

A New Method for Navigating Optimal Direction for Pulling Ligand from Binding Pocket: Application to Ranking Binding Affinity by Steered Molecular Dynamics

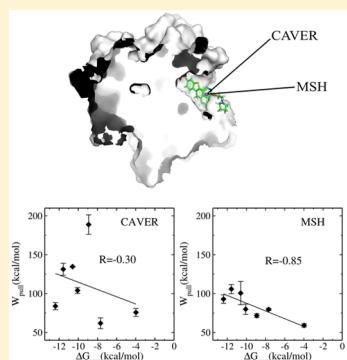
Quan Van Vuong,^{†,§} Tin Trung Nguyen,^{†,§} and Mai Suan Li^{*,†,‡}

[†]Institute for Computational Science and Technology, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam

[‡]Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Supporting Information

ABSTRACT: In this paper we present a new method for finding the optimal path for pulling a ligand from the binding pocket using steered molecular dynamics (SMD). Scoring function is defined as the steric hindrance caused by a receptor to ligand movement. Then the optimal path corresponds to the minimum of this scoring function. We call the new method MSH (Minimal Steric Hindrance). Contrary to existing navigation methods, our approach takes into account the geometry of the ligand while other methods including CAVER only consider the ligand as a sphere with a given radius. Using three different target + receptor sets, we have shown that the rupture force F_{\max} and nonequilibrium work W_{pull} obtained based on the MSH method show a much higher correlation with experimental data on binding free energies compared to CAVER. Furthermore, W_{pull} was found to be a better indicator for binding affinity than F_{\max} . Thus, the new MSH method is a reliable tool for obtaining the best direction for ligand exiting from the binding site. Its combination with the standard SMD technique can provide reasonable results for ranking binding affinities using W_{pull} as a scoring function.



INTRODUCTION

Discovery of a new drug approved by the FDA usually takes about 10 years and costs billions of US dollars. Recently, computers have become useful tools to speed up the drug discovery process reducing costs. The computer-aided drug design provides prediction of potential drugs based on the binding affinity of a ligand to a receptor. The binding affinity is estimated using different scoring functions. One of the most popular ones is the binding energy used in the simple docking method.^{1–4} However, the predictive power of this method is very limited due to crude approximations such as omission of dynamics of the receptor and restricted numbers of trials for ligand conformations. In return, due to the high speed the docking simulation is widely used for virtual screening when dealing with huge databases.

In order to go beyond the docking approximation one can compute the absolute binding free energy by all-atom molecular dynamics simulations using numerous methods such as free energy perturbation (FEP),⁵ thermodynamic integration (TI),⁶ linear response approximation (LRA),⁷ linear interaction energy (LIE),⁸ molecular mechanic- Poisson-Boltzmann surface area (MM-PBSA),⁹ etc. The main disadvantage of these methods is that they are computationally expensive,¹⁰ but in return, they are more accurate compared to the docking method.

Recently, the steered molecular dynamics (SMD) method^{11,12} has been shown to be a useful tool for drug design.¹³ In this method the scoring function is either the rupture force F_{\max}

defined as the maximum in the force-extension/time profile or unbinding work W_{pull} generated by the external force applied to the ligand. Strictly speaking, SMD was designed for studying nonequilibrium unfolding/unbinding processes. However, it may be used to understand ligand binding assuming that the larger the F_{\max} or W_{pull} , the stronger the binding is.^{13–16} For some systems SMD is as accurate as the MM-PBSA method in ranking binding affinities,^{13–15} but computationally it is about 1 order of magnitude faster. From this point of view, SMD can be used as a reliable tool which is complementary to the docking and equilibrium MD methods.

In contrast from the standard SMD applied to the protein mechanical unfolding problem,^{12,17} before driving the ligand from the crowded binding site one has to find a pathway along which the ligand is pulled out. Because the rupture force and work in the nonequilibrium unbinding process are sensitive to the choice of pulling direction, their accurate determination is pivotal for predicting binding affinity. There are several programs developed to identify exit tunnels for ligands, such as CAVER^{18,19} and MOLE.²⁰ However, none of them carefully considers the geometry of a ligand assuming it as a sphere with a given radius. Thus, our main goal is to overcome this drawback in navigating optimal pulling pathway from the binding pocket. For this purpose, we introduced a new method which minimizes the steric hindrance caused by a receptor on a

Received: June 17, 2015



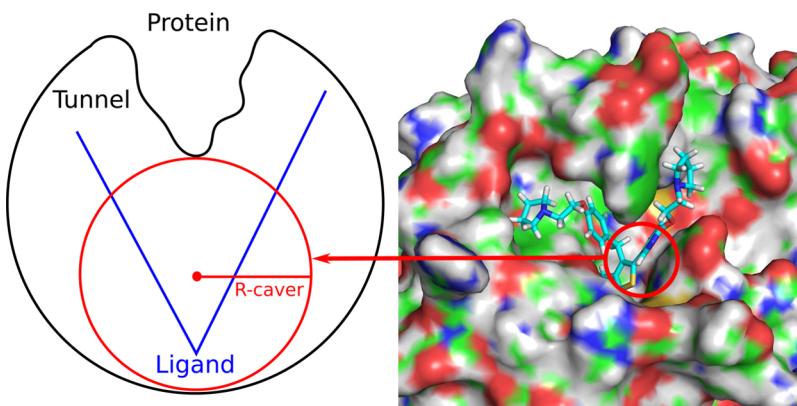


Figure 1. Two-dimensional schematic plot showing that ligand (blue) can go along the channel which has diameter less than the typical size of ligand (left). Such a pathway is not allowed in CAVER but acceptable in MSH. Direction of channel is perpendicular to the sheet. The corresponding three-dimensional image is shown in the right.

ligand while taking the ligand geometry into account. We will refer to it as MSH (Minimal Steric Hindrance).

The validity of our method for identifying the optimal pulling pathway was tested against three different sets of model systems for which structures of protein–ligand complexes and binding free energies are experimentally available. W_{pull} and F_{\max} calculated by the standard SMD method^{12,13} but with optimized pulling directions given by MSH and CAVER, were compared with the experimental free binding energies. One can show that the results obtained by our approach for both of these quantities correlate with experiments better than CAVER. Thus, MSH is presumably reliable in finding the pulling pathway of small molecules from the binding site.

■ NEW MSH METHOD

In this section we describe the scoring function for choosing an optimal direction for pulling the ligand and outline the main steps of MSH.

Scoring Function for Steric Hindrance. In the CAVER method the optimal path corresponds to the widest and shortest tunnel for ligand exit from the binding site neglecting important factors such as the steric hindrance of the receptor to ligand movement as well as the geometry of the ligand which was considered as a sphere. The drawback of omitting ligand geometry is illustrated in Figure 1, where the ligand can pass through the channel although its typical size is larger than the channel diameter obtained by CAVER.

Our main assumption is that the optimal pulling direction is a direction which minimizes the steric hindrance of the receptor to movement of the pulled ligand. In addition the realistic structure of the ligand is taken into account. For the complex of the receptor of N_r atoms and ligand of N_l atoms with origin O placed at the center of mass (COM) of the ligand, the scoring function for a given direction \vec{Ov} is defined as the total weighted hindrance acting on each atom of ligand:

$$S = \sum_{i=1}^{N_l} \frac{\text{Hdr}_i}{\sqrt{r_i}} \quad (1)$$

Here Hdr_i is the hindrance caused by a receptor due to the movement of atom i of the ligand in the \vec{Ov} direction (Figure 2). Its definition will be given below. r_i is the distance from atom i to axis of direction \vec{Ov} (Figure 2). For a pulled molecule, the atom located far away from the pulling direction axis is

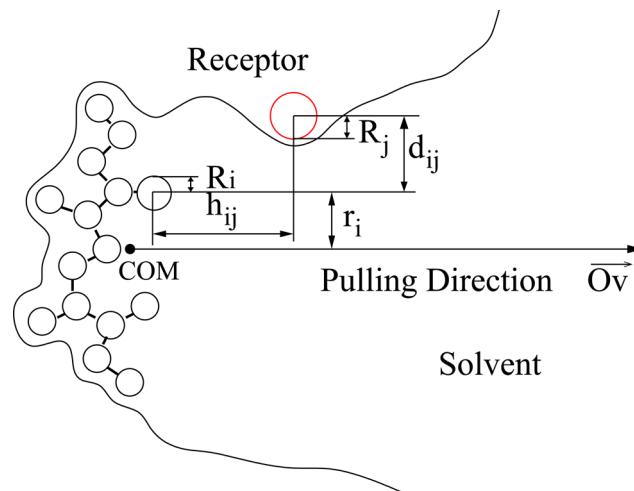


Figure 2. Terms used to evaluate hindrance caused by receptor atom j on movement of ligand atom i . Ligand is represented by black linked circles, while red circle represents atom of receptor.

more flexible than the close one. As a result the contribution of the further away atom to the total hindrance is less than the closer one. To mimic this effect $1/\sqrt{r_i}$ is used in eq 1. To avoid the possible divergence of S (eq 1) related to the limit $r_i \rightarrow 0$, we renormalize r_i in such a way that

$$r_i = \begin{cases} r_i & \text{if } r_i > 1 \\ 1 & \text{if } r_i \leq 1 \end{cases} \quad (2)$$

where r_i and other distances are measured in Å if not otherwise stated.

