor protein rigid. A few docking programs are able to have set side-chains of key residues in the receptor as flexible. However, most docking programs cannot sample the larger configuration space for the whole ligand-receptor complex. From our MD simulations, we find that Arg13 is quite flexible and can be adjusted to accommodate both large (*e.g.* Compound NTP) and small (*e.g.* Compound L1C4) compounds, which shows the advantages of using MD simulations over simple docking calculations.

Judging from the hydrogen bond analysis on Compound L1C4 and NTP, Arg47, Arg13, and Asn45 play crucial roles in the antithrombin binding process. Antithrombin provides multiple sulfate/phosphate binding sites, consisting of mostly positively charged residues (arginine, lysine) and neutral charged residues that can provide rich hydrogen bond donors/acceptors (asparagine). All four phosphate groups of

Compound L1C4 form hydrogen bonds with antithrombin while not all the sulfate groups of Compound NTP can form hydrogen bonds with antithrombin. As pointed out above, introducing a positively charged group in the ligand will result in a penalty for solvation free energy. If the positively charged group also cannot form favorable interactions (*e.g.* hydrogen bonding), ligand efficiency will be reduced, explaining why Compound L1C4 has higher ligand efficiency than Compound NTP.

Comparing the results from single-structure and ensemble-average MM/GBSA rescoring, the latter yields more accurate results. The ensemble-average MM/GBSA rescoring ranks the binding affinities of antithrombin ligands in the order that agrees with the experimental results. The advantage of ensemble-average MM/GBSA rescoring is that the binding affinity is averaged from an ensemble of structures extracted from long-time MD simulations. Long-time MD simulations can explore more configuration space and find energetically favorable configurations, which could offset the bias of initial structures. This can be verified in the MD trajectory of Compound L1C4. Arg13 is observed to form hydrogen bonds with the phosphate group of Compound L1C4 after 8 ns of MD simulation. Hou *et al.* have also identified the importance of MD sampling for recognizing the correct binding poses [11], although they carried out relatively short-time (300ps) MD simulations for the top three poses of 13 protein–ligand complexes and extracted only 60 snapshots from the trajectory for the ensemble-average MM/GBSA rescoring. They concluded even short MD simulation could help MM/GBSA to identify the experimentally determined binding conformations in most cases.

The ensemble-average MM/GBSA rescoring method is relatively accurate compared to single-structure MM/GBSA. However, the method to obtain an ensemble of structures, long-time MD simulations, is computationally intensive. Since the drug-like virtual libraries often contain millions of compounds, high-throughput virtual screening could be very costly, if using a more accurate, but expensive, method at the very beginning of the screening. To bridge the gap, our virtual screening pipeline uses a down-select scheme, a hierarchical approach, to screen large virtual compound libraries. A standard procedure to run the pipeline is to down-select compounds after they pass each screening method as implemented in the pipeline. The first screening method in the pipeline is VinaLC docking [2]. Top ranked poses of down-selected compounds after docking are rescored using a single-structure MM/GBSA rescoring method. From this study, we know it is important to keep a few top docking poses for rescoring because the best binding pose in rescoring is often not the top docking pose, especially for non-specific binding or not well-determined binding modes in the docking calculation. Finally, the most expensive ensemble-average MM/GBSA rescoring method in the pipeline can be applied to an amenable number of compounds down-selected after single-structure MM/GBSA rescoring, providing the accuracy needed for a fewer number of compounds.

CONCLUSION

In this article, we introduce a new addition, ensemble-average MM/GBSA rescoring, to our virtual screening pipe-

line. As a proof of concept, we calculated the binding affinities of eight antithrombin ligands by employing the previous single-structure MM/GBSA rescoring method and newly developed ensemble-average MM/GBSA rescoring method. A total of 20 MD simulations were performed for the ensemble-average MM/GBSA rescoring method. The correlation coefficient of calculated and experimental binding affinities is improved from 0.36 to 0.69 when using ensemble-average MM/GBSA rescoring. The rank order of calculated binding free energies using ensemble-average MM/GBSA rescoring exactly matches the experimentally derived free energies. We demonstrate that long-time MD trajectory can explore more configuration space and find energetically favorable configurations so that it can offset the bias of initial structures and improve the accuracy of binding affinity prediction. The electrostatic interactions in both solute and solvent contribute favorably to the binding free energy. Adding more negatively charged groups to the ligand provides more favorable electrostatic energy change. However, simultaneously it creates a higher penalty for the solvation free energy. The penalty can be compensated for by forming more hydrogen bonds as more negatively charged groups are added into the ligand. The negatively charged groups added to the ligand must actively interact with the receptor by forming hydrogen bonds to achieve high ligand efficiency. Compound L1C4 has higher ligand efficiency because it uses all its phosphate groups to form hydrogen bonds with antithrombin while Compound NTP does not. A hydrophobic compound, Penfluridol, has been docked to the antithrombin, which has abundant charge residues on the surface and in the active site of the protein. The docking results have shown the binding of Penfluridol is non-specific. The electrostatic energy changes, upon binding Penfluridol in both gas phase and GB solvent, are much less than those of good binders. The van der Waals energy change upon binding is the dominant factor for the Penfluridol binding affinity. The number of hydrogen bonds present in the Penfluridol MD simulations are fewer than that of the good binders. We have also demonstrated that including a few top docking poses for rescoring is important especially for non-specific binding or binding modes that are not well-determined in the docking calculation.

