

# Simulation-derived Protein Conformations for Energy Penalties in Molecular Docking

**Anna Sophia Kamenik**

Supervisor UCSF: Prof. Brian Shoichet, Ph. D.

Supervisor UIBK: Prof. DDr. Klaus R. Liedl

The logo for the University of California, San Francisco (UCSF), consisting of the letters 'UCSF' in a bold, blue, sans-serif font.

University of California  
San Francisco



Marshallplan-Jubiläumsstiftung  
Austrian Marshall Plan Foundation





# Contents

<b>I</b>	<b>Localizing Millisecond Dynamics</b>	
<b>1</b>	<b>Introduction</b> .....	<b>7</b>
<b>2</b>	<b>Theoretical Background</b> .....	<b>9</b>
<b>2.1</b>	<b>Accelerated Molecular Dynamics</b>	<b>9</b>
<b>3</b>	<b>Methods</b> .....	<b>10</b>
<b>3.1</b>	<b>Alanine Dipeptide</b>	<b>11</b>
<b>3.2</b>	<b>BPTI</b>	<b>11</b>
<b>3.3</b>	<b>Bet v 1a</b>	<b>12</b>
<b>4</b>	<b>Results</b> .....	<b>13</b>
<b>4.1</b>	<b>Alanine Dipeptide</b>	<b>13</b>
<b>4.2</b>	<b>BPTI</b>	<b>13</b>
<b>4.3</b>	<b>Bet v 1a</b>	<b>15</b>
<b>5</b>	<b>Discussion</b> .....	<b>18</b>
<b>6</b>	<b>Conclusion</b> .....	<b>21</b>
<b>II</b>	<b>Protein Dynamics and Molecular Docking</b>	
<b>7</b>	<b>Introduction</b> .....	<b>25</b>
<b>7.1</b>	<b>Computation and Drug Design</b>	<b>25</b>
<b>7.2</b>	<b>Molecular Docking and Dynamics</b>	<b>25</b>
<b>8</b>	<b>Theoretical Background</b> .....	<b>28</b>
<b>8.1</b>	<b>Docking Step-by-Step</b>	<b>28</b>
<b>8.2</b>	<b>Working with DOCK 3.7</b>	<b>29</b>
<b>8.2.1</b>	<b>Flexible Receptor Docking with DOCK 3.7</b> .....	<b>29</b>

<b>9</b>	<b>Methods</b> .....	<b>30</b>
<b>9.1</b>	<b>Molecular Dynamics Simulations</b>	<b>30</b>
<b>9.2</b>	<b>Conformational Energy Penalties from MD Simulation</b>	<b>30</b>
<b>9.3</b>	<b>Conformational Energy Penalties from Binding Free Energy</b>	<b>32</b>
<b>9.4</b>	<b>Retrospective virtual screening</b>	<b>32</b>
<b>10</b>	<b>Results</b> .....	<b>34</b>
<b>10.1</b>	<b>Conventional Rigid Docking</b>	<b>34</b>
<b>10.2</b>	<b>Penalties from MD Simulation</b>	<b>34</b>
<b>10.3</b>	<b>Penalties from Binding Affinity</b>	<b>35</b>
10.3.1	Flexible Receptor Docking .....	35
10.3.2	Multiple Receptor Docking .....	35
<b>11</b>	<b>Discussion</b> .....	<b>37</b>
<b>12</b>	<b>Conclusion and Outlook</b> .....	<b>38</b>
<b>13</b>	<b>Appendix</b> .....	<b>51</b>
<b>13.1</b>	<b>Calculation of aMD Parameters</b>	<b>51</b>
<b>13.2</b>	<b>Alanine Dipeptide</b>	<b>52</b>
<b>13.3</b>	<b>BPTI</b>	<b>55</b>
<b>13.4</b>	<b>Bet v 1a</b>	<b>58</b>



# Localizing Millisecond Dynamics

<b>1</b>	<b>Introduction</b> .....	<b>7</b>
<b>2</b>	<b>Theoretical Background</b> .....	<b>9</b>
2.1	Accelerated Molecular Dynamics	
<b>3</b>	<b>Methods</b> .....	<b>10</b>
3.1	Alanine Dipeptide	
3.2	BPTI	
3.3	Bet v 1a	
<b>4</b>	<b>Results</b> .....	<b>13</b>
4.1	Alanine Dipeptide	
4.2	BPTI	
4.3	Bet v 1a	
<b>5</b>	<b>Discussion</b> .....	<b>18</b>
<b>6</b>	<b>Conclusion</b> .....	<b>21</b>



# 1 Introduction

Macromolecules in solution steadily undergo conformational changes at room temperature.<sup>1</sup> Various structural studies on diverse model systems have shown that the conformational plasticity of proteins plays a key role in molecular mechanisms such as catalytic activity,<sup>2</sup> bio-molecular recognition,<sup>3–5</sup> and allosteric regulation.<sup>6</sup> Thus, characterizing dynamics of biological macromolecules is crucial to understand their biological activity and function.<sup>7,8</sup> Molecular dynamics (MD) simulations have proven to be an efficient tool to capture the flexibility of macromolecules.<sup>9</sup> Yet, state-of-the-art MD simulations routinely capture dynamics on the nanosecond to microsecond time scale, while many biologically significant motions appear on the millisecond time scale or slower.<sup>10</sup> Inherent limitations in conformational sampling can either be overcome by usage of dedicated simulation hardware<sup>11</sup> or by application of enhanced sampling algorithms.<sup>12,13</sup> Accelerated MD (aMD) is a promising enhanced sampling technique, which improves the efficiency of conventional MD (cMD) simulation without *a priori* knowledge of the potential energy surface.<sup>14</sup> Introduction of a continuous, non-negative bias potential increases escape rates from local energy basins. Thus, the conformational space is sampled more extensively at negligible computational overhead costs.<sup>15,16</sup> Subsequently the original energy landscape can be reconstructed by Boltzmann reweighting.<sup>17–19</sup>

The versatile applicability of aMD simulations has repeatedly been proven on manifold macromolecular systems.<sup>20–25</sup> Current aMD studies predominantly focus on analyzing the global dynamics of the obtained conformational ensembles.<sup>26–28</sup> Yet, for a comprehensive understanding of biomolecular properties it is crucial to be able to localize flexibility in specific protein domains.<sup>29–33</sup>

Current approaches estimating local dynamics in aMD simulations are limited to expensive large-scale calculations of amide order parameters from multiple aMD trajectories on various acceleration levels.<sup>34,35</sup> Rather than approximations of NMR observables a straight forward approach based on the thermodynamics of a system would be desirable. So far no metric is available to directly quantify local flexibility from aMD trajectories based on the captured thermodynamics of the system. We propose residue-wise dihedral entropy as first methodology to efficiently characterize local dynamics of macromolecules from aMD simulations.<sup>36,37</sup> To confirm the validity and efficiency of our approach we apply the metric to the model systems alanine dipeptide (DiAla) and bovine pancreatic trypsin inhibitor (BPTI). The conformational dynamics of BPTI have been investigated thoroughly in NMR experiments<sup>38–41</sup> as well as in large-scale computer simulations.<sup>11</sup> It has already been demonstrated in previous studies that metrics for local flexibility in a 1 ms cMD simulation of BPTI track the characteristic motions of BPTI, known from NMR and global flexibility studies.<sup>29,38–42</sup>

Here we show that residue-wise dihedral entropies deriving from a 500 ns aMD and a 1 ms cMD simulation of BPTI correlate remarkably. Application of our metric on aMD trajectories provides a possibility to track low-frequency local dynamics on the millisecond time scale.

Additionally we apply our metric to cMD and aMD simulations of the major birch pollen allergen Bet v 1.0101 (Bet v 1a). Bet v 1a is a highly immunogenic storage protein and most prominent for causing seasonal pollen allergy in the northern hemisphere.<sup>43,44</sup> Despite a sequence similarity of more than 95%<sup>45,46</sup> and minor differences in their 3D structures the more than 13 reported isoforms of Bet v 1a vary strongly in their immunogenicity.<sup>47</sup> Investigations on differences in proteolytic stability and ligand binding of different isoforms and mutants of Bet v 1a suggest a linkage between immunogenic potential and conformational flexibility.<sup>47-49</sup>

The accuracy of our results is underlined by comparison to NMR data,<sup>47</sup> displaying analogous trends in experimentally and computationally estimated flexibility. Our results show that dihedral entropies from aMD simulations are an efficient tool to describe local protein dynamics on the millisecond time scale.

## 2 Theoretical Background

### 2.1 Accelerated Molecular Dynamics

In aMD a boosting potential  $\Delta V(\vec{r})$  is added to the original potential energy  $V(\vec{r})$  to increase the sampling efficiency. Decreased barriers between local minima enhance the escape rate for the biased potential  $V^*(\vec{r})$ .<sup>14</sup>

**Definition 2.1.1 – The biased potential.**

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}) \quad (2.1)$$

A threshold energy  $E$  is defined, which determines the biased potential.

$$V^*(\vec{r}) = \begin{cases} V(\vec{r}), & V(\vec{r}) \geq E, \\ V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}), & V(\vec{r}) < E. \end{cases} \quad (2.2)$$

High-energy conformations, with energies greater than the threshold  $E$ , are sampled in the conventional way. If an energy is lower than  $E$  a boosting potential  $\Delta V(\vec{r})$  is added.

**Definition 2.1.2 – The boosting potential.**

$$\Delta V = \frac{(E - V(\vec{r}))^2}{\alpha + E - V(\vec{r})} \quad (2.3)$$

$\alpha$  is the so called acceleration factor or tuning parameter. The more  $\alpha$  approaches zero, the more it enforces a rigorous acceleration of the simulation. Considering equation 2.3 the modified forces acting on every atom  $i$ , for example, are given by

$$F_i^* = -\frac{d}{d(\vec{r})}[V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r})] = F_i \times \left[ \frac{\alpha^2}{(\alpha + E - V(\vec{r}))^2} \right] \quad (2.4)$$

The computational effort is comparable to conventional MD simulations. Yet the parameters  $E$  and  $\alpha$  have to be evaluated individually for each investigated system. The original potential energy surface can be reconstructed by reweighting the trajectory with the Boltzmann-factor.<sup>15</sup>

### 3 Methods

MD simulations were performed with the AMBER14 simulation package.<sup>50</sup> All shown structures were prepared in MOE (Molecular Operating Environment, Chemical Computing Group, version 2014.0901)<sup>51</sup> using the protonate3D tool.<sup>52</sup> With leap of the AmberTools15<sup>50</sup> package all three systems were soaked into a truncated octahedral solvent box of TIP3P water molecules.<sup>53</sup> For Di-Ala the minimum wall distance was set to 12 Å, for BPTI and Bet v 1a to 10 Å. Parameters for all three systems derive from the AMBER force field 99SB-ILDN.<sup>54</sup> All systems were carefully equilibrated using a multi step equilibration protocol.<sup>55</sup> Precedent cMD simulations as well as all aMD simulations were performed in NpT ensemble using pmemd.cuda.<sup>56</sup> Bonds involving hydrogen atoms were restrained by applying the SHAKE algorithm,<sup>57</sup> allowing a time step of 2.0 fs. Atmospheric pressure of the system was preserved by weak coupling to an external bath using the Berendsen algorithm.<sup>58</sup> The Langevin thermostat<sup>59</sup> was used to maintain the temperature during simulations at 300 K for Di-Ala and BPTI and 310 K for Bet v 1a (human body temperature). All shown aMD simulations were performed using the dual-boost protocol<sup>60</sup> implemented in pmemd.cuda.<sup>56</sup> Thereby the total potential is accelerated and an extra boosting is applied to the dihedral potential. It has been shown that dual-boost aMD simulations sample the diffusive solvent motions more extensively. Ensemble averages as well as entropy estimates converge faster than in dihedral-boost aMD simulations.<sup>26,60</sup> All simulations were analyzed using cpptraj<sup>61</sup> in AmberTools15,<sup>50</sup> the reweighting protocol provided by Miao *et al.*<sup>18</sup> and in-house scripts. The free energy profile was reconstructed from the aMD simulations via Boltzmann reweighting using a Maclaurin series expansion (up to the 10th order) as approximation for the exponential term, as suggested in previous studies.<sup>15</sup>

The local backbone flexibility profiles were estimated from the resulting reweighted one-dimensional free energy profiles, state populations respectively, of the backbone dihedrals  $\Phi$  and  $\Psi$ . The entropy is calculated by integration of the reweighted state populations of a given dihedral. A high entropy of a residue backbone dihedral indicates high local backbone flexibility.<sup>29,36</sup>

In the presented work we prioritized dihedral entropies  $S_\Psi$  over  $S_\Phi$  in the representation protein dynamics as it captures the backbone dynamics more comprehensively.<sup>62,63</sup> Yet,  $S_\Psi$  alone does not reflect the entire backbone dynamics and dihedral entropies  $S_\Phi$  were calculated as well (see appendix).

### 3.1 Alanine Dipeptide

For the reference cMD simulation of alanine dipeptide (Di-Ala) a 10  $\mu$ s trajectory comprising 100 000 frames, previously performed in our group was reanalyzed.<sup>29</sup> Residue-wise dihedral entropies of the reference cMD simulations were calculated from probability density functions reconstructed by non-parametric kernel density estimation.<sup>36,37</sup> As proposed by Botev *et al.*<sup>64</sup> we optimize the bandwidth of the kernel function by cross validation, resulting in a continuous probability density function for each dihedral. We periodically duplicate our data to minimize the overestimated flexibility at the boundaries. The entropy for each dihedral  $S_x$  was calculated by integration of  $k_B \cdot p(x) \cdot \log(p(x))$  on its probability density function  $p(x)$  as described in Huber *et al.*<sup>36,37</sup> The standard deviations were calculated by splitting the trajectory into 100 segments.

For the aMD calculations the solvated and equilibrated Di-Ala system was simulated for 1  $\mu$ s and stored as 500 000 equal-spaced snapshots (2 fs spacing). The aMD boosting parameters were calculated as suggested in previous studies<sup>15</sup> (see appendix for further information). To minimize the statistical noise (and capture dynamics corresponding to 10  $\mu$ s of cMD) the resulting trajectory was divided into 200 segments of 5 ns each (2 500 frames) using the segments for averaging (appendix Figure 13.4). From the reweighted dihedral populations a Ramachadran plot was created to assure sufficient and accurate sampling of conformational space (appendix Figure 13.1).

To compare aMD and cMD free energies we calculated the free energy of the cMD trajectory using the reweighting protocol by Miao *et al.*<sup>18</sup> We investigated several different bin sizes and the jaggedness of the cMD profile disappears only using a bin size above 20° (appendix Figure 13.2). However, as also suggested by Miao *et al.*<sup>18</sup> we also observe a bin size of 6° to be a reasonable compromise between accuracy and statistical noise.

### 3.2 BPTI

D. E. Shaw Research kindly provided us with a long time scale 1.03 ms trajectory comprising 4 140 000 snapshots as a cMD reference.<sup>11</sup> We used the same joint neutron/X-ray refined structure of BPTI (PDB: 5PTI)<sup>38</sup> as D. E. Shaw *et al.* as starting structure for our large scale cMD and aMD simulations. A 500 ns aMD simulation was performed and the trajectory stored in 250 000 snapshots. The aMD boosting parameters were calculated as suggested by Pierce *et al.*<sup>15</sup> and can be found in the appendix. To localize the observed dynamic hot-spots we calculated dihedral entropies as described earlier for 1 ms and 1  $\mu$ s of cMD and 500 ns of aMD sampling time.