Hdr_i is defined as follows:

$$\text{Hdr}_i = \max \left(\frac{w_j O_{ij} \sqrt{\epsilon_i \epsilon_j}}{\sqrt{h_{ij}^e}} \right); 1 \leq j \leq N_r \quad (3)$$

Here w_j is the weight for atom j of the receptor. If receptor atom j stands behind atom i of the ligand in the \vec{Ov} direction, then it causes no hindrance to the movement of atom i in this direction and w_j is set to zero. Otherwise w_j is set to 1.0 if atom j belongs to the side chain and is greater than 1.0 for the backbone atoms which is more rigid than the side chain ones. The choice of w_j for backbone atoms is somewhat arbitrary, but

we have taken it as large as 1.3. Variation of w_j in the interval of 1.1–1.5 does not significantly change the results. O_{ij} , which measures the maximum overlap of atom i and atom j when atom i moves along the \vec{Ov} direction, is defined as follows

$$O_{ij} = \begin{cases} (d_{ij} - R_i - R_j)^2 & \text{if } d_{ij} - R_i - R_j < 0 \\ 0 & \text{if } d_{ij} - R_i - R_j \geq 0 \end{cases} \quad (4)$$

where d_{ij} is the distance between projected images of atom i and atom j on a plane perpendicular to the \vec{Ov} direction (Figure 2). R_i and R_j are the van der Waals radius of atoms i, j which are taken from the Amber force field²¹ (Table 1).

Table 1. van der Waals Radius R and Depth of the Potential Well ϵ for Relevant Atoms

atom	R (Å)	ϵ (kcal/mol)
H	1.487	0.0157
C	1.908	0.1094
N	1.824	0.1700
O	1.721	0.2104
P	2.100	0.2000
S	2.000	0.2500
Cl	1.948	0.2650
Br	2.220	0.3200
I	2.350	0.4000

Clearly, O_{ij} is proportional to the steric hindrance of atom j on the movement of atom i in the \vec{Ov} direction. The effective distance between atoms i and j in the pulling direction h_{ij}^e is given by the following expression:

$$h_{ij}^e = \begin{cases} h_{ij} - R_i - R_j + 1 & \text{if } h_{ij} - R_i - R_j > 0 \\ 1 & \text{if } h_{ij} - R_i - R_j \leq 0 \end{cases} \quad (5)$$

Here h_{ij} is the distance between projected images of atoms i and j on axis \vec{Ov} (Figure 2). The larger the distance is between receptor atom j and ligand atom i , the lower the impact of atom j is on the movement of ligand atom i . ϵ_i and ϵ_j are depths of the Lennard-Jones potential well²² of atom i and j , respectively. They are taken from the Amber force field²¹ and listed in Table 1. The square root of ϵ_i and ϵ_j which are proportional to the repulsion between two atoms in the Lennard-Jones potential is adopted as the relative harness of atoms i and j , i.e. the harder atom posing more hindrance than the softer one.

Main Steps in MSH. In order to find the optimal pulling path by the MSH method one has to follow the following steps:

- A sphere surface grid map is built for scanning every possible direction. The center of sphere O is placed at the COM of the ligand. The degree of vertex O formed by O and two adjacent points on the grid map is grid map resolution (default value is 1°).

- For each grid point, the pulling direction \vec{Ov} is defined by the vector from center O to a given grid point.
- For each atom i of the ligand, calculate r_i .
- If receptor atom j stands in front of ligand atom i in direction \vec{Ov} , then calculate the following:

$$\frac{w_j O_{ij} \sqrt{\epsilon_i \epsilon_j}}{\sqrt{h_{ij}^e}}$$

- Compute hindrance $H_{dr,i}$ for every atom i of the ligand using the definition given by eq 3.
- Calculate score S for a given direction \vec{Ov} using eq 1.
- Choose the optimal direction by minimizing scoring function S .

Once the optimal pulling path was found, one can apply the standard SMD method to compute F_{\max} and W_{pull} for predicting binding affinity.

MATERIALS AND METHODS

Choice of Test Sets. Three test sets including α -thrombin (7 ligands), neuraminidase (8 ligands), and penicillopepsin (7 ligands) were selected from a paper²³ where authors used them to verify the accuracy of the MM-PBSA and MM-GBSA methods in estimating binding free energy. In fact, six sets have been studied, but we decided to choose only three of them. Avidin and P450cam sets were excluded due to the lack of experimental structures for some complexes. The set of Cytochrome C Peroxidase was also ignored because of the presence of molecule Protoporphyrin IX (HEM) containing an Fe atom which is very difficult to describe by classical molecular mechanics.

For convenience, the PDB IDs and experimental binding free energies of three chosen sets are given in Tables 2–4. The chemical structures and their protonated states for all ligands are shown in Tables S1–S3 in the Supporting Information (SI).

Preparation of Protein–Ligand Complexes. Atomic structures of protein–ligand complexes were taken from the RSCB Protein Data Bank²⁴ according to their PDB ID. Hydrogen is added to the receptor and ligand using the PDB 2PQR²⁵ server and Avogadro package,²⁶ respectively. The complexes were first solvated in water with the three-site TIP3P water model,²⁷ and then ions Cl^- or Na^+ were added to neutralize the systems. The size of the water box depends on

Table 2. Calculated Values of F_{\max} and W_{pull} and Experimental Binding Free Energies for α -Thrombin Set^a

PDB	ΔG^{exp}	F_{\max}^{caver}	$W_{\text{pull}}^{\text{caver}}$	F_{\max}^{msh}	$W_{\text{pull}}^{\text{msh}}$
1D3D	-12.39	660.47 ± 30.14	83.86 ± 4.59	649.75 ± 10.93	92.95 ± 5.49
1D3P	-10.08	669.91 ± 48.03	103.95 ± 3.73	521.38 ± 41.10	79.90 ± 6.63
1D3Q	-8.92	1103.56 ± 49.16	188.84 ± 12.44	553.86 ± 20.71	71.61 ± 2.53
1D3T	-7.68	462.19 ± 41.83	61.84 ± 6.92	615.36 ± 24.75	79.66 ± 1.43
1DWB	-3.98	643.48 ± 34.04	75.76 ± 5.05	668.29 ± 26.76	59.05 ± 2.18
1DWC	-10.60	891.32 ± 25.94	134.68 ± 0.67	790.65 ± 74.06	100.61 ± 14.94
1DWD	-11.57	785.91 ± 19.38	131.34 ± 7.66	665.14 ± 5.44	105.80 ± 5.86

^a F_{\max} is measured in pN, while W_{pull} and ΔG^{exp} are in kcal/mol.

Table 3. Calculated Values of F_{\max} and W_{pull} and Experimental Binding Free Energies for Neuraminidase Set^a

PDB	ΔG^{exp}	F_{\max}^{caver}	$W_{\text{pull}}^{\text{caver}}$	F_{\max}^{msh}	$W_{\text{pull}}^{\text{msh}}$
1NSC	-4.09	946.19 ± 46.49	128.86 ± 5.29	1047.59 ± 76.61	130.37 ± 10.81
1NSD	-7.23	795.42 ± 38.28	98.53 ± 1.09	939.49 ± 73.87	121.28 ± 15.38
2QWB	-3.74	865.31 ± 61.72	102.65 ± 6.34	749.66 ± 6.48	95.81 ± 3.43
2QWC	-4.84	895.98 ± 48.17	128.43 ± 1.82	781.74 ± 31.08	87.34 ± 4.88
2QWD	-6.61	959.96 ± 56.11	126.60 ± 2.01	961.15 ± 80.96	130.66 ± 10.32
2QWE	-10.20	935.28 ± 59.28	137.47 ± 12.90	1398.76 ± 112.92	193.47 ± 19.63
2QWF	-7.73	1326.04 ± 111.16	184.22 ± 18.29	1133.94 ± 104.02	168.35 ± 19.60
2QWG	-11.45	1301.74 ± 14.25	231.86 ± 19.47	1214.60 ± 75.21	174.67 ± 13.66

^a F_{\max} is measured in pN, while W_{pull} and ΔG^{exp} are in kcal/mol.