DISCLOSURE

We have previously reported 7 long-time MD simulations as part of conference proceedings [3], and here we expand upon this work to disclose a total of 20 long-time simulations.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The authors thank Livermore Computing for the computer time and Laboratory Directed Research and Development for funding (12-SI-004). We also thank Livermore Computing Grand Challenge for extensive computing resources. This work was performed under the auspices of the United States Department of Energy by the Lawrence

Livermore National Laboratory under Contract DE-AC52-07NA27344. Release Number LLNL-JRNL-666361. This work was partially supported by the Fundación Séneca la Región de Murcia under Project 18946/JLI/13. This work has been funded by the Nils Coordinated Mobility under grant 012-ABEL-CM-2014A, in part financed by the European Regional Development Fund (ERDF).

REFERENCES

- [1] Zhang, X.; Wong, S.E.; Lightstone, F.C. Toward fully automated high performance computing drug discovery: A massively parallel virtual screening pipeline for docking and molecular mechanics/generalized born surface area rescoring to improve enrichment. *J. Chem. Inf. Model.*, **2014**, *54*(1), 324-337.
- [2] Zhang, X.; Wong, S.E.; Lightstone, F.C. Message passing interface and multithreading hybrid for parallel molecular docking of large databases on petascale high performance computing machines. *J. Comput. Chem.*, **2013**, *34*(11), 915-927.
- [3] Zhang, X.; Pérez-Sánchez, H.; Lightstone, F. In *Bioinformatics and Biomedical Engineering*. Ortuño, F.; Rojas, I., Eds.; Springer International Publishing USA, **2015**; vol. *9044*, pp. 584-593.
- [4] Pérez-Sánchez, H.; Cano, G.; García-Rodríguez, J. Improving drug discovery using hybrid softcomputing methods. *Appl. Soft Comput.*, **2014**, *20*, 119-126.
- [5] Sanchez-Linares, I.; Perez-Sanchez, H.; Cecilia, J.; Garcia, J. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinform.*, **2012**, *13*(Suppl 14), S13.
- [6] Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.*, **2011**, *32*(4), 742-755.
- [7] Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.*, **2010**, *31*(4), 797-810.
- [8] Beveridge, D.L.; DiCapua, F.M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.*, **1989**, *18*(1), 431-492.
- [9] Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.*, **2011**, *32*(5), 866-877.
- [10] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **2000**, *33*(12), 889-897.
- [11] Hou, T.J.; Wang, J.M.; Li, Y.Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.*, **2011**, *51*(1), 69-82.
- [12] Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E), better ligands and decoys for better benchmarking. *J. Med. Chem.*, **2012**, *55*(14), 6582-6594.
- [13] Rastelli, G.; Rio, A.D.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.*, **2010**, *31*(4), 797-810.
- [14] Greenidge, P.A.; Kramer, C.; Mozziconacci, J.C.; Wolf, R.M. MM/GBSA Binding energy prediction on the PDBbind data set: Successes, failures, and directions for further improvement. *J. Chem. Inf. Model.*, **2013**, *53*(1), 201-209.
- [15] Lafont, V.; Armstrong, A.A.; Ohtaka, H.; Kiso, Y.; Mario Amzel, L.; Freire, E. Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem. Biol. Drug Des.*, **2007**, *69*(6), 413-422.
- [16] Turko, I.V.; Fan, B.; Gettins, P.G.W. Carbohydrate isoforms of antithrombin variant N135Q with different heparin affinities. *FEBS Lett.*, **335**(1), 9-12.
- [17] Jin, L.; Abrahams, J.P.; Skinner, R.; Petitou, M.; Pike, R.N.; Carrell, R.W. The anticoagulant activation of antithrombin by heparin. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*(26), 14683-14688.
- [18] Thunberg, L.; Bäckström, G.; Lindahl, U. Further characterization of the antithrombin-binding sequence in heparin. *Carbohydr. Res.*, **1982**, *100*(1), 393-410.
- [19] Verli, H.; Guimarães, J.A. Insights into the induced fit mechanism in antithrombin-heparin interaction using molecular dynamics simulations. *J. Mol. Graph. Model.*, **2005**, *24*(3), 203-212.
- [20] Pol-Fachin, L.; Franco Becker, C.; Almeida Guimarães, J.; Verli, H. Effects of glycosylation on heparin binding and antithrombin activation by heparin. *Proteins: Struct. Funct. Bioinf.*, **2011**, *79*(9), 2735-2745.
- [21] Bitomsky, W.; Wade, R.C. Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8. *J. Am. Chem. Soc.*, **1999**, *121*(13), 3004-3013.
- [22] Merlitz, H.; Wenzel, W. Comparison of stochastic optimization methods for receptor-ligand docking. *Chem. Phys. Lett.*, **2002**, *362*(3-4), 271-277.
- [23] Navarro-Fernández, J.; Pérez-Sánchez, H.; Martínez-Martínez, I.; Meliciani, I.; Guerrero, J.A.; Vicente, V.; Corral, J.; Wenzel, W. *In Silico* discovery of a compound with nanomolar affinity to antithrombin causing partial activation and increased heparin affinity. *J. Med. Chem.*, **2012**, *55*(14), 6403-6412.
- [24] Ponder, J.W.; Case, D.A. Force fields for protein simulations. *Protein Simul.*, **2003**, *66*, 27-85.
- [25] Wang, J.M.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.*, **2004**, *25*(9), 1157-1174.
- [26] Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, **2006**, *25*(2), 247-260.
- [27] Jakalian, A.; Bush, B.L.; Jack, D.B.; Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, **2000**, *21*(2), 132-146.
- [28] Case, D.A.; Cheatham, T.E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.*, **2005**, *26*(16), 1668-1688.
- [29] Onufriev, A.; Bashford, D.; Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins-Struct. Funct. Bioinf.*, **2004**, *55*(2), 383-394.
- [30] Reynolds, C.H.; Tounge, B.A.; Bembek, S.D. Ligand binding efficiency: Trends, physical basis, and implications. *J. Med. Chem.*, **2008**, *51*(8), 2432-2438.