### 3.3 Bet v 1a

Based on the crystal structure of Bet v 1a with PDB ID 4A88<sup>65</sup> we sampled 3  $\mu$ s of cMD simulations on Bet v 1a. Suitable aMD boosting parameters for Bet v 1a were determined by a systematic search (see appendix). We performed a 1  $\mu$ s aMD simulation stored as 100 000 equally spaced snapshots. The trajectory was split in 50 segments of 20 ns each (2 000 snapshots) and reweighted as described above. Subsequently dihedral entropies were calculated and averaged over all 50 segments. NMR order parameters were kindly provided by Grutsch *et al.*, the experimental setup has been described elsewhere.<sup>47</sup>

## 4 Results

### 4.1 Alanine Dipeptide

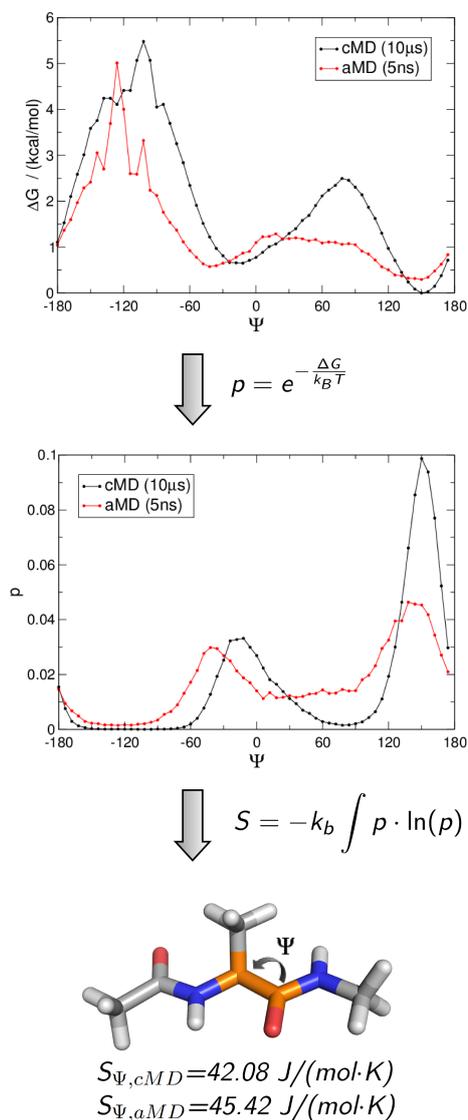
To establish our approach we calculated backbone dihedral entropies for  $\Phi$  and  $\Psi$  of alanine dipeptide from 5 ns aMD sampling and used a 10  $\mu$ s cMD simulation as a reference (Figure 4.1). We find a significant reproduction of local minima and overall shape of the potential energy surface for the backbone dihedral  $\Phi$  of Di-Ala. As reported before<sup>18</sup> we also observe a shift in the extrema of backbone dihedral  $\Psi$  to smaller  $\Psi$ -values in aMD simulations. Exemplary, in the cMD simulation an energy minimum is found at  $-24^\circ$ , while the corresponding minimum in the aMD results is recovered at  $-42^\circ$ . Still the overall energy landscape of the reweighted aMD trajectory is in reliable agreement with the cMD results.

Considering these errors, the resulting dihedral entropies,  $S_{\Phi,aMD}=41.03 (\pm 2.95)$  J/(mol·K) and  $S_{\Psi,aMD}=45.42 (\pm 3.25)$  J/(mol·K), agree very well with those of the cMD ensemble  $S_{\Phi,cMD}=40.20 (\pm 0.60)$  J/(mol·K) and  $S_{\Psi,cMD}=42.08 (\pm 0.39)$  J/(mol·K).

### 4.2 BPTI

To benchmark our local flexibility metric we analyzed a 500 ns aMD simulation of BPTI and compared it to a 1 ms cMD simulation provided by D. E. Shaw Research<sup>11</sup> (Figure 4.2). It has been demonstrated previously that 500 ns of BPTI aMD simulation cover a conformational space equivalent to a 1 ms cMD.<sup>15</sup> In addition to the findings of Pierce *et al.* we observe equal assessment of local backbone flexibility for the 500 ns aMD and 1 ms cMD simulation. In the 1 ms reference cMD maxima of  $S_{\Psi}$  are found in regions from residues 10–20, 32–44 and the C-terminal residues from 54–58. Each flexibility hot spot is reproduced in  $S_{\Psi}$  from 500 ns of aMD sampling. The high similarity of local flexibility in both simulation protocols is reflected in a Spearman rank correlation over the protein length of  $r=0.93$  for  $S_{\Psi}$  between cMD and aMD. Comparable agreement between the aMD and reference cMD simulations is also found for  $S_{\Phi}$  (see SI Figure 13.3). In contrast, when looking at dihedral entropies from a 1  $\mu$ s cMD simulation notable differences are found for  $S_{\Psi}$  in the regions of residues 10–14 and 32–44. In 1  $\mu$ s of cMD sampling no elevation in conformational plasticity is captured in these domains, though this is clearly observed on the millisecond time scale as well as in the aMD-derived ensemble. The deviation of the local dynamics pattern results in a Spearman rank correlation of  $r=0.65$  for  $S_{\Psi}$  in 500 ns of aMD and 1  $\mu$ s cMD sampling.

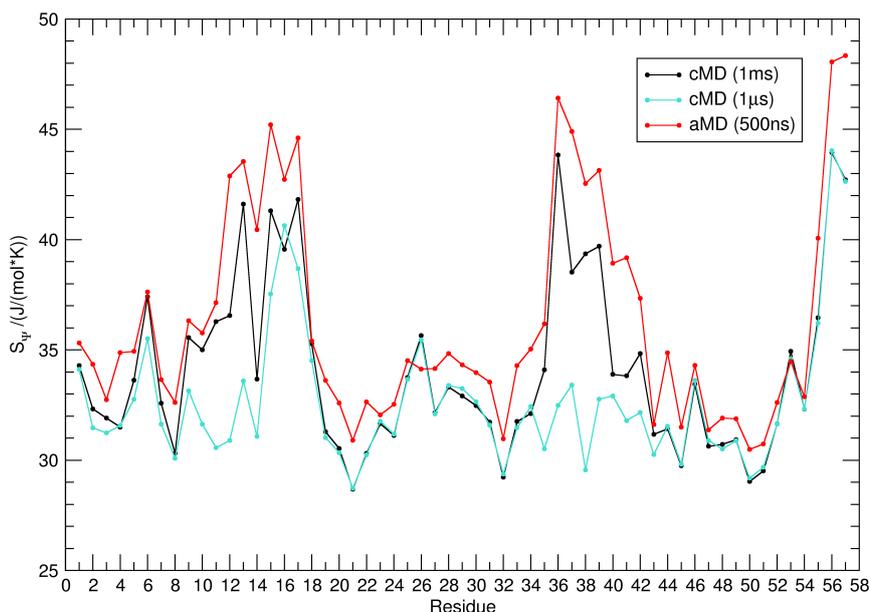
For the dihedral entropies  $S_{\Psi}$  and  $S_{\Phi}$  captured in 500 ns of aMD we find a correlation of  $r=0.77$ . The high correlation of  $S_{\Psi}$  and  $S_{\Phi}$  supports the assumption that a similar extent of flexibility is captured in both backbone dihedral angles. In Figure 4.3 the flexibility hot spots displayed as



**Figure 4.1:** From a reweighted aMD trajectory to dihedral entropies of alanine dipeptide: State populations  $p$  are calculated from free energy profiles of the backbone dihedral  $\Psi$  of a  $10 \mu\text{s}$  cMD (black) and a reweighted 5 ns aMD (red) trajectory. Integration of the resulting state populations leads to the dihedral entropy  $S_{\Psi,cMD} = 42.08 \text{ J}/(\text{mol}\cdot\text{K})$  and  $S_{\Psi,aMD} = 45.42 \text{ J}/(\text{mol}\cdot\text{K})$ .

graph in Figure 4.2 are color coded and projected on the BPTI fold. The differences in flexibility captured in 500 ns of aMD (B) and  $1 \mu\text{s}$  cMD (C) are highlighted in (D).  $S_{\Psi}$  of the  $1 \mu\text{s}$  cMD simulation were subtracted from  $S_{\Psi}$  resulting from 500 ns aMD. Positive values (blue) indicate domains, which are more flexible in the aMD than in the cMD simulation of the same sampling

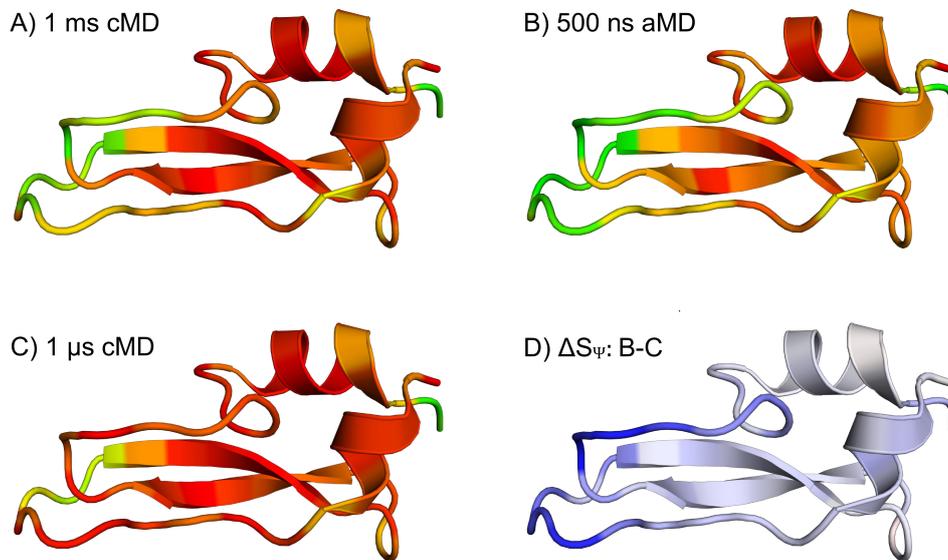
length. When looking at the structure the most prominent deviation between 500 ns of aMD and 1  $\mu$ s cMD simulation lies in the local flexibility of the loop regions. Clearly, the dynamic nature of the loop involving residues 32–44 found in the millisecond simulation is not adequately sampled in 1  $\mu$ s cMD, but accurately represented in aMD sampling of the same length.



**Figure 4.2:** Comparing local flexibility of BPTI captured in cMD and aMD simulations. Residue-wise dihedral entropies  $S_{\Psi}$  from a 1 ms cMD simulation (black) and 500 ns aMD simulation of BPTI (red) show remarkable rank correlation ( $r=0.93$ ). Local flexibility observed in a 1  $\mu$ s cMD simulation (turquoise) clearly differs from the aMD results ( $r=0.65$ ).

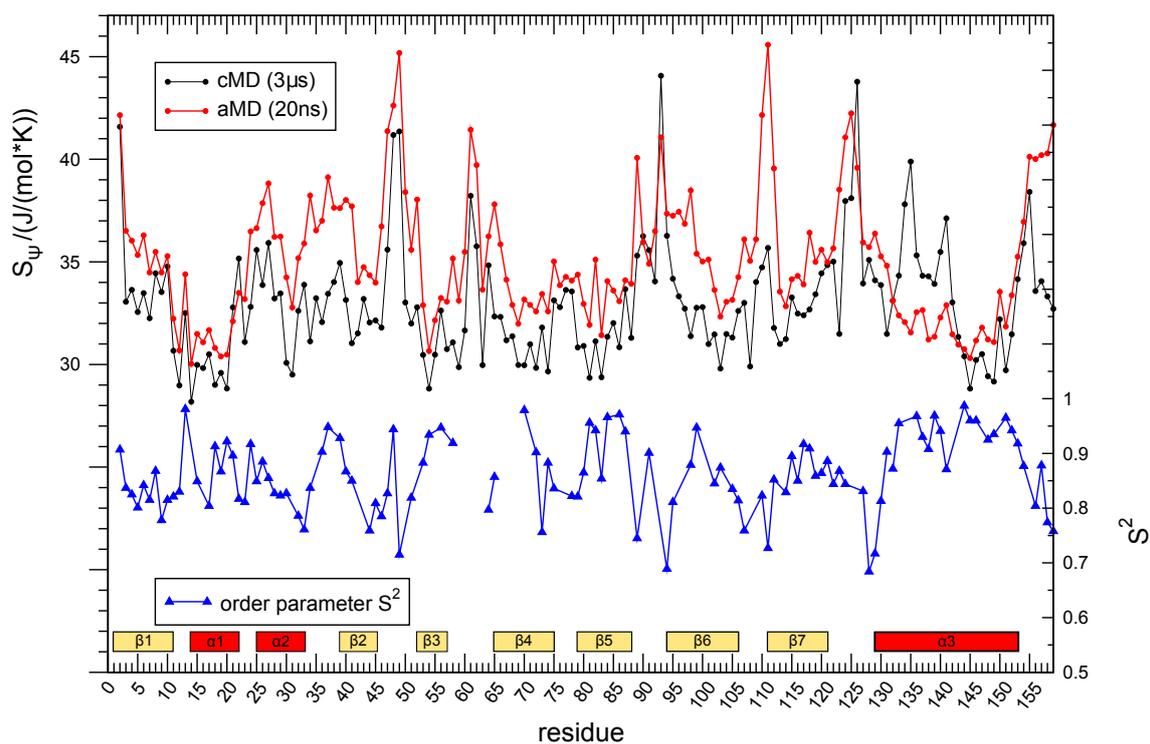
### 4.3 Bet v 1a

With 159 residues the major birch pollen allergen Bet v 1a is the largest system in our study. A 1  $\mu$ s aMD simulation was split into 50 segments, 20 ns each. A 3  $\mu$ s cMD simulation and order parameters  $S^2$  from backbone amide NMR relaxation experiments<sup>47</sup> act as references for our metric to quantify local motions in aMD simulations (Figure 4.4). Order parameters range from zero to one, indicating no or full constriction of internal mobility of backbone amide groups on the ps- to ns-timescale.<sup>66</sup> Thus, an anti-correlation is expected since high entropy indicates lower order. As observed for the BPTI system, the aMD-derived ensemble shows higher backbone flexibility



**Figure 4.3:** Flexibility hot spots of BPTI: Residue-wise dihedral entropies for  $\Psi$  ( $S_{\Psi}$ ) from the 1 ms cMD (A), 500 ns aMD simulation (B) and 1  $\mu$ s cMD (C) are projected on the structure of BPTI (PDB-ID: 5PTI). In Part A, B and C the color coding ranges from red ( $S_{\Psi} \leq 30$  J/(mol·K)) via yellow, to green ( $S_{\Psi} \geq 45$  J/(mol·K)). Thus, the most rigid residues are pictured in red, whereas the most flexible ones are colored green. Part D shows the differences in  $S_{\Psi}$  between 500 ns aMD and 1  $\mu$ s cMD ( $\Delta S_{\Psi} = B - C$ ). The color coding in D ranges from red ( $\Delta S_{\Psi} \leq 15$ ) to white, to blue ( $\Delta S_{\Psi} \geq 15$  J/(mol·K)), i.e., blue indicates regions where the aMD simulation captures a higher local flexibility. Thus, the cMD simulation clearly underestimates the conformational dynamics of BPTI in the region of residues 10–14 and 32–44.

for all residues compared to the cMD-derived one. The general flexibility patterns are conserved for most parts of the protein in both simulations. However, especially in the region from  $\alpha 1$  to  $\beta 2$  (residues 15–45) enhanced dynamics are visible in aMD, which are not reflected in the cMD simulation. Order parameters  $S^2$  show the expected anti-correlation with dihedral entropies of the core domains, i.e.,  $\alpha 3$ -helix and  $\beta$ -sheets 4–7 (residues 70–159). Yet in contrast to experimental findings we obtain lowered local dynamics for the  $\alpha 1$ -helix and the  $\beta 2$ -sheet, while for the loop region in between elevated flexibility is observed. These opposing qualitative observations are reflected in a Spearman rank correlation between the  $\Psi$  dihedral entropies from aMD simulations and amide order parameters  $S^2$  of  $r = -0.35$  for the whole fold. When restricting the correlation analysis to the core helix  $\alpha 3$  and  $\beta$ -sheets 5–7 (residues 70–159) the rank correlation is strengthened to  $r = -0.61$ .