Table 4. Calculated Values of F_{\max} and W_{pull} and Experimental Binding Free Energies for Penicillopepsin Set^a

PDB	ΔG^{exp}	F_{\max}^{caver}	$W_{\text{pull}}^{\text{caver}}$	F_{\max}^{msh}	$W_{\text{pull}}^{\text{msh}}$
1APT	-12.83	769.22 ± 5.77	139.30 ± 4.16	834.01 ± 66.46	191.15 ± 10.25
1APU	-10.51	579.91 ± 51.22	148.44 ± 17.28	778.62 ± 10.09	150.24 ± 2.84
1APV	-12.27	853.90 ± 50.50	191.76 ± 1.19	1232.14 ± 83.07	185.79 ± 13.91
1APW	-10.91	821.77 ± 79.36	166.79 ± 10.67	825.07 ± 32.98	133.76 ± 6.22
2WEA	-8.37	671.57 ± 55.41	117.25 ± 6.62	443.04 ± 39.49	78.13 ± 9.89
2WEB	-7.03	763.30 ± 32.06	165.29 ± 12.75	685.65 ± 45.47	127.34 ± 16.51
2WEC	-6.80	485.18 ± 28.25	72.87 ± 8.90	465.63 ± 39.45	77.61 ± 5.53

^a F_{\max} is measured in pN, while W_{pull} and ΔG^{exp} are in kcal/mol.

the ligand and receptor. For the studied systems the number of atoms varies from 48 262 to 68 084. The topology and coordinate files of complexes were generated by utilities of AmberTools1.5 in Amber format file with Amber99sb²⁸ and gaff²⁹ force fields assigned for receptor and ligands, respectively. Finally, the topology and coordinate files were converted to the GROMACS format by the acpype code.³⁰

SMD. The SMD method was developed to study mechanical unfolding of biomolecules¹² and ligand unbinding from a receptor along a given direction.^{13,31} Technically, the ligand is connected to a dummy atom which has no charge and zero radius by a spring with spring constant k . The spring is connected to either a selected atom of the ligand or the ligand's COM (Figure 3). Here the dummy atom is attached to the first

heavy atom of the ligand in the pulling direction. Moving along the pulling direction with a constant loading rate v , the dummy atom experiences elastic force $F = k(\Delta x - vt)$, where Δx is the displacement of the pulled atom from the starting position. We have chosen the spring constant $k = 600 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$ which is a typical value for the cantilever used in AFM experiments.^{32,33} As in our previous work,^{14,15} the loading speed was set equal to $v = 5 \text{ nm/ns}$. All C_{α} atoms of the receptor which are 3 Å behind the last atom of the ligand in the pulling direction were restrained (Figure 3) to keep the receptor almost at the same place while still maximally maintaining its flexibility.

First, the complex was minimized using the steepest descent method. Then, the position-restrained MD simulation was performed for 500 ps in the NVT ensemble with a velocity-rescaling thermostat³⁴ followed by 2 ns in the NPT ensemble with a Parrinello–Rahman barostat³⁵ to make sure that the system was stable. In the last step, for each complex we performed two sets of SMD simulations for two pulling directions determined by MSH and CAVER methods. For complete unbinding the duration of SMD runs was set equal to 600 ps for α -thrombin and neuraminidase and 800 ps for penicillopepsin sets. Ligands of the last set are significantly larger than those of the two former sets. To obtain reliable results, for each complex three independent SMD simulations were carried out using different initial random seed numbers. Final results were averaged over three trajectories. All MD and SMD simulations in this study were performed using the GROMACS package.³⁶

Scoring Functions F_{\max} and W_{pull} for Binding Affinity. We used both F_{\max} and W_{pull} as scoring functions for binding affinity. F_{\max} is the maximum force, experienced by the dummy atom, in the force–extension/time profile (Figure 4).

The pulling work W_{pull} is defined as performed work to pull a ligand out from the binding pocket. It is calculated by the following formula:

$$W_{\text{pull}} = \int_0^{x_{\max}} \vec{F} \cdot d\vec{x} \approx \int_0^{t_{\text{pull}}} Fv \, dt \approx \sum_{i=1}^{N_{\text{step}}} F_i v \Delta t \quad (6)$$

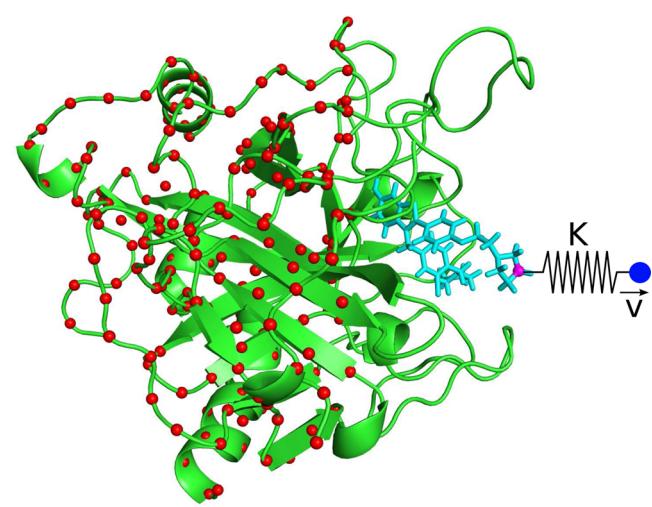


Figure 3. Schematic plot for pulling ligand out from a receptor by moving the dummy atom (blue one) which is connected to the front heavy atom (pink) of ligand by a spring with spring constant k in pulling direction \vec{v} . Receptor and ligand are highlighted by green and cyan, respectively. Red dots refer to restrained C_{α} atoms of receptor.

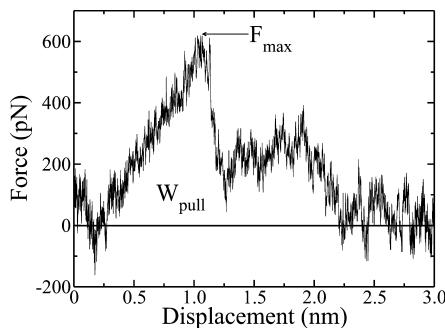


Figure 4. Typical dependence of force experienced by the dummy atom on the displacement from its initial position.

RESULTS AND DISCUSSIONS

MSH and CAVER provide different pulling directions. The upper panel of Figure 5 illustrates pulling directions

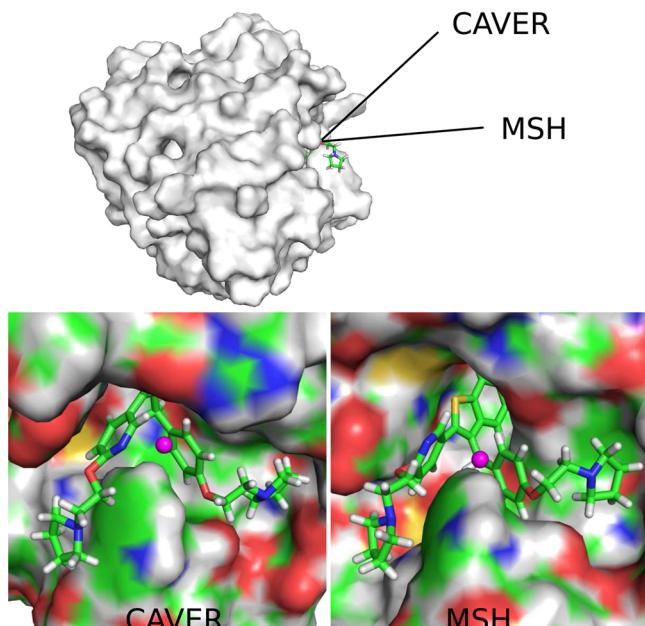


Figure 5. Upper panel: A schematic plot demonstrating the difference between pulling directions determined by MSH and CAVER methods. Lower panel: shown surface is perpendicular to the pulling direction. Small pink sphere represents COM of ligand.

obtained by MSH and CAVER for the ligand in the 1D3P complex from the α -thrombin set. As expected they are different, as the two methods use different scoring functions. This may be understood as follows. As evident from the lower panel (Figure 5), where pulling directions are perpendicular to the sheet, in the CAVER direction, the benzo[*b*]thiophene ring of the ligand is completely eclipsed by the wall of the binding pocket. Thus, the pulling ligand along this direction would encounter more hindrance than pulling it in the MSH direction where the benzo[*b*]thiophene ring is partially exposed.

More examples on pathways, obtained by CAVER and MSH, are shown in Figure S1 in the SI. For the neuraminidase set with sphere-like ligands, the pulling directions predicted by both methods are more similar than in the α -thrombin and penicillopepsin sets where the geometry of the ligands is complicated. Particularly, for 2QWE two pathways are nearly identical (Figure S1). The CAVER direction tends to direct the

ligand along a wider tunnel, but it is more eclipsed by the wall of the binding pocket than in the MSH case.

Dependence of Scoring Functions for Binding Affinity on Pulling Pathways. W_{pull} and F_{max} obtained by SMD simulations using MSH and CAVER pulling directions are shown in Tables 2–4 for the α -thrombin, neuraminidase, and penicillopepsin sets, respectively. There is a significant deviation in the performed work W_{pull} required to unbind ligands along MSH and CAVER pulling directions. The pulling work obtained in the MSH mode is smaller than that from CAVER for some complexes but greater for the others. However, overall, the unbinding process along CAVER directions consumes more energy than along MSH directions for all studied sets. The excess in W_{pull} is 190.69, 36.67, and 57.68 kcal/mol for the α -thrombin, neuraminidase, and penicillopepsin sets, respectively. This result provides additional support for our method that MSH energetically is more favorable than CAVER.