**Figure 4.4:** Local flexibility on different time scales in Bet v 1a: Dihedral entropies  $S_{\psi}$  from 20 ns aMD (red) and 3  $\mu\text{s}$  of cMD (black) are compared to experimental ps/ns dynamics<sup>66</sup> captured by NMR derived backbone amide order parameters  $S^2$  (blue).<sup>47</sup>

## 5 Discussion

Local flexibility is decisive for biomolecular recognition mechanisms like protein-protein interactions and ligand binding, as well as for protein folding.<sup>2-4,6</sup> It has been shown that the flexibility patterns of Bet v 1 are linked to its fold-stability<sup>49</sup> and allergenicity.<sup>47</sup> Furthermore, studies on the dynamics of proteases found a remarkable correlation between substrate specificity and local flexibility of protease active sites.<sup>5</sup> The associated dynamics include motions from the nanosecond to the millisecond time scale.<sup>10</sup> Metrics to quantify the amount of global and local motions are indispensable for a holistic understanding of macromolecular interactions.<sup>29</sup> Several expedient metrics have been developed to account for global and local flexibility in cMD simulations, such as root-mean-square fluctuation, locally and globally aligned b-factors, or torsional entropies.<sup>29,36,67</sup> Dynamics from aMD simulations are currently predominantly described on a global level only.<sup>26-28,68</sup> In our study on three model systems of increasing size, we establish an alignment-free internal coordinate-based metric estimating local flexibility in aMD simulations. As expected, enhanced local dynamics are captured using enhanced sampling. As demonstrated for global movements, we are able to describe local protein flexibility on the millisecond time scale after several hundred of nanoseconds of aMD sampling. The residue-wise quantification of motion in a protein backbone is constructed from torsional free energy profiles of reweighted aMD trajectories via calculation of dihedral entropies.<sup>36</sup>

Investigations on the smallest system in our study, Di-Ala, illustrate the applied work flow. As already outlined in a previous study<sup>18</sup> we also find a shift of minima in the free energy landscape of  $\Psi$  in the reweighted aMD trajectory compared to the cMD results. This deviation is most probably generated by the reweighting step when using Maclaurin series of the 10th order as approximation for the exponential. As shown by Miao *et al.* for alanine dipeptide, reweighting using cumulant expansion to the second order reconstructs the free energy surface of aMD simulations accurately. Yet, this approach requires the distribution of the boost potential to be exactly Gaussian.<sup>69</sup> This may be approximately the case for a small system like Di-Ala, but we observed that it is less successful for larger biomolecular systems such as BPTI (appendix Figure 13.7). Previous studies on BPTI and other proteins showed accurate results for Maclaurin series expansion of the 10th order.<sup>15,27</sup> Thus we prefer to use Maclaurin series expansion for all systems in our study for consistency.

Overall, a quantitative reproduction of the positions of the local extrema is not essential for our methodology, since their exact location has no influence on the resulting entropies. The state probability distribution is generally broader in the aMD ensemble, but results in statistically equal dihedral entropies. Thus, the same dynamic tendencies are estimated from the aMD and reference cMD simulation.

BPTI is widely used as a test system for NMR and protein dynamics in general and has been investigated thoroughly over the last decades.<sup>41</sup> The group of D.E. Shaw performed large scale computer simulations and extensive studies on the dynamics of this model system.<sup>11</sup> We have demonstrated in previous studies that metrics of local flexibility in a 1 ms cMD simulation of BPTI capture characteristic motions, known from NMR and global flexibility studies.<sup>29</sup> These prominent movement motifs comprise the isomerization of a disulfide bridge involving Cys-14 and Cys-38. With our metric we quantified local flexibility for a 500 ns aMD and a 1 ms cMD simulation of BPTI with a striking Spearman rank correlation of  $r=0.93$  for  $S_{\psi}$ . Thus, we are able to quantify and localize millisecond dynamics with aMD simulations of several hundred nanosecond length. When compared to a 1  $\mu$ s cMD simulation it is evident that these results cannot be obtained with state-of-the-art simulation protocols of the same length ( $r=0.65$ ). The differences in captured molecular motions in 500 ns aMD and 1  $\mu$ s cMD sampling can be traced back to the loop regions (residue 10–14 and 32–44), which comprise the switching disulfide bridge mentioned earlier (see also appendix Figure 13.8). The isomerization of this bond, which is described by NMR studies and the calculations of the Shaw group, implies an enhancement of local flexibility in the surrounding region. The characteristic dynamics of the domain are evidently quantified by our metric in 500 ns aMD sampling. These results show that the approach allows access to dynamics of low-frequency motions. Additionally, our metric provides a tool to localize the origin of elevated flexibility thereby identifying domains with prominent dynamics and allowing a residue-wise interpretation.

Bet v 1a is the largest and thereby most challenging system in our study. The boosting parameters for the Di-Ala and BPTI simulations were applied as suggested in previous studies.<sup>15,18</sup> For Bet v 1a we set up 6 simulations on different levels of acceleration to test the limit of applied acceleration without unfolding the protein (see appendix). It has been shown that the choice of parameter has crucial impact on the resulting trajectory and has to be evaluated carefully.<sup>16</sup>

To estimate the reliability of our findings we compared a 20 ns aMD trajectory to a conventional 3  $\mu$ s MD simulation and NMR-derived NH order parameters  $S^2$ . Flexibility patterns recovered from aMD and cMD simulation are overall in good agreement. Characteristic deviations are observed in the region from helix  $\alpha 1$  to the sheet  $\beta 2$  (residues 15–45). These observations are reflected by a Spearman rank correlation of  $r=0.51$  between the aMD and cMD simulation. We hypothesize that this rather low correlation can primarily be explained by the different time scales captured. Extrapolating from previous studies on BPTI, where 500ns of aMD sampling cover the dynamics of 1 ms cMD sampling, 20 ns of aMD should correspond to dynamics of around 40  $\mu$ s. It can be assumed that these flexibilities clearly deviate from motions captured in only 3  $\mu$ s sampling time. Comparison of the dihedral entropy profile to order parameters  $S^2$  of the backbone amide<sup>66</sup> leads to similar findings. NMR order parameters  $S^2$  are sensitive to ps- to ns-dynamics, capturing much

faster than motions the shown aMD data. We expect a coupling between slow backbone dynamics, profiled by aMD simulation, and fast motions, captured in NMR data.<sup>29</sup> Thus, NMR order parameters  $S^2$  represent a method to experimentally probe protein backbone dynamics on residue resolution and an insightful reference to estimate the reliability of our approach. As expected we observe an anti-correlation between the calculated dihedral entropies and the experimental order parameters. Again reasonable agreement is visible for the region reaching from residue 70 to the C-terminus, while for the domain from helix  $\alpha 1$  to the sheet  $\beta 2$  (residues 15–45) almost opposing trends are found. These qualitative observations become apparent in a Spearman rank correlation of  $r=-0.39$  between the aMD dihedral entropies and  $S^2$  when considering the whole protein. This is only a slight improvement over cMD simulations, where correlations of  $r=-0.23$  for torsional entropies and order parameters are obtained. This might result from dynamics captured by aMD being beyond the scope of the NMR timescale. It has been shown in previous studies that the region from  $\alpha 1$  to  $\beta 2$  undergoes a noticeable rigidification upon ligand binding.<sup>47</sup> Order parameters and relaxation dispersion profiles of the apo protein confirm the flexible nature of Bet v 1a on a pico- to nanosecond as well as on a micro- to millisecond timescale. Residues from  $\alpha 1$  to  $\beta 2$  show elevated dynamics in both experiments. For the remaining parts of the protein (residue 70–159) a correlation of  $r=-0.61$  is found between the aMD dihedral entropies and  $S^2$ . Here the correlation of cMD simulations and order parameters is still notably lower with  $r=-0.35$ .

Additionally dihedral entropies were calculated from aMD simulations of varying length ranging from 10 ns to 1  $\mu$ s (see appendix Figure 13.10). Again, the resulting entropies show similar flexibility profiles as experimental NMR studies, with exception of the discussed domain reaching from  $\alpha 1$  to  $\beta 2$ . This emphasizes presence of complex conformational dynamics on multiple time scales in this area.

With the presented metric we provide a tool to map low-frequency conformational dynamics of biomolecules at residue level. With increasing system size reproduction of the original flexibility profile becomes more challenging. The decorrelation time of aMD and cMD data has been investigated extensively in previous studies.<sup>16</sup> It has been shown that the aMD generally reduces the statistical inefficiency of a simulation. An extensive probing of the acceleration parameter is crucial for the reliability of any aMD trajectory. Aggressive boosting enables extensive speedup in conformational exploration, but can lead to a substantial loss of accuracy.<sup>16</sup> Particularly the reweighting step is a known, but yet not completely solved challenge.<sup>19</sup> Some approaches, like boosting of rotatable torsions only (RaMD),<sup>70</sup> Gaussian aMD<sup>71</sup> or selectively applied aMD,<sup>72</sup> alleviate the impact of the reweighting error.

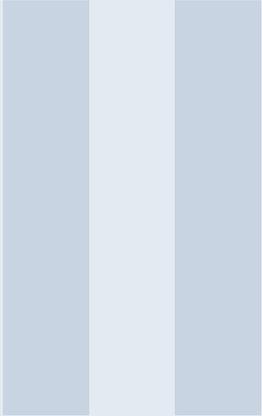
## 6 Conclusion

With the present study we introduce and validate a metric to characterize local protein dynamics on the millisecond time scale. Accelerated MD simulations provide access to time scales three orders of magnitude beyond state-of-the-art sampling time. Subsequent calculation of dihedral entropies from aMD trajectories quantifies backbone flexibility of each residue.

The general functionality of our approach is shown on the model system Di-Ala. We calculated dihedral entropies from a 1 ms cMD simulation of BPTI,<sup>11</sup> serving as reference to validate and benchmark method and metric. We were able to show that dihedral entropies from only 500 ns of aMD simulation identify the same flexibility hot spots, as observed in the 1 ms cMD trajectory. The results are supported by previous NMR studies,<sup>41</sup> which observe local conformational changes in the same regions characterized as most flexible in our study. We applied the procedure on the major birch pollen allergen Bet v 1a. Our study shows good agreement with local dynamics found in a 3  $\mu$ s cMD simulation as well as with NMR derived amide order parameters.<sup>47</sup>

We encourage the application of dihedral entropies as local flexibility metric on different aMD protocols. We anticipate our novel metric to facilitate characterizing and thus understanding the influence of molecular dynamics on biomolecular recognition and protein folding.





# Protein Dynamics and Molecular Docking

<b>7</b>	<b>Introduction</b> .....	<b>25</b>
7.1	Computation and Drug Design	
7.2	Molecular Docking and Dynamics	
<b>8</b>	<b>Theoretical Background</b> .....	<b>28</b>
8.1	Docking Step-by-Step	
8.2	Working with DOCK 3.7	
<b>9</b>	<b>Methods</b> .....	<b>30</b>
9.1	Molecular Dynamics Simulations	
9.2	Conformational Energy Penalties from MD Simulation	
9.3	Conformational Energy Penalties from Binding Free Energy	
9.4	Retrospective virtual screening	
<b>10</b>	<b>Results</b> .....	<b>34</b>
10.1	Conventional Rigid Docking	
10.2	Penalties from MD Simulation	
10.3	Penalties from Binding Affinity	
<b>11</b>	<b>Discussion</b> .....	<b>37</b>
<b>12</b>	<b>Conclusion and Outlook</b> .....	<b>38</b>
<b>13</b>	<b>Appendix</b> .....	<b>51</b>
13.1	Calculation of aMD Parameters	
13.2	Alanine Dipeptide	
13.3	BPTI	
13.4	Bet v 1a	



## 7 Introduction

### 7.1 Computation and Drug Design

The drug discovery and development process has changed remarkably over the last decades. Back in the 1950s researchers still relied solely on the results of *in vivo* testing.<sup>73</sup> Since the 1980s large scale protocols such as high-throughput *in vitro* screening (HTS) of extensive molecule libraries became popular techniques for drug discovery.<sup>74</sup> But the meager yields of hits combined with the high costs has decreased the HTS enthusiasm of the 1990s.<sup>75</sup> These days a typical drug discovery campaign starts with the identification of a bimolecular target. The toolbox to identify and validate biomolecular targets contains a plethora of gadgets ranging from classic gene knockout analysis in mice over phenotypic screening<sup>76</sup> to bioinformatic data mining protocols.<sup>77</sup> Once the target is classified a multidisciplinary team of scientists is delegated to design compounds that selectively bind to the target.<sup>78</sup> Again a vast variety of techniques intertwine to identify "hits", e.g. HTS, virtual screening, biological assays and many more.<sup>79</sup> A molecule usually is considered a hit at a concentration of 50% inhibition ( $IC_{50}$ ) around  $10\mu M$ . In the following hit-to-lead phase the  $IC_{50}$  is typically lowered 1 to 10 nm.<sup>78</sup> If possible the 3D structure of a protein-ligand complex and/or the apo target is determined, opening the door for "structure based drug design". The key idea in structure based drug design is that a target receptor and a potent binder complement each other in structure and chemical properties.<sup>80,81</sup> X-ray crystallography, NMR and computational methodologies work hand in hand, yielding in very efficient drug discovery and optimization routines.<sup>73</sup> The roles computational methods in modern drug design are manifold. Already the structure refinement relies on simulated annealing<sup>82</sup> and the development of the underlying molecular mechanics (MM) force fields. Virtual screens are used to scan colossal compound libraries and score each molecule based on its interaction potential with a target. In the more recent past, molecular dynamics (MD) simulations were recognized as valuable tool for modern drug design.<sup>83</sup> Altogether, today's computer aided campaigns, perform comparably well as HTS.<sup>2,78</sup> This, a versatile application of computational tools facilitates a more rapid and financially less intensive discovery and optimization of novel drug candidates.

### 7.2 Molecular Docking and Dynamics

Molecular docking is a very efficient and valuable computational component in state-of-the-art structure-based drug design. In virtual screens binding poses are predicted by a conformational search. Subsequently these poses are ranked based on a scoring function.<sup>84</sup> Usually the screened libraries comprise several millions of compounds.<sup>2</sup> Given the relentless growth of synthetically available chemical matter, scoring functions are bound to provide a rapid and distinct prioritization

of potential binding poses.<sup>85</sup> Consequently, they often lack accuracy due to partially neglected chemical terms.<sup>86</sup> There are various docking approaches and software packages using different scoring functions,<sup>78</sup> such as DOCK,<sup>87</sup> FlexX<sup>88</sup> and GOLD.<sup>89</sup>

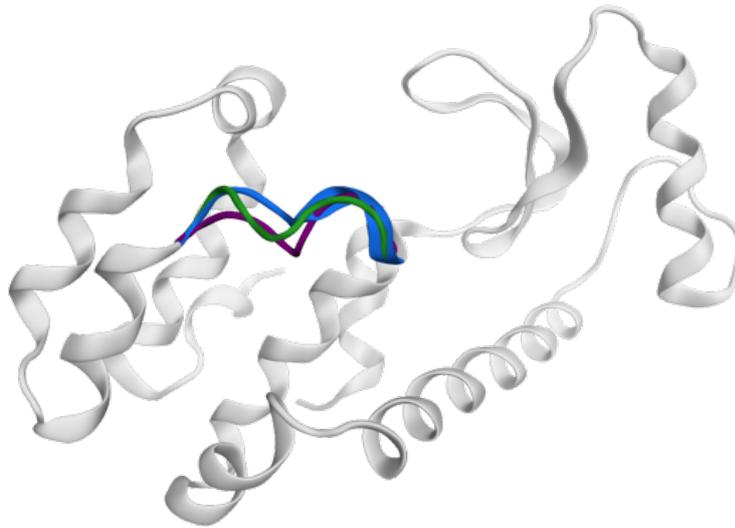
Traditionally a flexible ligand is docked into a rigid target cavity.<sup>90</sup> Recent approaches introduce receptor flexibility by consideration of several binding site conformations.<sup>91–93</sup> The sampling as well as the weighting of these receptor conformations remain challenging yet.<sup>94</sup> Nevertheless, the selection of a diverse set of target conformations enhances the search for new ligands.<sup>95</sup> Molecular dynamics (MD) simulations have proven to be an efficient tool to capture dynamics of macromolecules.<sup>9,96,97</sup> Yet, as mentioned above, state-of-the-art MD simulations routinely cover the microsecond time scale, while many biologically relevant motions appear on the millisecond time scale or slower.<sup>11</sup> Inherent limitations in conformational sampling can either be overcome by usage of dedicated simulation hardware or by application of enhanced sampling algorithms. Accelerated MD (aMD) is such algorithm and has been shown to extend the sampled conformational space drastically compared to conventional MD (cMD) simulation.<sup>21</sup> Representative states can be identified via clustering of the trajectory. Access to extended time scales of local dynamics is known to be beneficial in encompassing target flexibility in docking screens.<sup>86</sup>

Another challenge in flexible receptor docking is combining the score of each pose in each receptor conformation. The result in general is not a single scoring value but rather a scoring profile. The final rank of a pose can either be obtained by considering only the best score or by averaging over multiple scores.<sup>98</sup> The latter strategy assumes a normal energy distribution, or alternatively a Boltzmann distribution of scores which can lead to a reduction in accuracy.<sup>99,100</sup>

When applying the best score strategy the possibility of assigning high ranks to high-energy conformations arises. Despite the elevated internal energies of an unfavorable state it is still possible that it strongly binds the ligand.<sup>101</sup> Thus, it is appealing to include an energy penalty term for each receptor conformation in the scoring function.<sup>102</sup> One strategy is to weight the receptor conformations based on their occurrence during a simulation.<sup>103,104</sup>

The Shoichet group recently proposed the incorporation of experimentally deduced conformational energy penalties in docking scoring functions. Here, the Boltzmann-weighted energy penalties directly derive from crystallographic state occupancies.<sup>94</sup> This method decreases the predominance of higher-energy conformations and allows a prediction of conformational changes in the receptor upon ligand binding.

Many targets have been shown to undergo conformational changes upon ligand binding.<sup>105</sup> Still, for most of them no high-resolution crystal structures of the apo state exist in which several states are modeled. As extension to ensembles from X-ray structures we propose MD simulations to estimate the population of alternate receptor states. We probe our approach for the L99A mutant of T4 lysozyme. (see Figure 7.1) The system is very well characterized by multiple structural and



**Figure 7.1:** Structure of T4 Lysozyme (L99A) with three loop states modeled for the F-helix. (PDB 4W57)

binding studies<sup>106,107</sup> A recent study pointed out conformational differences of the T4 lysozyme cavity depending on the ligand bound.<sup>105</sup> Yet, there is no crystal structure of the apo protein that displays all three observed states. Altogether T4 lysozyme is an ideal candidate for our study.

## 8 Theoretical Background

### 8.1 Docking Step-by-Step

#### Retrospective Screening

The initial challenge in every virtual screening campaign is setting up the system. Three dimensional structures of proteins and DNA deriving from X-ray crystallography, NMR or electron microscopy can be found in the protein databank (PDB). The target structure needs to be prepared properly for docking, especially appropriate protonation of the binding site is crucial. In case a number ligands is already known, it is advantageous to carry out retrospective sanity checks.<sup>2</sup> For these screens each known ligand is complemented with 30-50 property matched decoys. DUDE<sup>108</sup> is a convenient web-tool to generate a decoy library. For every molecule poses and interaction energies are calculated and scored. The aim is to find an enrichment of known ligands in the top scoring compounds. Enrichment is mostly visualized in receiver operator curves (ROC).<sup>109</sup> Using a logarithmic scaling of the x-axis increases the weight on early enrichment. The adjusted area under the half-logarithmic curve (LogAUC) is a convenient method to quantify the enrichment. Here the LogAUC of a random enrichment (14.5%) is subtracted from the total LogAUC, so that positive values indicate enrichment better than random.<sup>110</sup>

#### Prospective Screening

With the system all set, one is ready to screen extensive compound libraries, such as ZINC<sup>111,112</sup> for novel binders. ZINC is a free database containing over 100 million molecules. All compounds in this library are commercially available and in a ready-to-dock, 3D format. Again potential binding poses are generated and ranked for each molecule. Subsequently the top scoring molecules need to be inspected and evaluated carefully. Docking algorithms are usually designed to emphasize on speed rather than accuracy.<sup>2</sup> Considering the simplicity of scoring functions and the extent of screened molecules it is not surprising that many compounds are misrepresented and artificially high-ranking. Thus, a careful manual curating of the list is an indispensable step.

#### Experimental Testing

Evaluation of the docking results with experimental testing of potential binders is essential. Activity alone can be tested in binding and functional assays. Those have to be evaluated carefully as they can be biased by experimental artifacts.<sup>113-116</sup> Experimental determination of ligand binding poses can be realized using X-ray crystallography and NMR.

## 8.2 Working with DOCK 3.7

In this study the DOCK 3.7 suite was used to conduct all presented docking screens. The workflow in DOCK is basically divided into two major parts:

- Orienting the ligand
- Evaluating the orientation

DOCK 3.7 is able to perform docking screens with or without a known crystal ligand. It can be advantageous to start from a structure of the target in complex with a ligand, but it will also bias the results towards the known ligand. The receptor pocket is represented with spheres, that are modeled based on the shape of the site and the pose of the crystal ligand (if given).<sup>87</sup> Poses generated by comparing the atom-atom distances to the distance between the sphere centers, following the longest distance heuristic. Using this method reduces the number possible poses.<sup>117</sup>

The scoring of each pose is based on several interaction grids. On each grid point the contribution of the receptor to the score is calculated and stored. So each grid is generated only once, which decreases the computational time. The calculation of Van der Waals interactions is based on an AMBER force field using a Lennard Jones Potential.<sup>118</sup> Electrostatics are approximated with QNIFFT using a Poisson Boltzmann model.<sup>119,120</sup> Another grid scores effects deriving from ligand desolvation.<sup>110</sup>

### 8.2.1 Flexible Receptor Docking with DOCK 3.7

In the flexible receptor (FlexRec) routine in DOCK 3.7 individual residues or loops are specified as flexible while the rest of the binding site is treated as rigid.<sup>94</sup> An invariant van der Waals and ligand desolvation grid is calculated for the rigid parts receptor excluding flexible residues entirely. Grids for the flexible domains of the receptor are calculated separately for each residue in each state. Ligand desolvation and van der Waals interactions are additive, thus, the individual grids can simply be added to the invariant grids.

Combining electrostatic terms is slightly more complex. To build the invariant electrostatic grid, flexible residues or loops are included and modeled in their most occupied state. The individual grids of each flexible residue in each state are calculated separately, while the rest of the flexible residues is represented in their most occupied state. Subsequent subtraction of the invariant grid leads to grids that reflect the impact of the movement on the electrostatics of each state.

The rank of each pose is ultimately the best score resulting from all possible combinations of conformational states.

## 9 Methods

### 9.1 Molecular Dynamics Simulations

Simulations from three different crystal structures of T4 Lysozyme (L99A) representing the closed, intermediate and open state of the F-Helix (PDB 4W52, 4W57 and 4W59) were performed. As described in part 1, MD simulations were performed with the AMBER14 simulation package.<sup>50</sup> MOE (Molecular Operating Environment, Chemical Computing Group, version 2014.0901)<sup>51</sup> and the protonate3D tool<sup>52</sup> were used to prepare the structures. With tleap of the AmberTools15<sup>50</sup> package all systems were soaked into a truncated octahedral solvent box of TIP3P water molecules<sup>53</sup> with a minimum wall distance was set to 10 Å. The parameters derive from the AMBER force field 99SB-ILDN.<sup>54</sup> Each system was carefully equilibrated using a multi step equilibration protocol.<sup>55</sup> Precedent cMD simulations as well as all aMD simulations were performed in NpT ensemble using pmemd.cuda.<sup>56</sup> Bonds involving hydrogen atoms were restrained by applying the SHAKE algorithm,<sup>57</sup> allowing a time step of 2.0 fs. Atmospheric pressure of the system was preserved by weak coupling to an external bath using the Berendsen algorithm.<sup>58</sup> The Langevin thermostat<sup>59</sup> was used to maintain the temperature during simulations at 300 K.

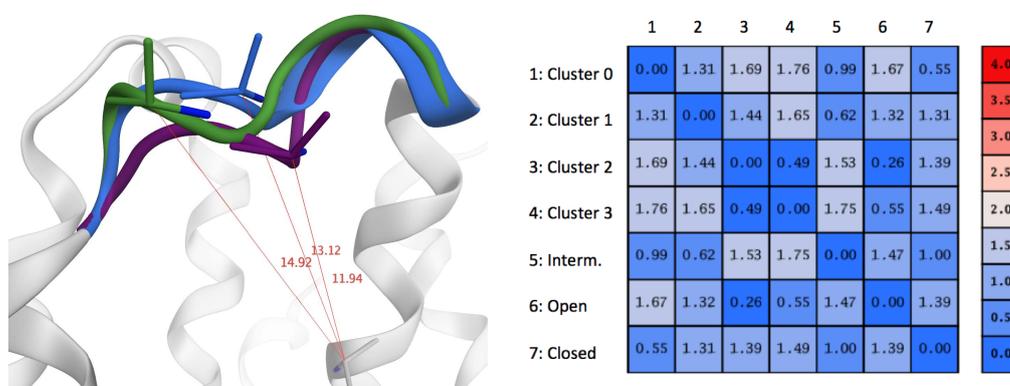
All aMD simulations were performed using the dihedral as well as the dual-boost protocol<sup>60</sup> implemented in pmemd.cuda.<sup>56</sup> The latter accelerates the total potential and applies an extra boosting to the dihedral potential. It has been shown that dual-boost aMD simulations sample the diffusive solvent motions more extensively. Ensemble averages as well as entropy estimates converge faster than in dihedral-boost only aMD simulations.<sup>26,60</sup> All simulations were analyzed using cpptraj<sup>61</sup> in AmberTools15.<sup>50</sup> The 1  $\mu$ s cMD and aMD trajectories were stored as 100,000 equal-spaced snapshots (2.0 fs spacing).

### 9.2 Conformational Energy Penalties from MD Simulation

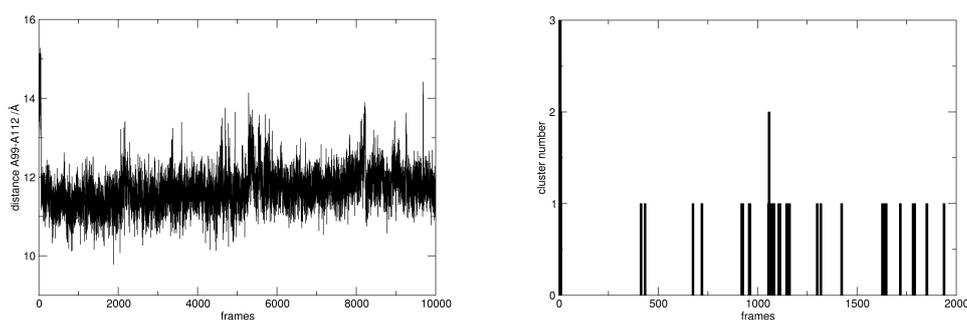
A distance criterion was applied to distinguish between the individual loop state. Alanine 99 (A99) is located in the rigid core of the binding pocket, while Alanine 112 (A112) is a central part of the flexible loop. In the crystal structure of T4 lysozyme L99A in complex with n-butylbenzene (PDB 4W57) all three loop states are modeled. The distance between the C $\alpha$  of A99 and A112 in this structure is 11.94 Å for the closed, 13.12 Å for the intermediate and 14.92 for the open state. We observe clear trends towards distinct loop states in the cMD simulations, whereas all aMD simulations showed a very broad and noisy distance distribution. We used a hierarchical agglomerative (bottom-up) approach to cluster the trajectories based on the A99-A112 distance. No discrete loop state, similar to any of the crystal structure conformations, could be reconstructed from the aMD trajectories. However, the 1  $\mu$ s cMD simulation starting from the open state showed

transitions between all three loop states and was therefore used for further evaluation.

Analyzing the cluster centroids, we find conformations of the F-helix similar to the three distinct states of the known crystal structures (root mean square deviation (RMSD) less or equal to 0.62 Å, Figure 9.1). The closed loop conformation is represented by cluster 0, the intermediate by cluster 1 and the open state by cluster 2 and 3. Figure 9.2 shows the sequence of visited clusters. The simulation that starts from the open state (cluster 3) immediately collapsed into the closed state (cluster 0). There are several transitions between the closed and intermediate state (cluster 1) throughout the whole simulation. After 500 ns the open state (cluster 2) is revisited. The cluster size is the normalized cluster occupancy, which derives from the number of frames sampled in each cluster divided by the total number of frames (100 000). Through Boltzmann weighting of the cluster size the receptor energy penalty for the docking runs was calculated. (Table 9.1)



**Figure 9.1:** Left: Binding site with all three loop states and distances A99-A112 (PDB 4W57) Right: RMSD between cluster and crystal structures of F-helix.



**Figure 9.2:** 1  $\mu$ s cMD simulation. Left: Distance A99-A112 evolution in 1  $\mu$ s cMD (sieved to 10 000 frames). Right: Cluster occupancy over time  $\mu$ s cMD (sieved to 2 000 frames), several transitions between all three states.

**Table 9.1:** Normalized cluster occupancies and energy penalties from MD simulation

State	Cluster number	Clustersize	Penalty /(kcal/mol)
closed	0	0.971	0.0177
intermediate	1	0.025	2.21
open	2, 3	0.005	3.18

### 9.3 Conformational Energy Penalties from Binding Free Energy

Previous studies on the L99A mutant of T4 Lysozyme showed significant differences in the increase of binding affinity and water-octanole transfer energy for ligands growing from benzene, to toluol, ethyl-, n-propyl and n-butylbenzene. Elongation of the alkyl chain raises the affinity of the ligands to the hydrophobic pocket, yet only half as much as expected from water-octanol transfer energies.<sup>105,106</sup> It has been proposed that the difference in free energy is consumed by the conformational change needed to fit the larger ligand in the pocket.<sup>105</sup>

We calculated the free energy difference and related it to the crystallographic ligand occupancy of each loop state.<sup>105</sup> (Table 9.2) We introduced an "invisible state" *i*, which represents the uncertainty in crystallographic occupancies. Solving the system of equations results in the contribution of each conformation to the binding free energy. We thereby obtain a penalty energy of 0.81 kcal/mol for the closed state, 1.24 kcal/mol for the intermediate state and 3.37 kcal/mol for the open state.

**Table 9.2:** Crystallographic ligand occupancies and differences in free energy.

Ligand	Closed	Intermediate	Open	Invisible state	$\Delta\Delta G$ /(kcal/mol)
Benzene	0.7			0.3	2.284
Toluol	0.8			0.2	1.794
Ethylbenzene	0.55	0.35		0.1	1.514
n-Butylbenzene		0.6	0.3	0.1	0.874

### 9.4 Retrospective virtual screening

DOCK 3.7 was used for all presented docking screens. The energy penalties, calculated as described above, were incorporated in the implemented flexible receptor docking protocol. We used the crystal structure of T4 lysozyme in complex with n-butylbenzene (PDB 4W57), in which all three loop states are modeled. For ligand enrichment calculations 127 known ligands and 6700 decoys generated with DUDE were docked.<sup>108</sup>

Further we performed standard rigid docking screens for each loop state. We used the most occu-

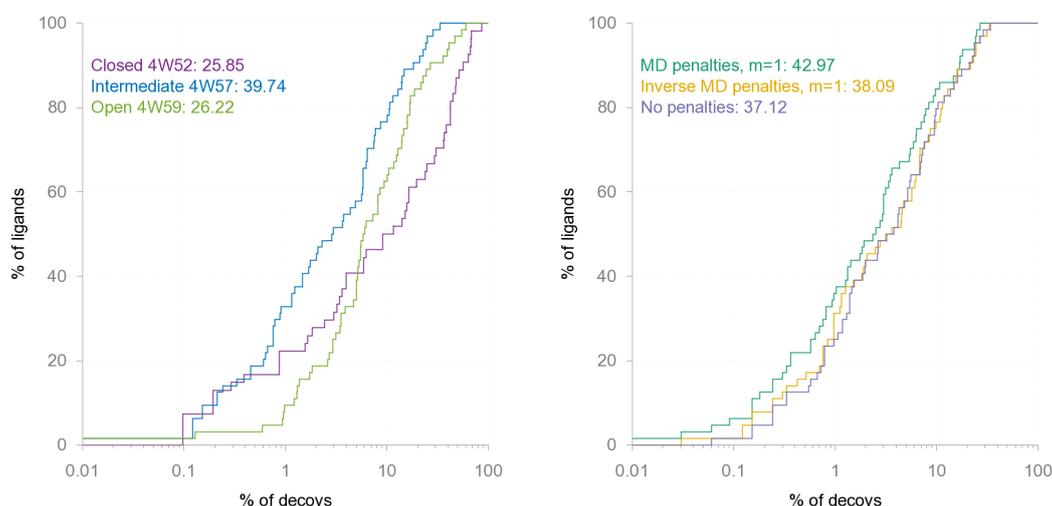
---

ped loop state with benzene, n-butylbenzene and n-hexylbenzene bound as crystal ligands (PDB 4W52, 4W57, 4W59). The poses in each screen were scored and evaluated individually. Subsequently we applied the penalties described above. The scores including the penalty were combined and reevaluated. The last approach is further referred to as multiple receptor docking.

# 10 Results

## 10.1 Conventional Rigid Docking

To benchmark the flexible receptor docking screens we carried out a conventional rigid docking screen with each loop conformation. Poses for ligands and decoys were generated and scored based on three different crystal structures. T4-Lysozyme in complex with benzene, n-butylbenzene and n-hexylbenzene were chosen to represent the closed, intermediate and open state (PDB 4W52, 4W57 and 4W59). Docking screens using the closed and open state of the F-Helix yield similar results (Figure 10.1). With a LogAUC of 25.85 for the closed and 26.22 for the open state, both show weak enrichment compared to docking to the intermediate state which results in a LogAUC of 39.74. Especially when looking at the early enrichment in the top 1% of docked compounds we find very few of the known binders.