Comparison of ligand structures (Tables S1–S3 in the Supporting Information) reveals that the diversity in W_{pull} depends on the complexity of the ligand geometry. The variation is greatest for the α -thrombin set in which the geometries of ligands are the most complicated, as six of seven ligands compose three branches of the star geometry. The minimal diversity is observed for the neuraminidase set where the geometries of ligands are fairly round. Ligands from the penicillopepsin set although are larger than ligands of the α -thrombin set, but their geometries are simpler. Thus, the diversity of W_{pull} for the penicillopepsin set is less than that for the α -thrombin set.

This result is predictable because CAVER only considers the ligand as a sphere with a given radius causing the irreversible unbinding process along its pulling direction to consume more waste energy when ligand geometries are more complicated. In contrast to CAVER, MSH which takes into account the complexity of the ligand geometry minimizes the waste energy which comes from the steric hindrance.

MSH gives better correlation with experiments than CAVER. Figures 6 and 7 plot W_{pull} and F_{max} obtained in MSH and CAVER modes, as a function of experimental binding free energy ΔG^{exp} . For W_{pull} computed using the MSH pulling direction we have obtained the correlation level $R = -0.85$, -0.85 , and -0.87 for the α -thrombin, neuraminidase, and penicillopepsin sets, respectively. This correlation is clearly higher than $R = -0.30$, -0.70 , and -0.57 followed from the CAVER simulation. In the case of the α -thrombin set, CAVER gives no correlation with ΔG^{exp} as $R = -0.30$.

On the other hand, F_{max} obtained by both MSH and CAVER pulling directions, are correlated with the experimental ΔG^{exp} for neuraminidase and penicillopepsin sets (Figure 7). For α -thrombin one has $R = -0.24$ and $= -0.08$ for CAVER and MSH, respectively. Similar to scoring function W_{pull} , F_{max} based on MSH also shows a higher correlation with experiments than CAVER because $R = -0.82$ against $R = -0.56$ for the neuraminidase set and $R = -0.77$ against $R = -0.58$ for the penicillopepsin set. These results again confirm the advantage of the MSH method over CAVER.

W_{pull} is a better score than F_{max} . In this section we only consider the W_{pull} and F_{max} obtained by MSH pulling directions. From the perspective of ranking binding affinity, W_{pull} appears to be better than F_{max} , as it shows a high correlation with ΔG^{exp} for all sets, whereas F_{max} fails to rank ligand binding affinities in the α -thrombin set (Figures 6 and 7). The failure of F_{max} in this

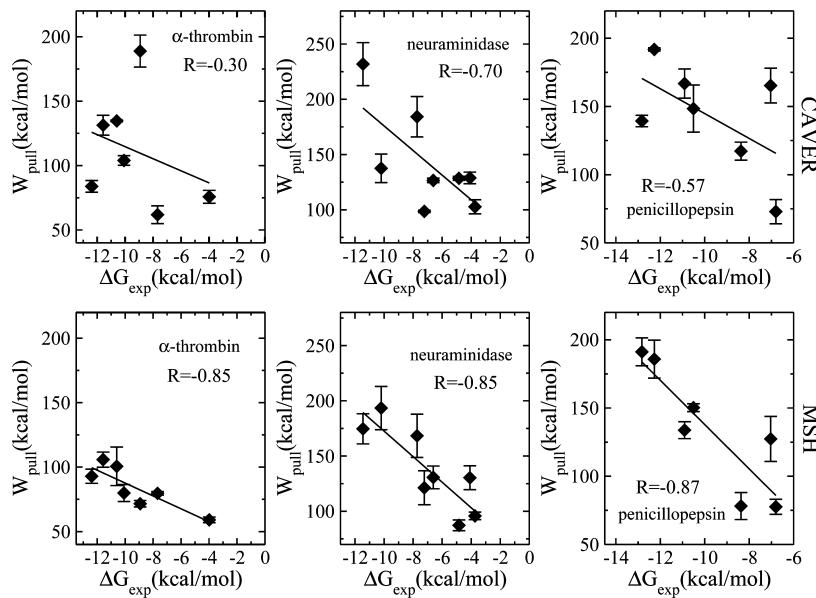


Figure 6. Correlation between unbinding work W_{pull} and experimental binding energies for α -thrombin, neuraminidase, and penicillopepsin sets.

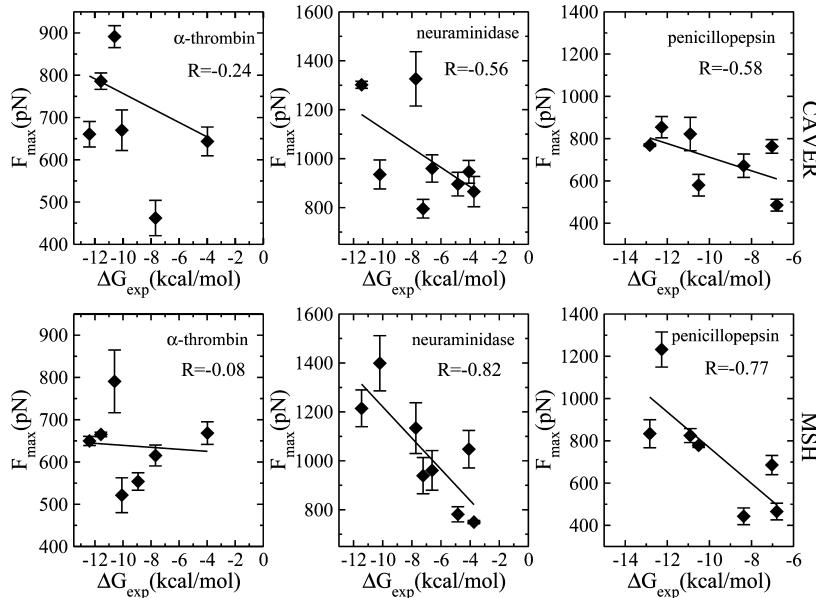


Figure 7. Correlation between unbinding work F_{max} and experimental binding energies for α -thrombin, neuraminidase, and penicillopepsin sets.

set stresses its weakness as a scoring function. In the α -thrombin set, the ligand of complex 1DWB, which has the lowest experimental binding affinity $\Delta G^{\text{exp}} = -3.98 \text{ kcal/mol}$, is ranked as the second by F_{max} ($F_{\text{max}} = 668.29 \text{ pN}$). Experimentally, the ligand from the 1D3D complex has the highest binding affinity, $\Delta G^{\text{exp}} = -12.39 \text{ kcal/mol}$, but it is ranked fourth having $F_{\text{max}} = 649.75 \text{ pN}$ (Table 2). The difference in ranking is due to the different unbinding mechanisms for these ligands (Figure 8). Being small and rigid, the ligand of the 1DWB complex binds to α -thrombin by three hydrogen bonds. Thus, its forced unbinding requires breaking three hydrogen bonds simultaneously. As a result the force increases steeply reaching a high peak and then drops rapidly (Figure 8). In contrast, the ligand of 1D3D is much larger and more flexible. Its force-induced unbinding process proceeded gradually, and the monotonic decrease of force occurs after passing the maximum (Figure 8). Therefore, the

work expended in the 1D3D case, $W_{\text{pull}} = 92.95 \text{ kcal/mol}$, is much larger than $W_{\text{pull}} = 59.05 \text{ kcal/mol}$ for 1DWB, even though the F_{max} of the former is smaller than the latter. Taken together, our simulations suggest that F_{max} is not suitable as a scoring function for ranking the binding affinity of ligands which are dramatically different from each other. Instead nonequilibrium work W_{pull} should be a better choice.

Robustness of Results against Model Parameters. Here we study the sensitivity of results to the spring constant and loading speed which are two main parameters in SMD simulations. Because the results are similar for three sets we focus on the neuraminidase case, where pulling directions were determined only by the MSH method.