**Figure 10.1:** LogROC plot and adjusted LogAUC values. Left: Conventional rigid docking; docking to the intermediate state (blue) leads to highest enrichment (LogAUC 39.74). Right: Flexible Receptor Docking with penalties from MD simulations; incorporation of penalties increases enrichment.

## 10.2 Penalties from MD Simulation

The distance between Alanine 99 (A99) and Alanine 112 (A112) was chosen as criteria to define and distinguish the three loop states. Clustering of a 1  $\mu$ s cMD simulation based on this distance led to the occupancy and subsequently the energy penalty of each state (see Methods). This penalty was then applied to the flexible receptor algorithm implemented in DOCK 3.7 (FlexRec)

by Fischer et. al.<sup>94</sup> The LogROC and the according LogAUC show high and early enrichment of known ligands. Compared to the best results from the conventional docking with a LogAUC of 39.74 for the intermediate loop, the LogAUC increases 42.97 by using FlexRec docking. In addition, more known ligands rank in the top 0.1% when using FlexRec compared to conventional rigid docking.

To investigate the impact of the introduced penalty energies we also performed a FlexRec docking without any penalty applied. We further "inverted" the original penalties by switching the highest with the lowest penalty. In both cases we obtain less enrichment of known ligands, especially in the top 0.1% of docked molecules. (Figure 10.1)

Besides enrichment of known ligands we also investigated the accuracy of pose and loop state prediction abilities of the methodology. We compared docked poses of eight homologous ligands to their crystal structures.<sup>105</sup> The dominant receptor conformation in these complexes/structures was found to shift from the closed to the open state with increasing ligand size. Loop states for five out of eight of the best scoring poses agree with the most occupied states found in the crystal structures. For n-propylbenzene two crystal ligand poses are observed and an occupancy of 60% for the closed state. The best scoring pose in the FlexRec docking was found for the intermediate state. More definite are the differences for n-pentyl- and n-hexylbenzene. Both bind exclusively to the open conformation in the crystal structure, but score best in the intermediate state.

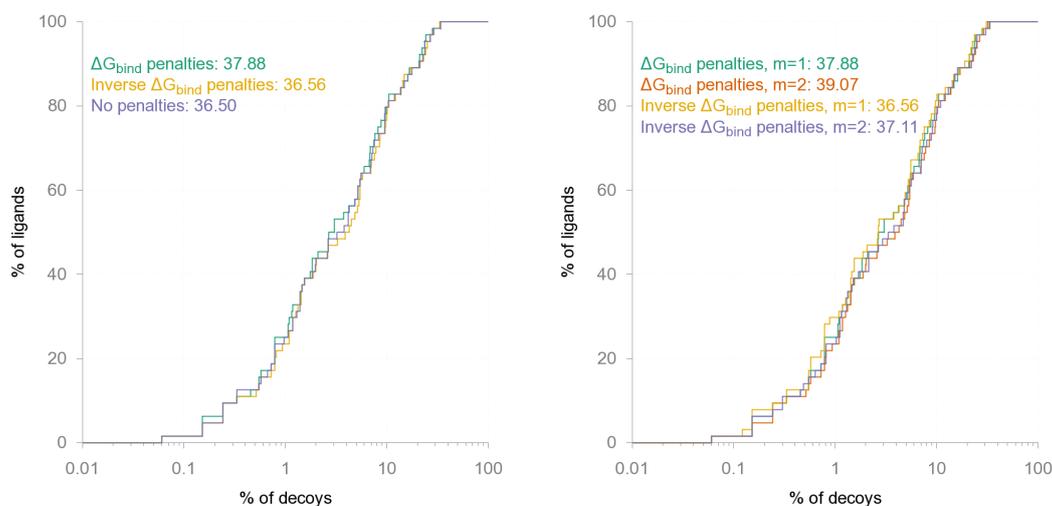
## 10.3 Penalties from Binding Affinity

### 10.3.1 Flexible Receptor Docking

The FlexRec docking screens using energy penalties based on binding affinity differences result in a LogAUC of 37.99 (Figure 10.2). The LogROC shows less enrichment of ligands in the top 0.1% compared to the results from MD simulation based penalties. Furthermore, we do not observe significant differences in the enrichment results on applying no penalty at all or inverting the weighting. Also introduction of a multiplier, as suggested by Fischer et al.,<sup>94</sup> hardly changes the results. Regarding the reproduction of crystal poses and preferred loop states the approach performs similar to the MD-FlexRec protocol.

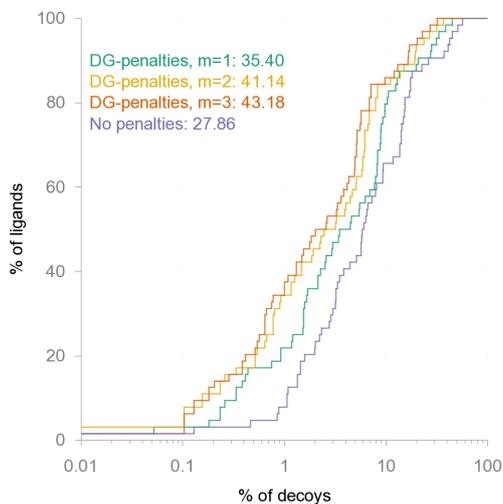
### 10.3.2 Multiple Receptor Docking

To customize the docking spheres used to sample poses in each loop state we extended the FlexRec methodology to a multiple receptor docking scheme (see Methods). This approach leads to a LogAUC of 35.40 for the binding affinity based penalties (Figure 10.3). Applying a multiplier of  $m=2$  as proposed by Fischer et al. 2014 increases the LogAUC to 41.34. Combining the scores without penalties leads to a clear decrease of the LogAUC to 27.89. The significant impact of



**Figure 10.2:** LogROC plot and adjusted LogAUC values of flexible receptor docking screens using penalties estimated from binding affinity. Left: Incorporation of penalties slightly improves enrichment. Right: Improvement of enrichment by applying a multiplier

the penalty is also apparent in the ligands found in the top 1% of docked molecules, as can be seen in the LogROC. Docking based on three different loop structures is further the only approach discussed that finds the correct loop state for n-pentyl and n-hexylbenzene. On the other hand it falsely favors the intermediate over the closed state for n-propylbenzene and toluol.



**Figure 10.3:** LogROC plot and adjusted LogAUC values of multiple receptor docking screens using penalties estimated from binding affinity.

## 11 Discussion

Protein flexibility, i.e. the conformations of the binding site, plays a key role in ligand binding.<sup>9,83,98,121,122</sup> In this work we compare various docking setups for the flexible binding site of T4 lysozyme. We estimate the success of retrospective docking campaigns based on LogROC plots, LogAUC, and the accuracy in reproduction of crystal ligand poses.

The standard rigid docking protocol performs fairly well for the intermediate loop state. This result is hardly surprising considering that the vast majority of known ligands was reported bound to this state. Clustering of all T4 lysozyme structures in the PDB identified the intermediate state as the most common one.<sup>105</sup> Thus, the test set is presumably biased towards this conformation.

The target of this work is to provide an efficient methodology to increase the novelty in high ranking ligands and decrease the bias to known structures. The original work from Fischer et al.<sup>94</sup> on which this work is based upon, achieves this aim by using a high resolution room temperature apo crystal structure of cytochrome C peroxidase. Here we aim to establish a robust protocol that estimates receptor flexibility for proteins without the support of multi-state structural data.

MD simulations have been proposed to assist flexible receptor docking screens in several previous expedient studies.<sup>91,123,124</sup> We envisaged to extend the covered conformational changes using aMD. Unfortunately, we did not succeed in defining discrete receptor states from aMD trajectories. We observed the most significant increase of dynamics compared to the cMD simulations in the N-domain of the protein far away from the binding site. For the F-helix itself we found a broad range of conformations, but were not able to reconstruct the experimentally observed states. These results indicate a too vigorous biasing of the simulations.

With a 1  $\mu$ s of cMD simulation starting from the open state we were able to sample transitions between all three loop states. This additionally indicates that the distinct states are accessible without biasing the underlying potential.

The ratio of sampled states leads to energy penalties that result in very reasonable enrichment. The very early enrichment of known ligands in the top 0.1% to 1% of docked molecules is clearly improved by applying the penalty. We find comparable enrichment for docking to the rigid intermediate state only. Similar findings were reported in the work of Fisher et al.<sup>94</sup> In their study they did not observe an increased enrichment of ligands, but an increase in the covered chemical space. We anticipate that a prospective screen followed by experimental testing would lead to a similar outcome for our protocol.

In a further approach we incorporated more experimental information in the estimation of the receptor penalties. Determining the bias of each conformation based on a binding affinity study led to acceptable results with a LogAUC of 37.88. Yet, this docking screen is hardly any more successful than applying no penalty at all, which results in a LogAUC of 36.50. Also when we

introduce a multiplier of  $m=2$  the impact of the weighted penalty is subtle. Yet, we observe a slight improvement in enrichment when comparing the suggested penalty (LogAUC: 39.07) to the inverted ones (LogAUC: 37.11). In docking screens using inverted penalties the highest and lowest energy are switched, penalizing the closed state the most. These results indicate that the ratio of the penalties deriving from the binding affinity study might not be robust enough to prioritize the individual states.

In both of the described protocols one sphere set was used to sample conformations of each loop state. These spheres were based on the intermediate state of the F-helix. Ligands binding to the open state hardly fit into this sphere set. We assume this is at least partially the reason why the loop preferences were predicted incorrectly for the more bulky ligands like *n*-pentyl- and *n*-hexylbenzene. To overcome this restraint we combined results from docking ligands to each loop state individually. We added the binding affinity based energy penalty to each score and reevaluated the enrichment. Using this protocol we found a major improvement on applying the penalty (LogAUC: 35.40) compared to not prioritizing any state (LogAUC: 27.86). Increasing the weight with a multiplier further increases the enrichment to a LogAUC of 41.14 for  $m=2$  and 43.18 for  $m=3$ . Thus, the results of the multiple receptor docking approach are comparable to the best rigid docking and MD-FlexRec enrichment. Further we correctly predict the binding mode of *n*-pentyl- and *n*-hexylbenzene. Yet, small ligands binding to the closed state are falsely predicted to bind to the intermediate state. Further, we observe large ligands scoring significantly better, indicating a bias towards high molecular weight. Altogether, we observe an acceptable enrichment and diversity in high ranking ligands docking to multiple receptor conformations, but we also significantly increase the calculation time.

## 12 Conclusion and Outlook

Incorporation of protein flexibility into DOCK 3.7 has shown to increase the chemical space of newly discovered binders.<sup>94</sup> In this work we perform and compare several flexible receptor docking protocols. We use MD simulations and binding studies to weight known conformational states of T4 lysozyme L99A. The results are comparable to what was found for conformational penalties from X-ray occupancies. Thus, in absence of appropriate experimental data MD simulations represent a feasible method to prioritize receptor conformations for molecular docking.

We will continue the project to conduct prospective screens and experimental testing of compounds. Numerous pharmaceutically highly relevant proteins, such as neuraminidases, have been found to change their conformation upon ligand binding.<sup>105,125</sup> We envisage to extend our set of model systems to test and substantiate our approach.

## Bibliography

- [1] J. J. Falke, “A moving story,” *Science*, vol. 295, no. 5559, pp. 1480–1481, 2002.
- [2] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Huebner, and D. Kern, “Intrinsic motions along an enzymatic reaction trajectory,” *Nature*, vol. 450, no. 7171, pp. 838–844, 2007.
- [3] D. D. Boehr, R. Nussinov, and P. E. Wright, “The role of dynamic conformational ensembles in biomolecular recognition,” *Nat. Chem. Biol.*, vol. 5, no. 12, pp. 954–954, 2009.
- [4] M. Fischer, R. G. Coleman, J. S. Fraser, and B. K. Shoichet, “Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery,” *Nat. Chem.*, vol. 6, no. 7, pp. 575–583, 2014.
- [5] J. E. Fuchs, R. G. Huber, B. J. Waldner, U. Kahler, S. von Grafenstein, C. Kramer, and K. R. Liedl, “Dynamics govern specificity of a protein-protein interface: Substrate recognition by thrombin,” *PLoS One*, vol. 10, no. 10, 2015. e0140713.
- [6] J. R. Schnell, H. J. Dyson, and P. E. Wright, “Structure, dynamics, and catalytic function of dihydrofolate reductase,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 33, pp. 119–140, 2004.
- [7] R. B. Fenwick, S. Esteban-Martin, and X. Salvatella, “Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles,” *Eur. Biophys. J. Biophys. Lett.*, vol. 40, no. 12, pp. 1339–1355, 2011.
- [8] U. Hensen, T. Meyer, J. Haas, R. Rex, G. Vriend, and H. Grubmuller, “Exploring protein dynamics space: The dynasome as the missing link between protein structure and function,” *PLoS One*, vol. 7, no. 5, 2012. e33931.
- [9] J. D. Durrant and J. A. McCammon, “Molecular dynamics simulations and drug discovery,” *BMC Biol.*, vol. 9, 2011. 1-9.
- [10] K. Henzler-Wildman and D. Kern, “Dynamic personalities of proteins,” *Nature*, vol. 450, no. 7172, pp. 964–972, 2007.
- [11] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. B. Shan, and W. Wriggers, “Atomic-level characterization of the structural dynamics of proteins,” *Science*, vol. 330, no. 6002, pp. 341–346, 2010.

- [12] R. C. Bernardi, M. C. R. Melo, and K. Schulten, "Enhanced sampling techniques in molecular dynamics simulations of biological systems," *Biochim. Biophys. Acta, Gen. Subj.*, vol. 1850, no. 5, pp. 872–877, 2015.
- [13] W. Wenzel and K. Hamacher, "Stochastic tunneling approach for global minimization of complex potential energy landscapes," *Physical Review Letters*, vol. 82, no. 15, pp. 3003–3007, 1999.
- [14] D. Hamelberg, J. Mongan, and J. A. McCammon, "Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules," *J. Chem. Phys.*, vol. 120, no. 24, pp. 11919–11929, 2004.
- [15] L. C. T. Pierce, R. Salomon-Ferrer, C. A. F. de Oliveira, J. A. McCammon, and R. C. Walker, "Routine access to millisecond time scale events with accelerated molecular dynamics," *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 2997–3002, 2012.
- [16] C. A. F. de Oliveira, D. Hamelberg, and J. A. McCammon, "On the application of accelerated molecular dynamics to liquid water simulations," *J. Phys. Chem. B*, vol. 110, no. 45, pp. 22695–22701, 2006.
- [17] R. I. Cukier and M. Morillo, "A targeted reweighting method for accelerating the exploration of high-dimensional configuration space," *J. Chem. Phys.*, vol. 123, no. 23, 2005.
- [18] Y. L. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker, and J. A. McCammon, "Improved reweighting of accelerated molecular dynamics simulations for free energy calculation," *J. Chem. Theory Comput.*, vol. 10, no. 7, pp. 2677–2689, 2014.
- [19] T. Y. Shen and D. Hamelberg, "A statistical analysis of the precision of reweighting-based simulations," *J. Chem. Phys.*, vol. 129, no. 3, p. e034103, 2008.
- [20] D. Bucher, B. J. Grant, P. R. Markwick, and J. A. McCammon, "Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics," *PLoS Comput. Biol.*, vol. 7, no. 4, p. e1002034, 2011.
- [21] P. R. L. Markwick and J. A. McCammon, "Studying functional dynamics in bio-molecules using accelerated molecular dynamics," *Phys. Chem. Chem. Phys.*, vol. 13, no. 45, pp. 20053–20065, 2011.
- [22] Y. Wang, P. R. L. Markwick, C. A. F. de Oliveira, and J. A. McCammon, "Enhanced lipid diffusion and mixing in accelerated molecular dynamics," *Biophys. J.*, vol. 102, no. 3, pp. 413A–413A, 2012.

- [23] K. Kappel, Y. Miao, and J. A. McCammon, "Accelerated molecular dynamics simulations of ligand binding to a muscarinic g-protein-coupled receptor," *Q. Rev. Biophys.*, vol. 48, no. 4, pp. 479–487, 2015.
- [24] A. Kalenkiewicz, B. J. Grant, and C.-Y. Yang, "Enrichment of druggable conformations from apo protein structures using cosolvent-accelerated molecular dynamics," *Biology*, vol. 4, no. 2, pp. 344–66, 2015.
- [25] C. Mucksch and H. M. Urbassek, "Enhancing protein adsorption simulations by using accelerated molecular dynamics," *PloS one*, vol. 8, no. 6, p. e64883, 2014.
- [26] D. D. L. Minh, D. Hamelberg, and A. McCammon, "Accelerated entropy estimates with accelerated dynamics," *J. Chem. Phys.*, vol. 127, no. 15, 2007. e154105.
- [27] J. R. Thomas, P. C. Gedeon, B. J. Grant, and J. D. Madura, "Leut conformational sampling utilizing accelerated molecular dynamics and principal component analysis," *Biophys. J.*, vol. 103, no. 1, pp. L01–L03, 2012.
- [28] Q. Bai, Y. Zhang, X. Li, W. Chen, H. Liu, and X. Yao, "Computational study on the interaction between ccr5 and hiv-1 entry inhibitor maraviroc: insight from accelerated molecular dynamics simulation and free energy calculation," *Phys. Chem. Chem. Phys.*, vol. 16, no. 44, pp. 24332–24338, 2014.
- [29] J. E. Fuchs, B. J. Waldner, R. G. Huber, S. von Grafenstein, C. Kramer, and K. R. Liedl, "Independent metrics for protein backbone and side-chain flexibility: Time scales and effects of ligand binding," *J. Chem. Theory Comput.*, vol. 11, no. 3, pp. 851–860, 2015.
- [30] J. E. Fuchs, S. von Grafenstein, R. G. Huber, H. G. Wallnoefer, and K. R. Liedl, "Specificity of a protein-protein interface: Local dynamics direct substrate recognition of effector caspases," *Proteins: Struct., Funct., Bioinf.*, vol. 82, no. 4, pp. 546–555, 2014.
- [31] M. J. Edgeworth, J. J. Phillips, D. C. Lowe, A. D. Kippen, D. R. Higazi, and J. H. Scrivens, "Global and local conformation of human igg antibody variants rationalizes loss of thermodynamic stability," *Angew. Chem., Int. Ed.*, vol. 54, no. 50, pp. 15156–15159, 2015.
- [32] N. Sinha and S. J. Smith-Gill, "Molecular dynamics simulation of a high-affinity antibody-protein complex - the binding site is a mosaic of locally flexible and preorganized rigid regions," *Cell Biochem. Biophys.*, vol. 43, no. 2, pp. 253–273, 2005.

- [33] N. Sinha, S. Mohan, C. A. Lipschultz, and S. J. Smith-Gill, "Differences in electrostatic properties at antibody-antigen binding sites: Implications for specificity and cross-reactivity," *Biophys. J.*, vol. 83, no. 6, pp. 2946–2968, 2002.
- [34] L. Salmon, L. Pierce, A. Grimm, J.-L. O. Roldan, L. Mollica, M. R. Jensen, N. van Nuland, P. R. L. Markwick, J. A. McCammon, and M. Blackledge, "Multi-timescale conformational dynamics of the sh3 domain of cd2-associated protein using nmr spectroscopy and accelerated molecular dynamics," *Angew. Chem., Int. Ed.*, vol. 51, no. 25, pp. 6103–6106, 2012.
- [35] P. R. L. Markwick, C. F. Cervantes, B. L. Abel, E. A. Komives, M. Blackledge, and J. A. McCammon, "Enhanced conformational space sampling improves the prediction of chemical shifts in proteins," *J. Am. Chem. Soc.*, vol. 132, no. 4, pp. 1220–1221, 2010.
- [36] R. G. Huber, C. Eibl, and J. E. Fuchs, "Intrinsic flexibility of nlrp pyrin domains is a key factor in their conformational dynamics, fold stability, and dimerization," *Protein Sci.*, vol. 24, no. 2, pp. 174–181, 2015.
- [37] R. G. Huber, J. E. Fuchs, S. von Grafenstein, M. Laner, H. G. Wallnoefer, N. Abdelkader, R. T. Kroemer, and K. R. Liedl, "Entropy from state probabilities: Hydration entropy of cations," *J. Phys. Chem. B*, vol. 117, no. 21, pp. 6466–6472, 2013.
- [38] A. Wlodawer, J. Walter, R. Huber, and L. Sjolín, "Structure of bovine pancreatic trypsin-inhibitor - results of joint neutron and x-ray refinement of crystal form-ii," *J. Mol. Biol.*, vol. 180, no. 2, pp. 301–329, 1984.
- [39] G. Wagner, A. Demarco, and K. Wuthrich, "Dynamics of aromatic amino-acid residues in globular conformation of basic pancreatic trypsin-inhibitor (bpti) .1. h-1 nmr-studies," *Biophys. Struct. Mech.*, vol. 2, no. 2, pp. 139–158, 1976.
- [40] G. Wagner, D. Bruhwiler, and K. Wuthrich, "Reinvestigation of the aromatic side-chains in the basic pancreatic trypsin-inhibitor by heteronuclear two-dimensional nuclear-magnetic-resonance," *J. Mol. Biol.*, vol. 196, no. 1, pp. 227–231, 1987.
- [41] G. Otting, E. Liepinsh, and K. Wuthrich, "Disulfide bond isomerization in bpti and bpti(g36s) - an nmr-study of correlated mobility in proteins," *Biochemistry*, vol. 32, no. 14, pp. 3571–3582, 1993.
- [42] Y. Xue, J. M. Ward, T. Yuwen, I. S. Podkorytov, and N. R. Skrynnikov, "Microsecond time-scale conformational exchange in proteins: Using long molecular dynamics trajectory to simulate nmr relaxation dispersion data," *J. Am. Chem. Soc.*, vol. 134, no. 5, pp. 2555–2562, 2012.

- [43] H. Ipsen and H. Lowenstein, "Isolation and immunochemical characterization of the major allergen of birch pollen (betula-verrucosa)," *J. Allergy Clin. Immunol.*, vol. 72, no. 2, pp. 150–159, 1983.
- [44] R. Moverare, K. Westritschnig, M. Svensson, B. Hayek, M. Bende, G. Pauli, R. Sorva, T. Haahtela, R. Valenta, and L. Elfman, "Different ige reactivity profiles in birch pollen-sensitive patients from six european populations revealed by recombinant allergens: An imprint of local sensitization," *Int. Arch. Allergy Immunol.*, vol. 128, no. 4, pp. 325–335, 2002.
- [45] F. Ferreira, K. Hirtenlehner, A. Jilek, J. GodnikCvar, H. Breiteneder, R. Grimm, K. HoffmannSommergruber, O. Scheiner, D. Kraft, M. Breitenbach, H. J. Rheinberger, and C. Ebner, "Dissection of immunoglobulin e and t lymphocyte reactivity of isoforms of the major birch pollen allergen bet v 1: Potential use of hypoallergenic isoforms for immunotherapy," *J. Exp. Med.*, vol. 183, no. 2, pp. 599–609, 1996.
- [46] I. Swoboda, A. Jilek, F. Ferreira, E. Engel, K. Hoffmannsommergruber, O. Scheiner, D. Kraft, H. Breiteneder, E. Pittenauer, E. Schmid, O. Vicente, E. Heberlebers, H. Ahorn, and M. Breitenbach, "Isoforms of bet-v-1, the major birch pollen allergen, analyzed by liquid-chromatography, mass-spectrometry, and cdna cloning," *J. Biol. Chem.*, vol. 270, no. 6, pp. 2607–2613, 1995.
- [47] S. Grutsch, J. E. Fuchs, R. Freier, S. Kofler, M. Bibi, C. Asam, M. Wallner, F. Ferreira, H. Brandstetter, K. R. Liedl, and M. Tollinger, "Ligand binding modulates the structural dynamics and compactness of the major birch pollen allergen," *Biophys. J.*, vol. 107, no. 12, pp. 2963–2972, 2014.
- [48] R. Freier, E. Dall, and H. Brandstetter, "Protease recognition sites in bet v 1a are cryptic, explaining its slow processing relevant to its allergenicity," *Sci. Rep.*, vol. 5, p. 12707, 2015.
- [49] Y. Machado, R. Freier, S. Scheiblhofer, T. Thalhamer, M. Mayr, P. Briza, S. Grutsch, L. Ahammer, J. E. Fuchs, H. G. Wallnoefer, A. Isakovic, V. Kohlbauer, A. Hinterholzer, M. Steiner, M. Danzer, J. Horejs-Hoeck, F. Ferreira, K. R. Liedl, M. Tollinger, P. Lackner, C. M. Johnson, H. Brandstetter, J. Thalhamer, and R. Weiss, "Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen," *J. Allergy Clin. Immunol.*, pp. 1525–34, 2015.
- [50] D. Case, J. Berryman, R. Betz, D. Cerutti, T. Cheatham, III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, , A. Kovalenko, T. Lee,

- S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. Merz, G. Monard, P. Needham, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, R. Salomon-Ferrer, C. Simmerling, W. Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, D. York, and P. Kollman, *AMBER 2015, University of California, San Francisco*.
- [51] “Molecular operating environment (moe), 2014.09.” Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7.
- [52] P. Labute, “Protonate3d: Assignment of ionization states and hydrogen coordinates to macromolecular structures,” *Proteins: Struct., Funct., Bioinf.*, vol. 75, no. 1, pp. 187–205, 2009.
- [53] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.
- [54] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, “Improved side-chain torsion potentials for the amber ff99sb protein force field,” *Proteins: Struct., Funct., Bioinf.*, vol. 78, no. 8, pp. 1950–1958, 2010.
- [55] H. G. Wallnoefer, K. R. Liedl, and T. Fox, “A challenging system: Free energy prediction for factor xa,” *J. Comput. Chem.*, vol. 32, no. 8, pp. 1743–1752, 2011.
- [56] R. Salomon-Ferrer, A. W. Goetz, D. Poole, S. Le Grand, and R. C. Walker, “Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald,” *J. Chem. Theory Comput.*, vol. 9, no. 9, pp. 3878–3888, 2013.
- [57] G. Ciccotti and J. P. Ryckaert, “Molecular-dynamics simulation of rigid molecules,” *Comput. Phys. Rep.*, vol. 4, no. 6, pp. 345–392, 1986.
- [58] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, “Molecular-dynamics with coupling to an external bath,” *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [59] S. A. Adelman and J. D. Doll, “Generalized langevin equation approach for atom-solid-surface scattering - general formulation for classical scattering off harmonic solids,” *J. Chem. Phys.*, vol. 64, no. 6, pp. 2375–2388, 1976.
- [60] D. Hamelberg, C. A. F. de Oliveira, and J. A. McCammon, “Sampling of slow diffusive conformational transitions with accelerated molecular dynamics,” *J. Chem. Phys.*, vol. 127, no. 15, p. e155102, 2007.

- [61] D. R. Roe and I. Cheatham, Thomas E., "Ptraj and cpptraj: Software for processing and analysis of molecular dynamics trajectory data," *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, 2013.
- [62] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *J. Mol. Biol.*, vol. 7, no. 1, pp. 95–99, 1963.
- [63] M. J. Wood and J. D. Hirst, "Protein secondary structure prediction with dihedral angles," *Proteins: Struct., Funct., Bioinf.*, vol. 59, no. 3, pp. 476–481, 2005.
- [64] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Ann. Math. Stat.*, vol. 38, no. 5, pp. 2916–2957, 2010.
- [65] S. Kofler, C. Asam, U. Eckhard, M. Wallner, F. Ferreira, and H. Brandstetter, "Crystallographically mapped ligand binding differs in high and low ige binding isoforms of birch pollen allergen bet v 1," *J. Mol. Biol.*, vol. 422, no. 1, pp. 109–123, 2012.
- [66] A. Mittermaier and L. E. Kay, "Review - new tools provide new insights in nmr studies of protein dynamics," *Science*, vol. 312, no. 5771, pp. 224–228, 2006.
- [67] A. Kidera and N. Go, "Normal mode refinement - crystallographic refinement of protein dynamic structure .1. theory and test by simulated diffraction data," *J. Mol. Biol.*, vol. 225, no. 2, pp. 457–475, 1992.
- [68] Y. Miao, S. E. Nichols, and J. A. McCammon, "Free energy landscape of g-protein coupled receptors, explored by accelerated molecular dynamics," *Phys. Chem. Chem. Phys.*, vol. 16, no. 14, pp. 6398–6406, 2014.
- [69] Z. Jing and H. Sun, "A comment on the reweighting method for accelerated molecular dynamics simulations," *J. Chem. Theory Comput.*, vol. 11, no. 6, pp. 2395–2397, 2015.
- [70] U. Doshi and D. Hamelberg, "Improved statistical sampling and accuracy with accelerated molecular dynamics on rotatable torsions," *J. Chem. Theory Comput.*, vol. 8, no. 11, pp. 4004–4012, 2012.
- [71] Y. Miao, V. A. Feher, and J. A. McCammon, "Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation," *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3584–3595, 2015.
- [72] J. Wereszczynski and J. A. McCammon, "Using selectively applied accelerated molecular dynamics to enhance free energy calculations," *J. Chem. Theory Comput.*, vol. 6, no. 11, pp. 3285–3292, 2010.

- [73] B. E. Maryanoff
- [74] J. G. Lombardino and r. Lowe, J. A., "The role of the medicinal chemist in drug discovery—then and now," *Nat Rev Drug Discov*, vol. 3, no. 10, pp. 853–62, 2004.
- [75] R. Lahana, "How many leads from hts?," *Drug Discovery Today*, vol. 4, no. 10, pp. 447 – 448, 1999.
- [76] G. Kurosawa, Y. Akahori, M. Morita, M. Sumitomo, N. Sato, C. Muramatsu, K. Eguchi, K. Matsuda, A. Takasaki, M. Tanaka, Y. Iba, S. Hamada-Tsutsumi, Y. Ukai, M. Shiraishi, K. Suzuki, M. Kurosawa, S. Fujiyama, N. Takahashi, R. Kato, Y. Mizoguchi, M. Shamoto, H. Tsuda, M. Sugiura, Y. Hattori, S. Miyakawa, R. Shiroki, K. Hoshinaga, N. Hayashi, A. Sugioka, and Y. Kurosawa, "Comprehensive screening for antigens overexpressed on carcinomas via isolation of human mabs that may be therapeutic," *Proceedings of the National Academy of Sciences*, vol. 105, no. 20, pp. 7287–7292, 2008.
- [77] Y. Yang, S. J. Adelstein, and A. I. Kassis, "Target discovery from data mining approaches," *Drug Discovery Today*, vol. 14, no. 3–4, pp. 147 – 154, 2009.
- [78] W. L. Jorgensen, "The many roles of computation in drug discovery," *Science*, vol. 303, no. 5665, pp. 1813–1818, 2004.
- [79] J. Hughes, S. Rees, S. Kalindjian, and K. Philpott, "Principles of early drug discovery," *British Journal of Pharmacology*, vol. 162, no. 6, pp. 1239–1249, 2011.
- [80] P. M. Dean and P.-L. Chau, "Molecular recognition: optimized searching through rotational 3-space for pattern matches on molecular surfaces," *Journal of Molecular Graphics*, vol. 5, no. 3, pp. 152–158, 1987.
- [81] I. D. Kuntz, "Structure-based strategies for drug design and discovery," *Science*, vol. 257, no. 5073, pp. 1078–1082, 1992.
- [82] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore, "The xplor-nih nmr molecular structure determination package," *Journal of magnetic resonance*, vol. 160, no. 1, pp. 65–73, 2003.
- [83] C. F. Wong and J. A. McCammon, "Protein flexibility and computer-aided drug design," *Annu Rev Pharmacol Toxicol*, vol. 43, pp. 31–45, 2003.
- [84] A. R. Leach, B. K. Shoichet, , and C. E. Peishoff, "Prediction of protein-ligand interactions. docking and scoring: Successes and gaps," *Journal of Medicinal Chemistry*, vol. 49, no. 20, pp. 5851–5855, 2006.