In addition to loading speed $v = 5 \text{ nm/ns}$, we have performed simulations for $v = \frac{5}{3} \text{ nm/ns}$ and $5 \times 3 \text{ nm/ns}$ keeping spring constant $k = 600 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$. As v decreases W_{pull} lessens,

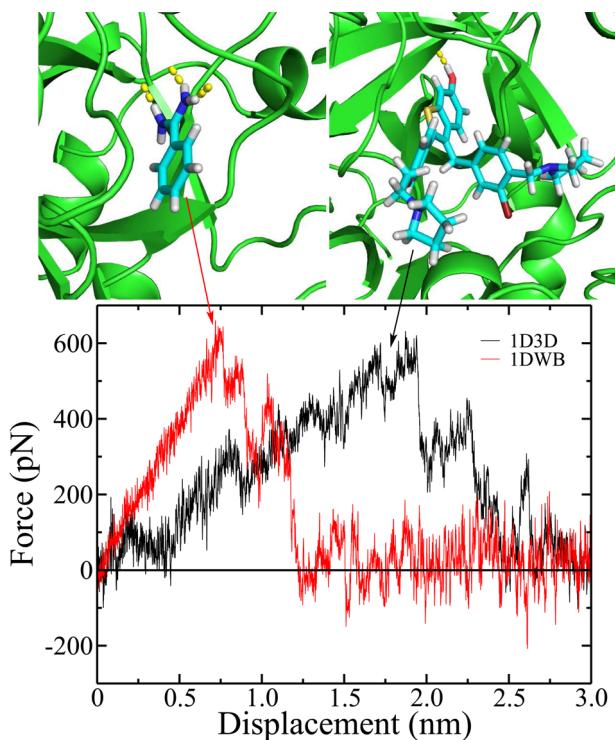


Figure 8. Dependence of force experienced by the dummy atom on the displacement from its initial position for complexes 1DWB and 1D3D in the α -thrombin set.

but the correlation level slightly changes from $R = -0.80$ for $5 \times 3 \text{ nm/ns}$ to -0.87 for $\frac{5}{3} \text{ nm/s}$ (Figure S2 in SI). This result is reasonable because upon $v \rightarrow 0$ one approaches equilibrium enhancing the correlation with experimental results obtained in equilibrium. One has to bear in mind that the decrease of loading speed also increases computing time. Therefore, the choice of v should depend on available recourses, but the typical value $v = 5 \text{ nm/ns}$ is a good option.

In order to investigate the robustness of results versus the spring constant the additional simulations were carried at fixed loading speed $v = 5 \text{ nm/ns}$. The results for three representative values of k are shown in Figure S3 in SI. As expected, for the range of $400 \leq k \leq 1400 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$ which is typical for AFM experiments,^{32,33} the dependence of R on the pulling speed is rather weak (Figure S4 in SI). This is because in this region the spring constant is rigid enough to have any visible effect on a nonequilibrium work.

Comparison with Other Methods. Docking Method. Autodock Vina version 1.1⁴ was used to estimate the binding energies of ligands to the receptor binding sites which are known from PDB holostructures for all three sets. Accurate results can be obtained setting the exhaustiveness of the global search equal to 1000. The maximum energy difference between the worst and best binding modes was set to 7 kcal/mol. A total of 10 binding modes were generated with random starting positions of ligand and fully flexible torsion degrees of freedom. The center of the grids was placed at the center of the binding site with grid dimensions large enough to cover the whole pocket.

The lowest binding energies ΔE_{dock} obtained in the best docking mode are compared with experimental binding free energies (Figure S5 in SI). Surprisingly, the docking binding energy shows a very high correlation with ΔG^{\exp} for α -thrombin

with $R = 0.9$. However, it becomes very bad for the neuraminidase set ($R = -0.01$) and even worse for penicillopepsin, where ΔE_{dock} anticorrelated with experiment ($R = 0.72$). Similar results were also obtained by Kim and Skolnik³⁷ showing that the docking simulation provides biased results. Thus, in general the docking method is not suitable for estimating the binding energy as well as for ranking binding affinities due to crude approximations involving omission of receptor dynamics and a limited number of trial positions for the ligand.

MM-PBSA Method. The correlation of binding free energies of three sets with experimental data has been studied via the MM/PBSA method by Hou et al.²³ They obtained $|R| = 0.80$, 0.68, and 0.41 for α -thrombin, neuraminidase, and penicillopepsin, respectively. For the α -thrombin set their result is compatible with ours, $|R| = 0.85$, but our correlation level, $|R| = 0.85$ and 0.87, for neuraminidase and penicillopepsin is higher. Thus, the SMD provides a higher correlation with the experiments than MM-PBSA for the three studied sets, but for other systems the SMD proved to be as accurate as the MM-PBSA method.¹³

CONCLUSION

We have proposed a new method to obtain the optimal direction to pull a ligand out from the binding pocket. In contrast to the widely used CAVER method and other methods, the geometry of the ligand, which is crucial for ranking binding affinity based on the pulling work or rupture force, was taken into account. It was shown that the results obtained based on MSH correlate with the experimental available data better than those from CAVER. F_{\max} can be used as a scoring function for some cases where the geometries of the ligands are not diverse, but W_{pull} is a superior choice in general.

The correlation between SMD and experimental results is improved with a decrease in loading speed because the system approaches closer to equilibrium where experimental measurements have been performed. However, the choice of v should be chosen in such a way that one can hold the balance between the computational time and accuracy of results. Our results suggest that $v \approx 1 \text{ nm/ns}$ is suitable for estimating W_{pull} for ranking binding affinities. Because the level of correlation with experiments is not sensitive to the spring constant in the AFM relevant interval, $400 \leq k \leq 1400 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$, one can use any k value from this region for SMD simulations.

In summary, our results suggest that MSH is a reliable method to determine the ligand pulling direction minimizing steric hindrance. The combination of MSH and standard SMD would be a useful tool for the drug design problem.

ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00386.

Chemical structures and protonated states of ligands in the α -thrombin, neuraminidase, and penicillopepsin sets (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: masli@ifpan.edu.pl. Phone: +(48 22) 843 66 01 (3326). Fax: +(48 22) 847 52 23.

Author Contributions

[§]Q.V.V. and T.T.N. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by Department of Science and Technology at Ho Chi Minh City, Vietnam and Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant Number 106-YS.02-2013.01. T.T.N. thanks Ngo Thanh Cong for providing computational resources.

REFERENCES

- (1) Clauben, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient Molecular Docking Considering Protein Structure Variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- (2) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (3) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: Combining Techniques to Model RNA-small Molecule Complexes. *RNA* **2009**, *15*, 1219–1230.
- (4) Trott, O.; Olson, A. J. Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.
- (5) Zwanzig, R. High Temperature Equation of State by a Perturbation Method. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (6) Kirkwood, J. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (7) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. Calculations of Antibody-Antigen Interactions: Microscopic and Semi-microscopic Evaluation of the Free Energies of Binding of Phosphorylcholine Analogs to McPC603. *Protein Eng., Des. Sel.* **1992**, *5*, 215–228.
- (8) Aqvist, J.; Medina, C.; Samuelsson, J. New Method for Prediction Ainding Affinity in Computer-aided Drug Design. *Protein Eng., Des. Sel.* **1994**, *7*, 385–391.
- (9) 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*, 889–897.
- (10) Chipot, C. Frontiers in Free Energy Calculations of Biological Systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 71–89.
- (11) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer: Berlin, Heidelberg, 1999; Vol. 4; pp 39–65.
- (12) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224–230.
- (13) Li, M. S.; Mai, B. K. Steered Molecular Dynamics-A Promising Tool for Drug Design. *Curr. Bioinf.* **2012**, *7*, 342–351.
- (14) Mai, B. K.; Viet, M. H.; Li, M. S. Top Leads for Swine Influenza A/H1N1 Virus Revealed by Steered Molecular Dynamics Approach. *J. Chem. Inf. Model.* **2010**, *50*, 2236–2247.
- (15) Mai, B. K.; Li, M. S. Neuraminidase Inhibitor R-125489 - A Promising Drug for Treating Influenza Virus: Steered Molecular Dynamics Approach. *Biochem. Biophys. Res. Commun.* **2011**, *410*, 688–691.
- (16) Marzinek, J. K.; Bond, P. J.; Lian, G.; Zhao, Y.; Han, L.; Noro, M. G.; Pistikopoulos, E. N.; Mantalaris, A. Free Energy Predictions of Ligand Binding to an α -Helix Using Steered Molecular Dynamics and Umbrella Sampling Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 2093–2104.
- (17) Kumar, S.; Li, M. S. Biomolecules under Mechanical Force. *Phys. Rep.* **2010**, *486*, 1–74.
- (18) Medek, P.; Benes, P.; Sochor, J. Computation of Tunnels in Protein Molecules Using Delaunay Triangulation. *Journal WSCG* **2007**, *15* (1–3), 107–114.
- (19) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, e1002708.
- (20) Sehnal, D.; Svobodova Varekova, R.; Berka, K.; Pravda, L.; Navratilova, V.; Banas, P.; Ionescu, C.-M.; Otyepka, M.; Koca, J. MOLE 2.0: Advanced Approach for Analysis of Biomacromolecular Channels. *J. Cheminf.* **2013**, *5*, 39.
- (21) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- (22) Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proc. R. Soc. London, Ser. A* **1924**, *106*, 463–477.
- (23) Hou, T.; Wang, J.; Li, 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*, 69–82.
- (24) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (25) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525.
- (26) Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. Avogadro: an Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4*, 17.
- (27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (28) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- (29) Wang, J.; 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*, 1157–1174.
- (30) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367.
- (31) Grubmuller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271*, 997–999.
- (32) Florin, E.; Moy, V.; Gaub, H. Adhesion Forces between Individual Ligand-Receptor Pairs. *Science* **1994**, *264*, 415–417.
- (33) Moy, V.; Florin, E.; Gaub, H. Intermolecular Forces and Energies between Ligands and Receptors. *Science* **1994**, *266*, 257–259.
- (34) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (35) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (36) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (37) Kim, R.; Skolnick, J. Assessment of Programs for Ligand Binding Affinity Prediction. *J. Comput. Chem.* **2008**, *29*, 1316–1331.