- [85] M. Stahl and M. Rarey, "Detailed analysis of scoring functions for virtual screening," *Journal of medicinal chemistry*, vol. 44, no. 7, pp. 1035–1042, 2001.
- [86] W. Sinko, S. Lindert, and J. A. McCammon, "Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design," *Chemical Biology & Drug Design*, vol. 81, no. 1, pp. 41–49, 2013.
- [87] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, "A geometric approach to macromolecule-ligand interactions," *J Mol Biol*, vol. 161, no. 2, pp. 269–88, 1982.
- [88] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *Journal of molecular biology*, vol. 261, no. 3, pp. 470–489, 1996.
- [89] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *Journal of molecular biology*, vol. 267, no. 3, pp. 727–748, 1997.
- [90] S.-Y. Huang and X. Zou, "Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking," *Proteins-Structure Function and Bioinformatics*, vol. 66, no. 2, pp. 399–421, 2007.
- [91] C. B-Rao, J. Subramanian, and S. D. Sharma, "Managing protein flexibility in docking and its applications," *Drug Discovery Today*, vol. 14, no. 7-8, pp. 394–400, 2009.
- [92] P. R. L. Markwick and J. A. McCammon, "Studying functional dynamics in bio-molecules using accelerated molecular dynamics," *Physical Chemistry Chemical Physics*, vol. 13, no. 45, pp. 20053–20065, 2011.
- [93] N. Bansal, Z. Zheng, and K. M. M. Jr., "Incorporation of side chain flexibility into protein binding pockets using {MTflex}," *Bioorganic and Medicinal Chemistry*, pp. –, 2016.
- [94] M. Fischer, R. G. Coleman, J. S. Fraser, and B. K. Shoichet, "Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery," *Nature Chemistry*, vol. 6, no. 7, pp. 575–583, 2014.
- [95] M. Rueda, M. Totrov, and R. Abagyan, "ALiBERO: Evolving a Team of Complementary Pocket Conformations Rather than a Single Leader," *Journal of Chemical Information and Modeling*, vol. 52, no. 10, pp. 2705–2714, 2012.

- [96] C. S. Tautermann, D. Seeliger, and J. M. Kriegl, "What can we learn from molecular dynamics simulations for gpcr drug design?," *Computational and structural biotechnology journal*, vol. 13, pp. 111–121, 2015.
- [97] L. E. W. LaConte, V. A. Voelz, W. D. Nelson, and D. D. Thomas, "Molecular dynamics simulation of site-directed spin labeling: Experimental validation in muscle fibers," *BiophysJ*, vol. 82, 2002.
- [98] S. von Grafenstein, J. E. Fuchs, and K. R. Liedl, "(how to) profit from molecular dynamics-based ensemble docking," in *Application of Computational Techniques in Pharmacy and Medicine* (L. Gorb, V. Kuz'min, and E. Muratov, eds.), vol. 17 of *Challenges and Advances in Computational Chemistry and Physics*, pp. 501–538, Springer Netherlands, 2014.
- [99] L. S. Cheng, R. E. Amaro, D. Xu, W. W. Li, P. W. Arzberger, and J. A. McCammon, "Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase," *Journal of Medicinal Chemistry*, vol. 51, no. 13, pp. 3878–3894, 2008.
- [100] J. L. Paulsen and A. C. Anderson, "Scoring Ensembles of Docked Protein:Ligand Interactions for Virtual Lead Optimization," *Journal of Chemical Information and Modeling*, vol. 49, no. 12, pp. 2813–2819, 2009.
- [101] M. Fischer, B. K. Shoichet, and J. S. Fraser, "One crystal, two temperatures: Cryocooling penalties alter ligand binding to transient protein sites," *ChemBioChem*, vol. 16, no. 11, pp. 1560–1564, 2015.
- [102] X. Barril and X. Fradera, "Incorporating protein flexibility into docking and structure-based drug design," *Expert Opinion on Drug Discovery*, vol. 1, no. 4, pp. 335–349, 2006.
- [103] J. Lin, A. Perryman, J. Schames, and J. McCammon, "Computational drug design accommodating receptor flexibility: The relaxed complex scheme," *Journal of the American Chemical Society*, vol. 124, no. 20, pp. 5632–5633, 2002.
- [104] R. E. Amaro, R. Baron, and J. A. McCammon, "An improved relaxed complex scheme for receptor flexibility in computer-aided drug design," *Journal of Computer-Aided Molecular Design*, vol. 22, no. 9, pp. 693–705, 2008.
- [105] M. Merski, M. Fischer, T. E. Balius, O. Eidam, and B. K. Shoichet, "Homologous ligands accommodated by discrete conformations of a buried cavity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 16, pp. 5039–5044, 2015.

- [106] A. Morton and B. W. Matthews, "Specificity of ligand binding in a buried nonpolar cavity of t4 lysozyme: Linkage of dynamics and structural plasticity," *Biochemistry*, vol. 34, no. 27, pp. 8576–8588, 1995.
- [107] A. Morton, W. A. Baase, and B. W. Matthews, "Energetic origins of specificity of ligand binding in an interior nonpolar cavity of t4 lysozyme," *Biochemistry*, vol. 34, no. 27, pp. 8564–8575, 1995.
- [108] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of useful decoys, enhanced (dud-e): Better ligands and decoys for better benchmarking," *Journal of Medicinal Chemistry*, vol. 55, no. 14, pp. 6582–6594, 2012.
- [109] A. Nicholls, "What do we know and when do we know it?," *J Comput Aided Mol Des*, vol. 22, no. 3-4, pp. 239–55, 2008.
- [110] M. M. Mysinger and B. K. Shoichet, "Rapid context-dependent ligand desolvation in molecular docking," *J Chem Inf Model*, vol. 50, no. 9, pp. 1561–73, 2010.
- [111] T. Sterling and J. J. Irwin
- [112] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman
- [113] C. H. Arrowsmith, J. E. Audia, C. Austin, J. Baell, J. Bennett, J. Blagg, C. Bountra, P. E. Brennan, P. J. Brown, M. E. Bunnage, C. Buser-Doepner, R. M. Campbell, A. J. Carter, P. Cohen, R. A. Copeland, B. Cravatt, J. L. Dahlin, D. Dhanak, A. M. Edwards, M. Frederiksen, S. V. Frye, N. Gray, C. E. Grimshaw, D. Hepworth, T. Howe, K. V. Huber, J. Jin, S. Knapp, J. D. Kotz, R. G. Kruger, D. Lowe, M. M. Mader, B. Marsden, A. Mueller-Fahrnow, S. Muller, R. C. O'Hagan, J. P. Overington, D. R. Owen, S. H. Rosenberg, B. Roth, R. Ross, M. Schapira, S. L. Schreiber, B. Shoichet, M. Sundstrom, G. Superti-Furga, J. Taunton, L. Toledo-Sherman, C. Walpole, M. A. Walters, T. M. Willson, P. Workman, R. N. Young, and W. J. Zuercher, "The promise and peril of chemical probes," *Nat Chem Biol*, vol. 11, no. 8, pp. 536–41, 2015.
- [114] B. K. Shoichet, "Screening in a spirit haunted world," *Drug Discov Today*, vol. 11, no. 13-14, pp. 607–15, 2006.
- [115] J. B. Baell and G. A. Holloway, "New substructure filters for removal of pan assay interference compounds (pains) from screening libraries and for their exclusion in bioassays," *J Med Chem*, vol. 53, no. 7, pp. 2719–40, 2010.

- [116] W. P. Walters and M. Namchuk, "Designing screens: how to make your hits a hit," *Nat Rev Drug Discov*, vol. 2, no. 4, pp. 259–66, 2003.
- [117] D. A. Gschwend, *DOCK version 3.5 reference manual*.
- [118] D. A. Pearlman, D. A. Case, J. C. Caldwell, G. L. S. GL, and U. C. S. et al. ., *AMBER 4.0*.
- [119] K. Gallagher and K. Sharp, "Electrostatic contributions to heat capacity changes of dna-ligand binding," *Biophys J*, vol. 75, no. 2, pp. 769–76, 1998.
- [120] K. A. Sharp, "Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear poisson–boltzmann model," *Biopolymers*, vol. 36, no. 2, pp. 227–243, 1995.
- [121] H. Zhao and A. Caflisch, "Molecular dynamics in drug design," *European Journal of Medicinal Chemistry*, vol. 91, no. SI, pp. 4–14, 2015.
- [122] A. Babakhani, T. T. Talley, P. Taylor, and J. A. McCammon, "A virtual screening study of the acetylcholine binding protein using a relaxed-complex approach," *Comput Biol Chem*, vol. 33, 2009.
- [123] M. Totrov and R. Abagyan, "Flexible ligand docking to multiple receptor conformations: a practical alternative," *Current Opinion in Structural Biology*, vol. 18, no. 2, pp. 178–184, 2008.
- [124] J. Hritz, A. de Ruyter, and C. Ostenbrink, "Impact of Plasticity and Flexibility on Docking Results for Cytochrome P450 2D6: A Combined Approach of Molecular Dynamics and Ligand Docking," *Journal of Medicinal Chemistry*, vol. 51, no. 23, pp. 7469–7477, 2008.
- [125] S. von Grafenstein, H. G. Wallnoefer, J. Kirchmair, J. E. Fuchs, R. G. Huber, M. Schmidtke, A. Sauerbrei, J. M. Rollinger, and K. R. Liedl, "Interface dynamics explain assembly dependency of influenza neuraminidase catalytic activity," *Journal of Biomolecular Structure & Dynamics*, vol. 33, no. 1, pp. 104–120, 2015.

## 13 Appendix

### 13.1 Calculation of aMD Parameters

All parameters were calculated according to the formulas below.

$$E_{\text{thresh}D} = \bar{E}_{\text{DIHED}} + a_1 \cdot N_{\text{RES}} \quad (13.1)$$

$$\alpha_D = a_2 \cdot \frac{N_{\text{RES}}}{5} \quad (13.2)$$

$$E_{\text{thresh}P} = \bar{E}_{\text{TOT}} + b_1 \cdot N_{\text{ATOMS}} \quad (13.3)$$

$$\alpha_P = b_2 \cdot N_{\text{ATOMS}} \quad (13.4)$$

$\bar{E}_{\text{DIHED}}$  is the average dihedral and  $\bar{E}_{\text{TOT}}$  the total potential energy resulting from previous cMD simulations.  $N_{\text{RES}}$  and  $N_{\text{ATOMS}}$  are the number of residues and atoms respectively, in each system. The variables  $a_1, a_2, b_1$  and  $b_2$  were altered systematically for each system to optimize the boosting level. The applied values for the aMD simulations are shown in table 13.1. For Bet v 1a we

**Table 13.1:** Parameters for aMD simulation of each system.

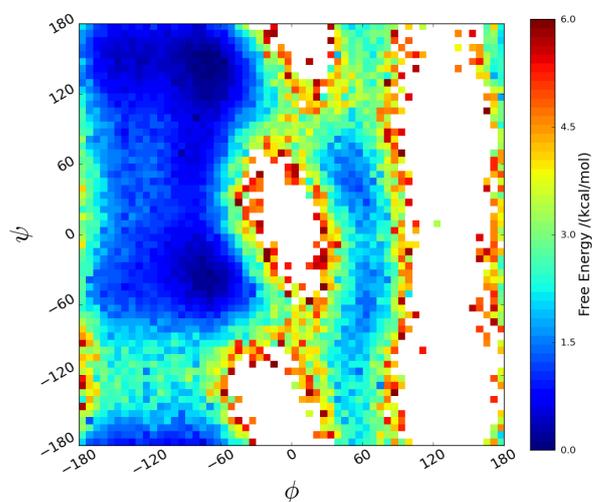
System	$a_1$	$a_2$	$E_{\text{thresh}D}$	$\alpha_D$	$b_1$	$b_2$	$E_{\text{thresh}P}$	$\alpha_P$
Di-Ala	3.5	3.5	18.7	1.4	0.175	0.175	-343	532.2
BPTI	4	4	837.8	46.4	0.16	0.16	-40754	2198.6
Bet v 1a	3	3	2189.2	95.4	0.30	0.16	-73135	4333.6

systematically tested 5 sets of boosting parameters. Further we set up an aMD simulation in which we only boosted the dihedral potential. We started with a set of boosting parameter that we considered to only cause a slight boosting effect and slowly increased the intensity of the boosting until the protein would unfold (Set 5). We chose to continue with the most aggressive boosting parameters, that would not unfold the protein after 1  $\mu$ s of aMD simulation. Despite the awareness of the resulting increase in inaccuracy we tried to maximize the effect of the aMD method. We considered this approach to give us the most information on potential slow conformational changes in Bet v 1a, as well as on the robustness of our metric.

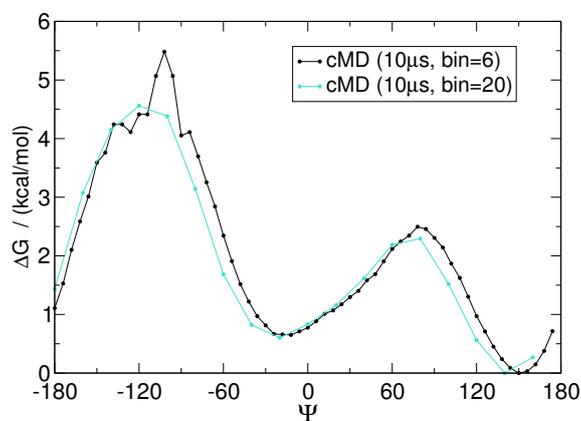
**Table 13.2:** Parameters tested for aMD simulation of Bet v 1a.

Set	$a_1$	$a_2$	$E_{\text{threshD}}$	$\alpha_D$	$b_1$	$b_2$	$E_{\text{threshP}}$	$\alpha_P$
1	4	4	2348.2	127.2	0.30	0.20	-73135	8125.5
2	4	4	2348.2	127.2	0.20	0.20	-75843	5417
3	3	2	2189.2	95.4	0.20	0.20	-75843	5417
4	3	3	2189.2	95.4	0.30	0.16	-73135	4333.6
dh	3	3	2189.2	95.4	-	-	-	-

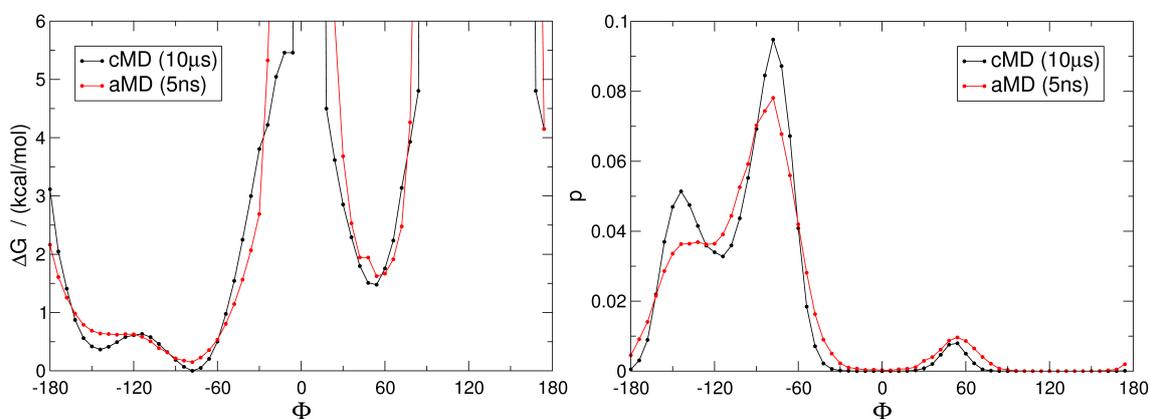
## 13.2 Alanine Dipeptide



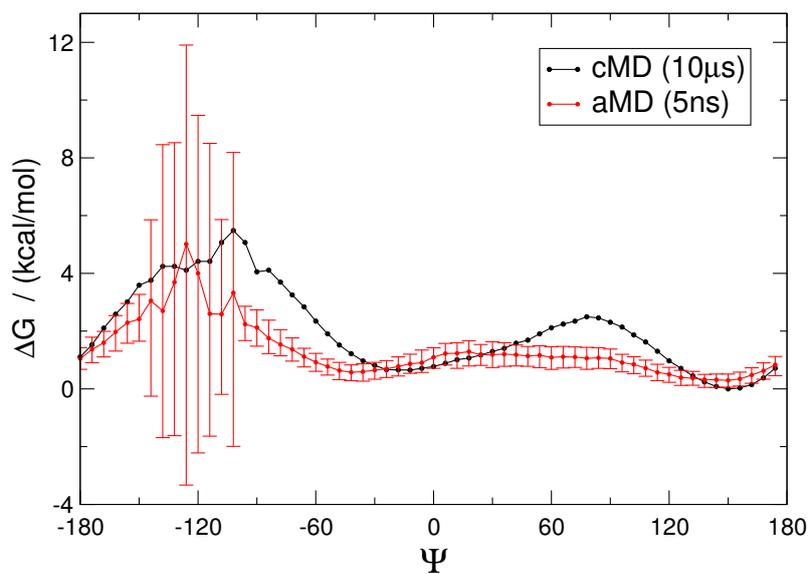
**Figure 13.1:** Conformational space sampled in 1  $\mu$ s of aMD. Blue regions indicate the most favorable states with the lowest energy. Unfavorable torsional states with a free energy higher than 6 kcal/mol (red) are depicted in white.