Supporting information for:

A New Method for Navigating Optimal Direction for Pulling Ligand from Binding Pocket: Application to Ranking Binding Affinity by Steered Molecular Dynamics

Quan Van Vuong,^{†,‡} Tin Trung Nguyen,^{†,‡} and Mai Suan Li^{*,†,¶}

Institute for Computational Science and Technology, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam, Contribution equally to the work, and Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland

E-mail: masli@ifpan.edu.pl

Phone: +(48 22) 843 66 01 (3326). Fax: +(48 22) 847 52 23

Supporting Information Available

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

^{*}To whom correspondence should be addressed

[†]Institute for Computational Science and Technology, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam

[‡]Contribution equally to the work

[¶]Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland

Table S1: Chemical structures and protonated states of ligands in the α -thrombin set.

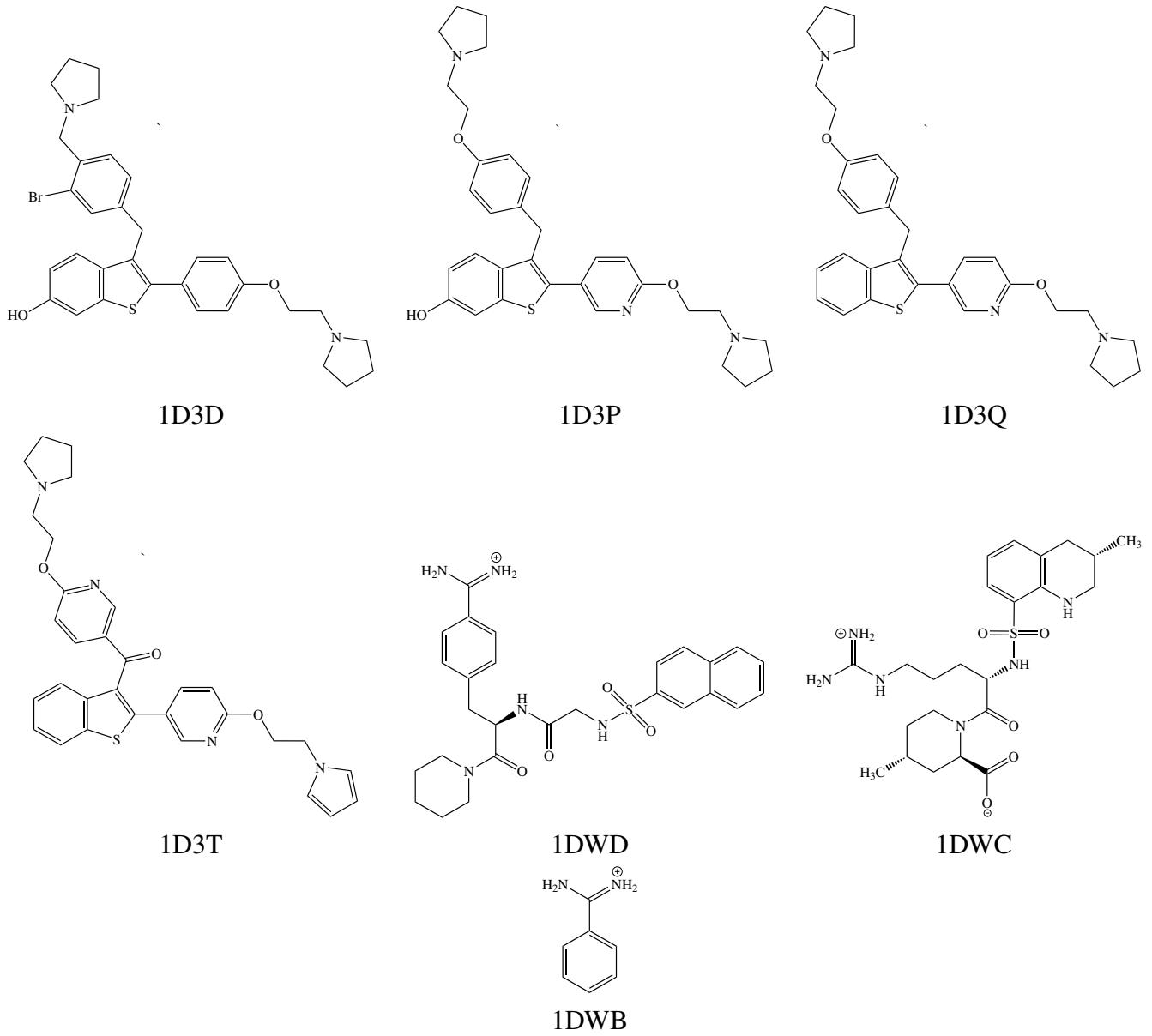


Table S2: Chemical structures and protonated states of ligands in the neuraminidase set.

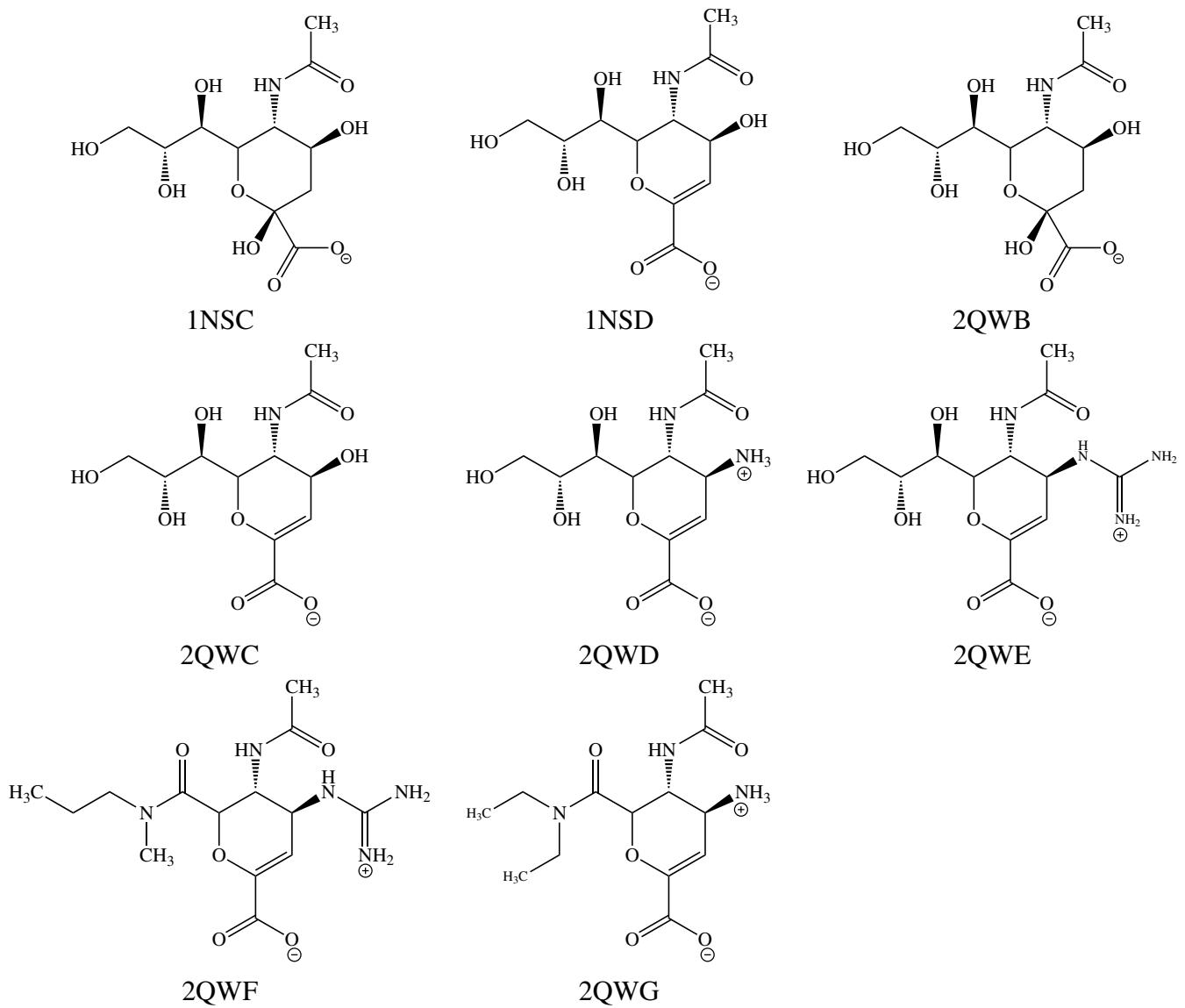
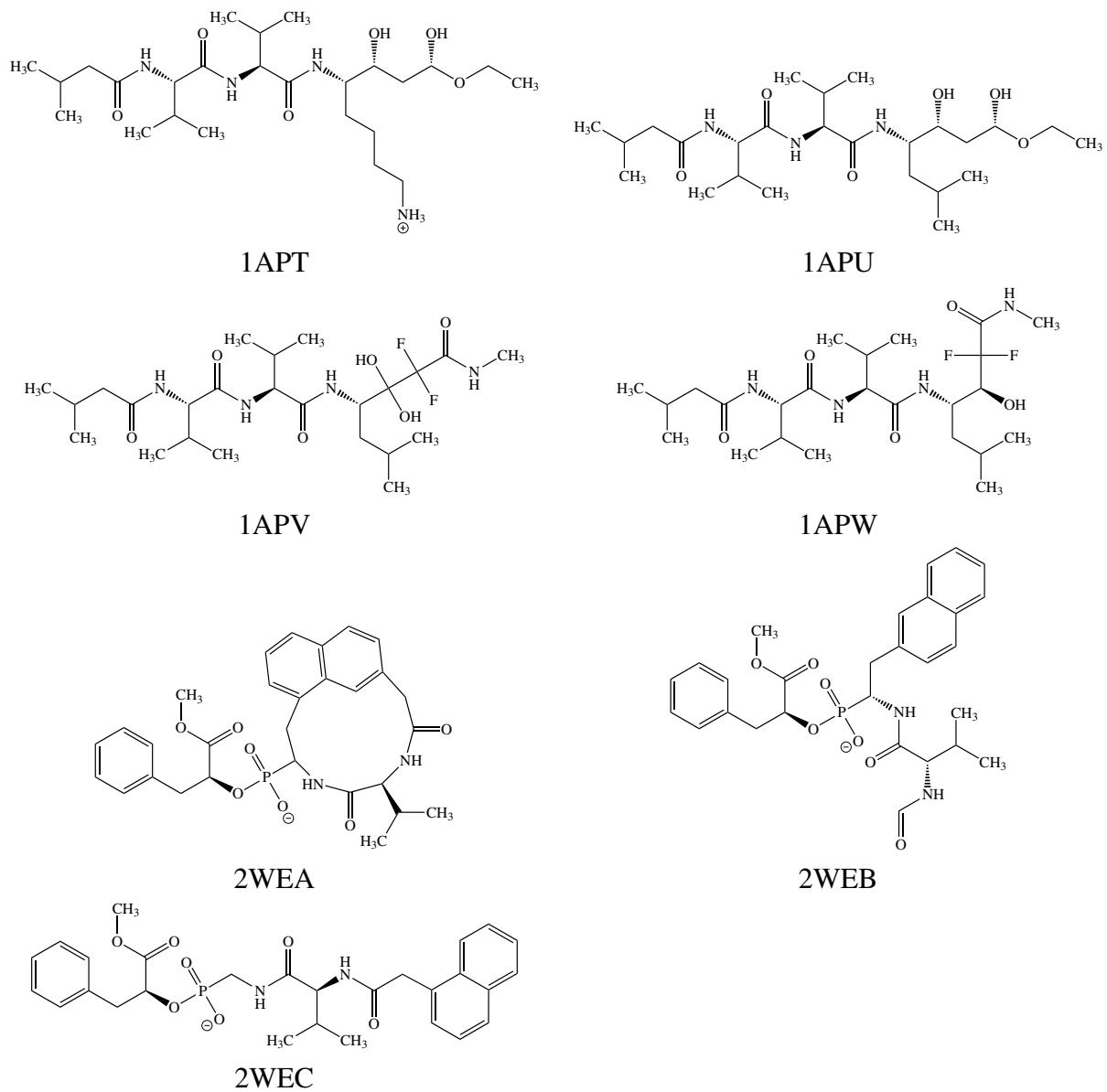