**Figure 13.2:** Free energy of  $\Psi$  calculated using a bin size of  $6^\circ$  (black) and  $20^\circ$  (turquoise). The increase of bin sizes smoothens the free energy surface but causes a shift of minima. The jaggedness of the profile using a bin width of  $6^\circ$  is most likely due to the limited number of recorded frames (100 000).

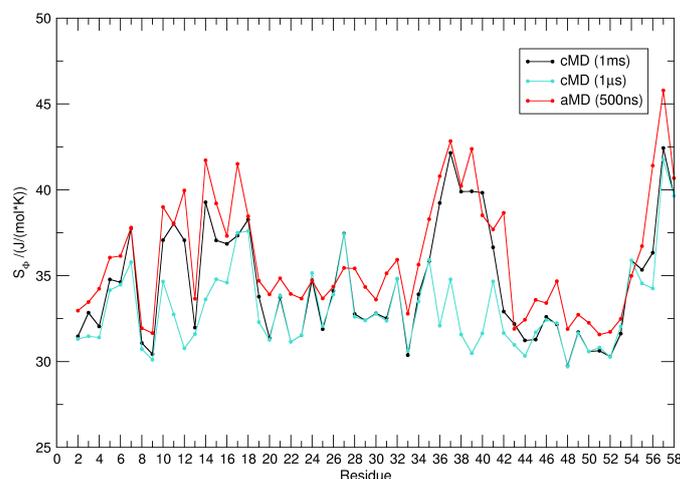


**Figure 13.3:** Free energy and state populations of  $\Phi$  in Di-Ala. Left: Free energy distribution of  $\Phi$  from a  $10 \mu\text{s}$  cMD (black) and  $5 \text{ ns}$  aMD (red) simulation of Di-Ala. Rarely or not visited dihedral states showing highly unfavorable free energies were cut off at  $6 \text{ kcal/mol}$ . Right: State populations calculated from the free energies of  $\Phi$  as shown in Figure 1



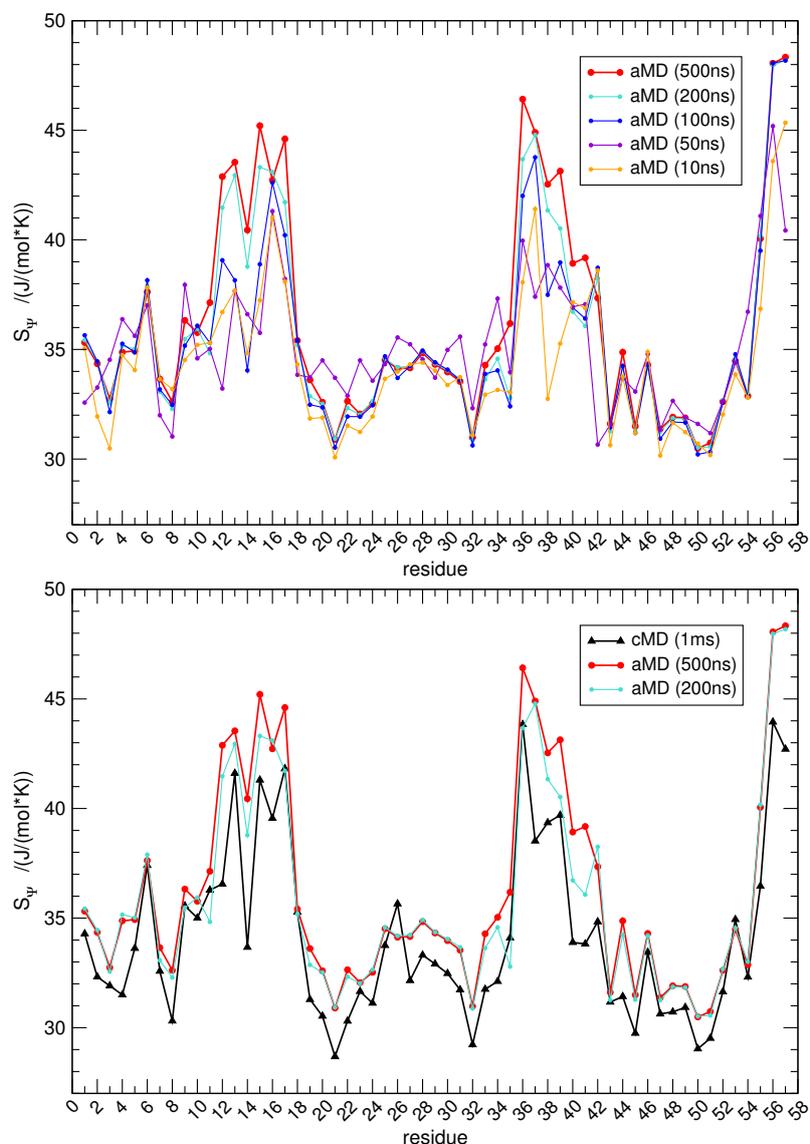
**Figure 13.4:** Reweighted free energies of Di-Ala. A  $1 \mu\text{s}$  trajectory was split into 200 segments of 5 ns each (2 500 frames) using the segments for averaging. (red) The free energy landscape of  $\Psi$  was reconstructed using reconstructed using Maclaurin series in the reweighting protocol. As a reference the free energy surface of  $\Psi$  from a  $10 \mu\text{s}$  cMD simulation (black) is shown. The standard deviation of the aMD trajectory shows the stability of the results in well sampled areas and highlights strong fluctuations for less sampled ones.

## 13.3 BPTI

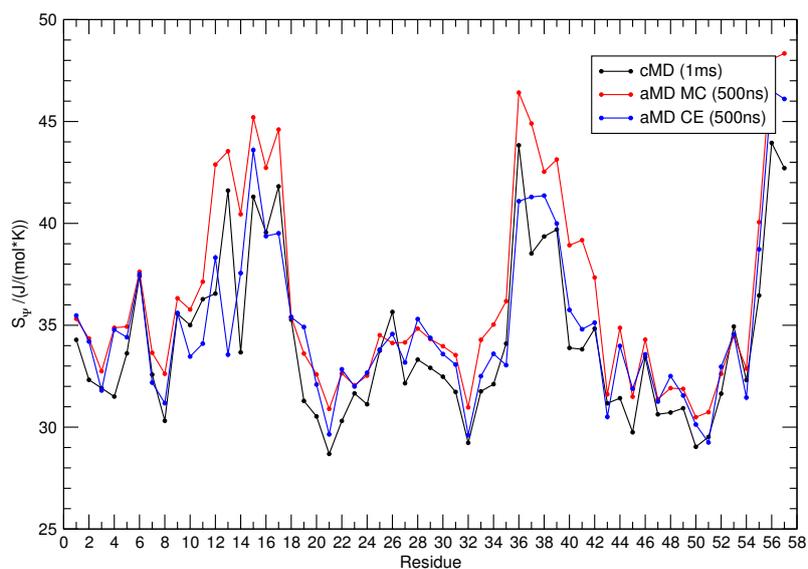


**Figure 13.5:** Comparing local flexibility of BPTI captured in cMD and aMD simulations. Residue-wise dihedral entropies  $S_{\Phi}$  from a 1 ms cMD simulation (black) and 500 ns aMD simulation of BPTI (red) show remarkable rank correlation. Local flexibility observed in a 1  $\mu$ s cMD simulation (turquoise) clearly differs from the aMD results.

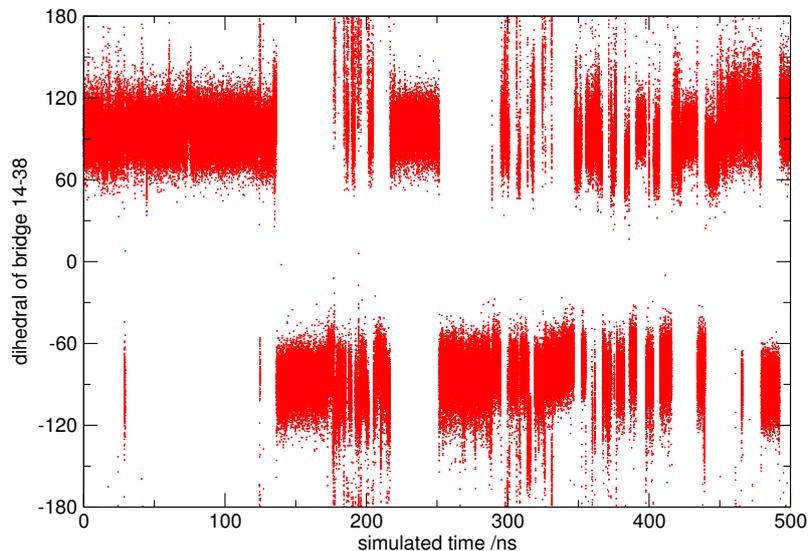
Comparing dihedral entropies from  $\Phi$  and  $\Psi$  we find a Spearman rank correlation  $r=0.77$  between  $S_{\Psi}$  and  $S_{\Phi}$  of BPTI. For Bet v 1a we observe correlation of  $r=0.86$ . These results support the assumption of a similar extent of motions captured in both backbone dihedrals phi and psi. When considering the information displayed in Ramachandran plots of single amino acids, the  $\Psi$ -axis generally shows a broader distribution than the  $\Phi$ -axis.<sup>62</sup> So most amino acids secondary structure elements, such as alpha-helices and beta-sheets, can be distinguished solely by looking at the psi-distribution.<sup>63</sup> Hence, not the whole backbone dynamics are reflected by the  $\Psi$  angle.  $\Phi$  dihedral distribution were calculated as well. Yet, for the representation of protein dynamics based on dihedral entropies we prioritized  $\Psi$  over  $\Phi$  as it captures the backbone dynamics more comprehensively.



**Figure 13.6:** Benchmarking sampling time of BPTI. Dihedral entropies of BPTI were evaluated after 10, 50, 100, 200 and 500 ns aMD simulation time. TOP: 200 ns and 500 ns result in similar flexibility patterns, while shorter sampling runs capture only small increase of flexibility from residues 10–20 and 32–44. BOTTOM: Comparison of dihedral entropies from 200 ns (blue) and 500 ns (red) aMD to 1 ms cMD (black) sampling. In both aMD simulations the same regions are captured as flexible, yet after 500 ns the shape of the 1 ms simulation is reproduced more accurately.

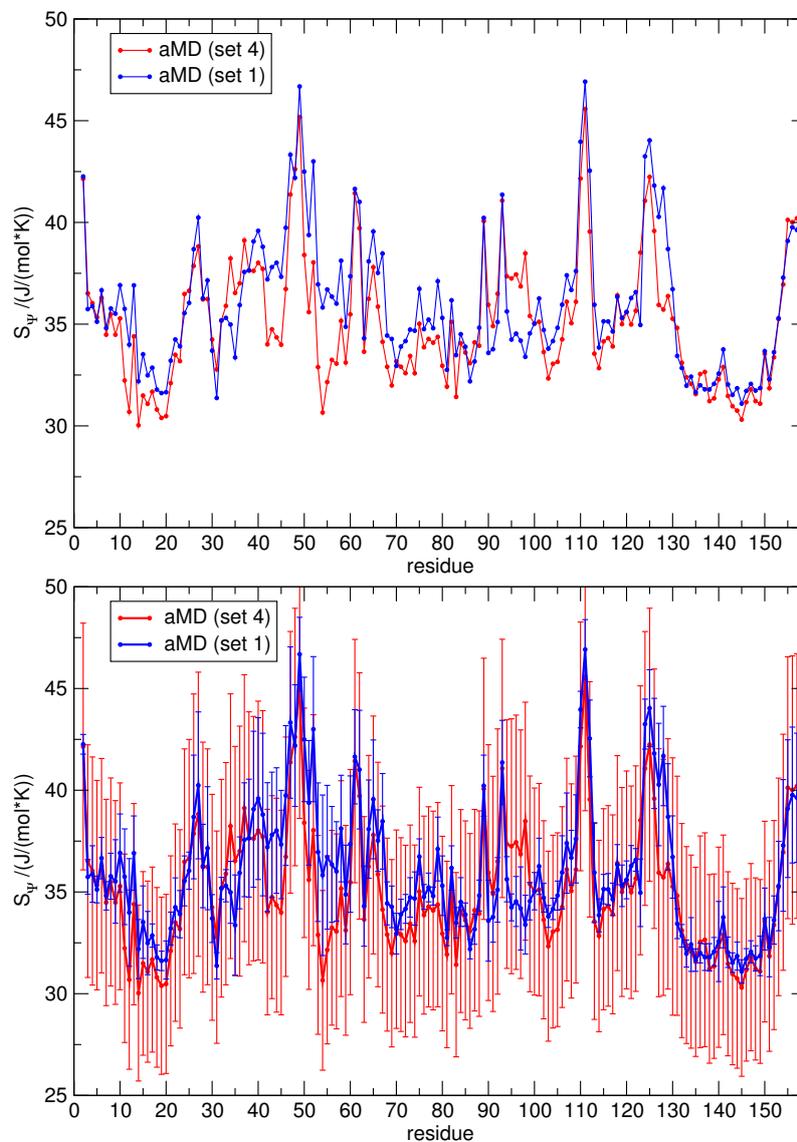


**Figure 13.7:** Probing different reweighting protocols. Dihedral entropies  $S_\Psi$  were calculated from the same 500 ns trajectory of BPTI using Maclaurin series (red) and cumulant expansion (blue) to approximate the exponential in the reweighting protocol. Using cumulant expansion we find a Spearman rank correlation of  $r=0.85$  between the 1 ms control (black) and 500 ns aMD simulation. Reweighting with Maclaurin series increases the correlation between aMD and cMD results to  $r=0.90$ .

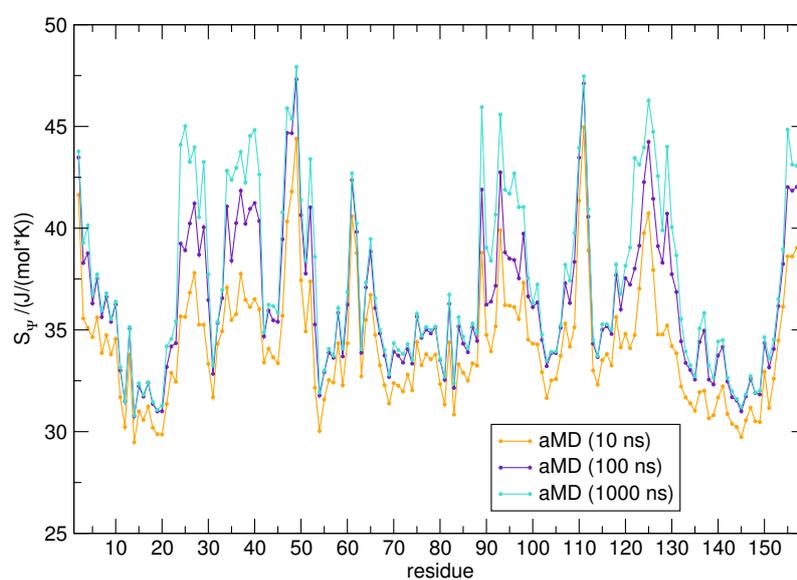


**Figure 13.8:** Isomerization of disulfide bridge CYS14-CYS38. During 500 ns aMD simulation the dihedral of the disulfidebridge between CYS14 and CYS38 switches multiple times between values around 100 degrees to -100 degrees. The two populated dihedral states represent the cis- and trans-conformation of the disulfide bridge.

## 13.4 Bet v 1a



**Figure 13.9:** Probing different levels of acceleration in Bet v 1a aMD simulations. Dihedral entropies  $S_{\Psi}$  for two sets of aMD parameter are shown ( $r=0.80$ ) one with less boosting (blue, set 1) and a more aggressively boosted one (red, set 4). The error shown in the bottom derives from trajectory splitting. The  $1 \mu\text{s}$  trajectories are split into 50 segments, resulting in the shown average and standard deviation representing 20 ns of aMD sampling. It is clear to see that the error of an aMD simulation is strongly dependent on the chosen boosting parameters, i.e. level of acceleration.



**Figure 13.10:** Benchmarking sampling time of Bet v 1a. Evaluation of local flexibility in Bet v 1 a on different time scales of aMD simulation. Most notable differences in the three aMD simulations are found from residue 15-45.