Table S3: Chemical structures and protonated states of ligands in the penicillopepsin set.



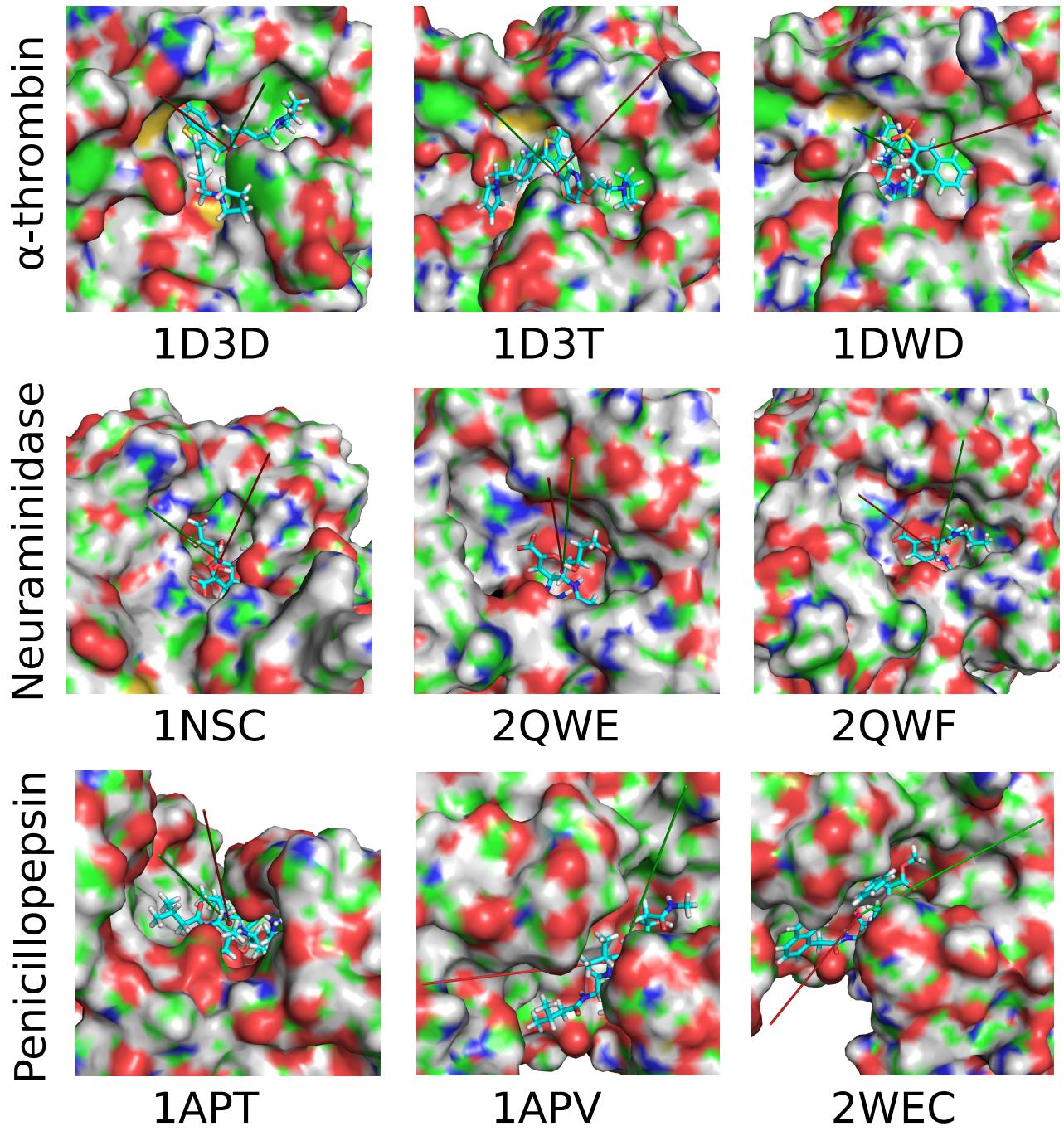


Figure S1: Pulling directions in CAVER and MSH for different systems. The green and red lines refer to pulling direction of MSH and CAVER, respectively.

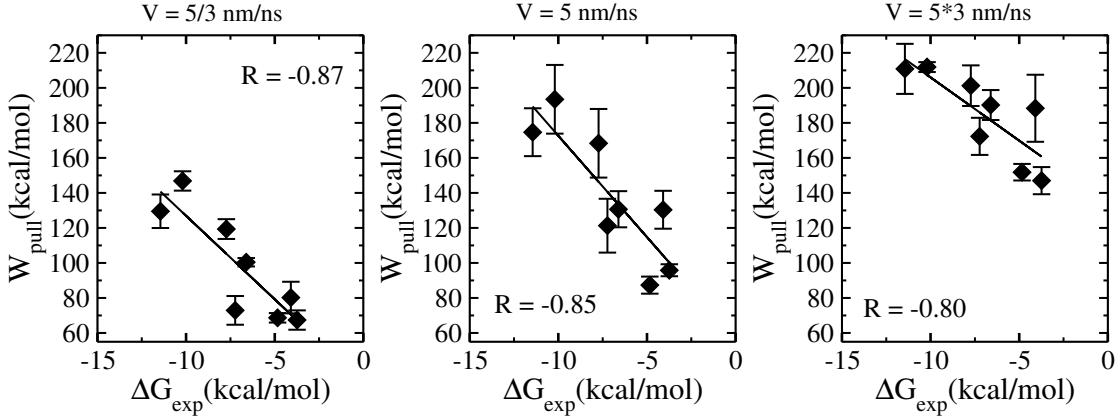


Figure S2: Dependence of correlation between W_{pull} and experimental results on pulling speed v . Results were obtained for the neuraminidase set and spring constant $k = 600 \text{ kJ}/(\text{mol} \cdot \text{nm}^2)$. Error bars come from averaging over 5 independent SMD runs.

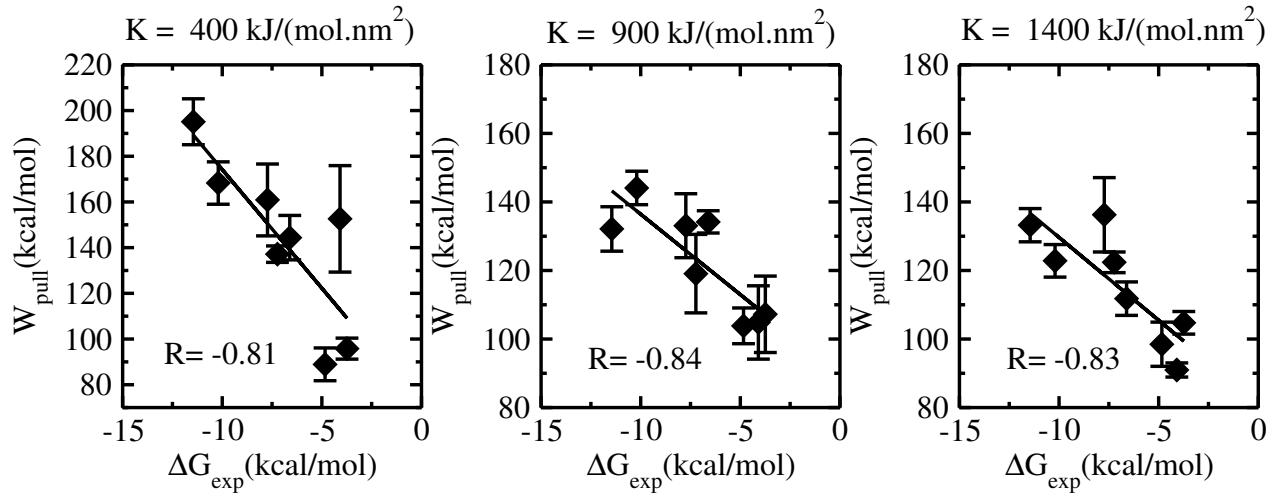


Figure S3: Dependence of correlation between W_{pull} and experimental results on spring constant k . Results were obtained for the neuraminidase set and loading speed $v = 5 \text{ nm/ns}$. Error bars come from averaging over 5 independent SMD runs. $R = -0.81, -0.84$ and -0.83 for $k=400, 900$ and $1400 \text{ kJ}/(\text{mol} \cdot \text{nm}^2)$, respectively.

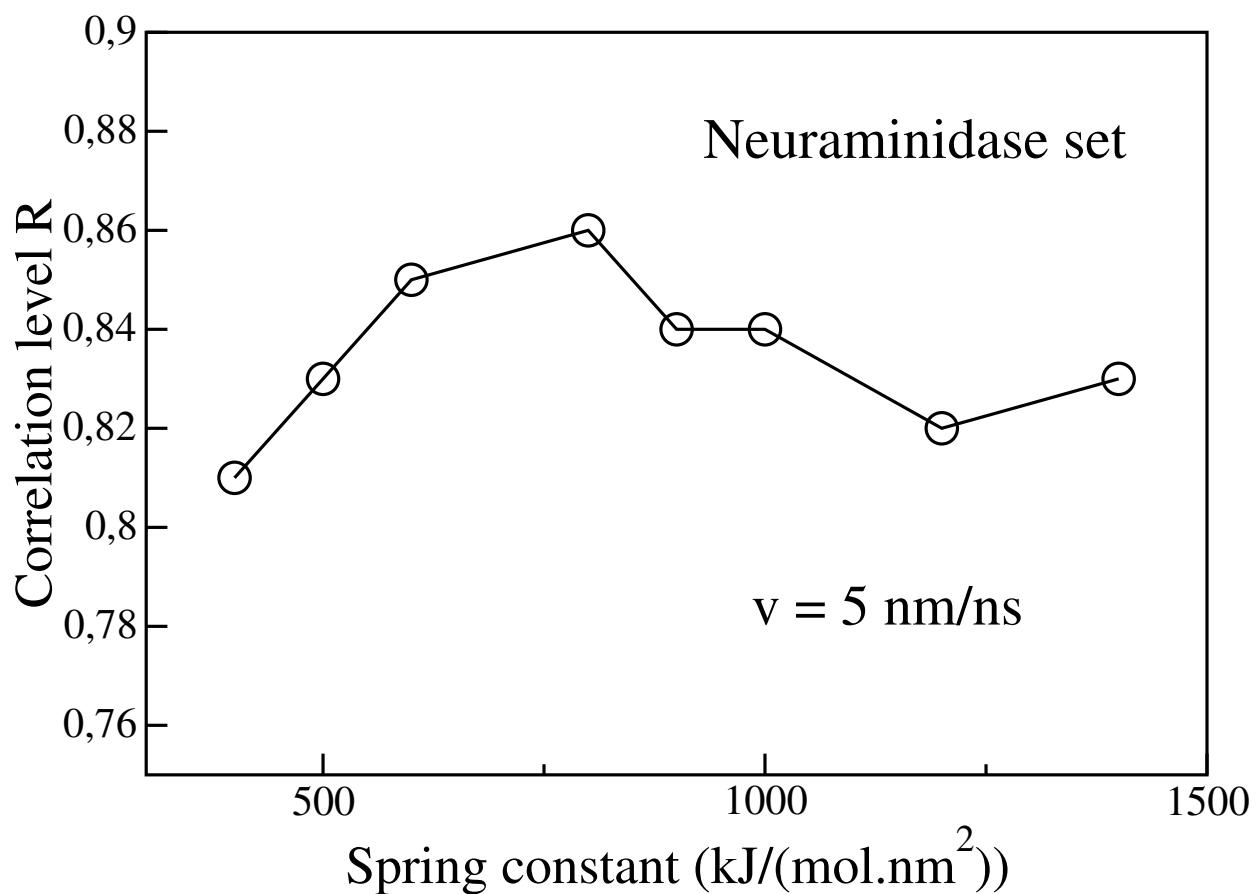


Figure S4: Dependence of correlation level R on spring constant k . Results were obtained for the neuraminidase set and loading speed $v = 5 \text{ nm/ns}$.

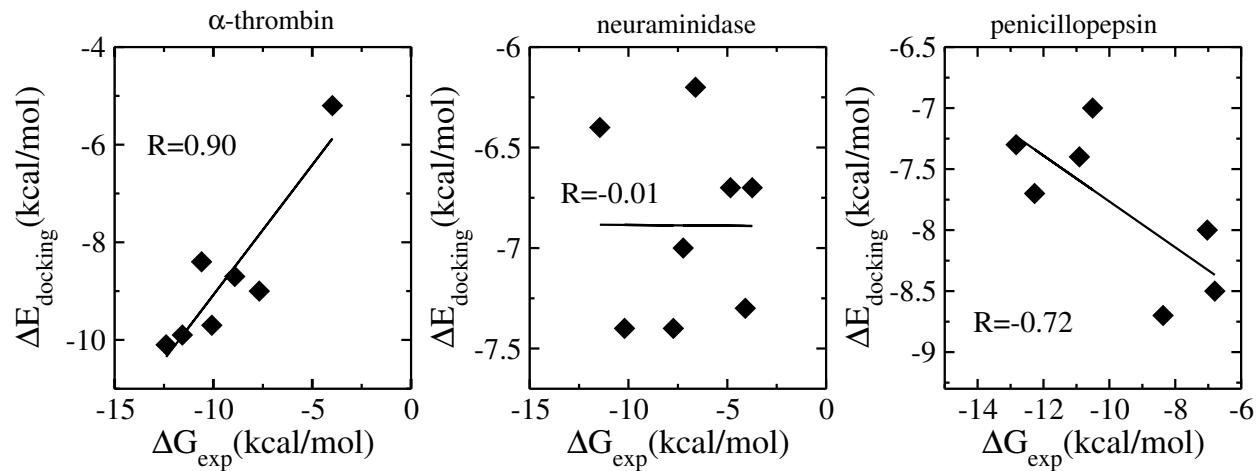


Figure S5: Correlation between the docking binding energies and experimental free energies. $R = 0.90, -0.01$, and -0.72 for the α -thrombin, neuraminidase and penicillopepsin, respectively.