JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Independent Metrics for Protein Backbone and Side-Chain Flexibility: Time Scales and Effects of Ligand Binding

Julian E. Fuchs,[†,‡] Birgit J. Waldner,[†] Roland G. Huber,[†,§] Susanne von Grafenstein,[†] Christian Kramer,[†] and Klaus R. Liedl*[,†]

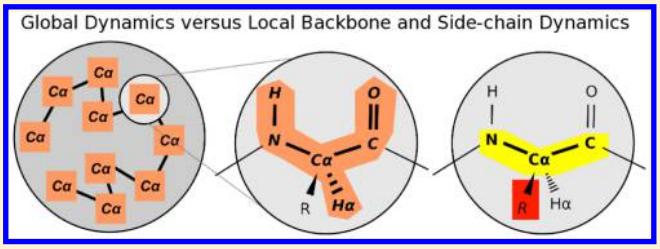
[†]Institute of General, Inorganic and Theoretical Chemistry, and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 80-82, A-6020 Innsbruck, Tyrol, Austria

[‡]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[§]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, Matrix #07-01, 138671 Singapore

Supporting Information

ABSTRACT: Conformational dynamics are central for understanding biomolecular structure and function, since biological macromolecules are inherently flexible at room temperature and in solution. Computational methods are nowadays capable of providing valuable information on the conformational ensembles of biomolecules. However, analysis tools and intuitive metrics that capture dynamic information from *in silico* generated structural ensembles are limited. In standard work-flows, flexibility in a conformational ensemble is represented through residue-wise root-mean-square fluctuations or B-factors following a global alignment. Consequently, these approaches relying on global alignments discard valuable information on local dynamics. Results inherently depend on global flexibility, residue size, and connectivity. In this study we present a novel approach for capturing positional fluctuations based on multiple local alignments instead of one single global alignment. The method captures local dynamics within a structural ensemble independent of residue type by splitting individual local and global degrees of freedom of protein backbone and side-chains. Dependence on residue type and size in the side-chains is removed via normalization with the B-factors of the isolated residue. As a test case, we demonstrate its application to a molecular dynamics simulation of bovine pancreatic trypsin inhibitor (BPTI) on the millisecond time scale. This allows for illustrating different time scales of backbone and side-chain flexibility. Additionally, we demonstrate the effects of ligand binding on side-chain flexibility of three serine proteases. We expect our new methodology for quantifying local flexibility to be helpful in unraveling local changes in biomolecular dynamics.



INTRODUCTION

Biological macromolecules persistently undergo conformational transitions at room temperature. Random thermal motions involve positional fluctuations of individual atoms, bonds, functional groups, side-chains, local regions of the backbone, secondary structure elements, domains, and entire folded regions.[1] Exploring the conformational space sampled by a protein is of crucial importance for understanding its function.[2−4]

A steadily growing pool of techniques is available to experimentally probe protein dynamics: It includes spectroscopic techniques such as nuclear magnetic resonance (NMR) spectroscopy[5] or fluorescence correlation spectroscopy.[6] Hydrogen−deuterium exchange mass spectrometry allows capturing the dynamics of proteins on different time scales.[7] B-factors (temperature factors, thermal factors, Debye−Waller factors) from X-ray structures contain limited dynamic information. X-ray B-factors are calculated as fitting parameters in structure refinement. Poorly defined atomic positions lead to higher B-factors. Crystallographic B-factors can be modeled as mean square displacements of atoms.[8] It was shown that B-factors reflect dynamic regions in a protein comparable to NMR experiments.[9,10] Nevertheless, B-factors only capture parts of protein dynamics as they do not include larger loop motions which are prevented in the crystal lattice. Moreover, X-ray temperature factors heavily depend on quality and experimental conditions of the underlying crystal structure,[11] thus impairing direct quantitative interpretation. Normalization strategies have been followed allowing for a rough comparison of local dynamics in proteins.[12]

All available experimental approaches are limited in their resolution concerning system size and/or time scales. Therefore, computational investigation of biomolecular flexibility in terms of molecular dynamics (MD) simulations is an attractive alternative.[13] To this day, MD simulations allow for routinely

Received: July 18, 2014
Published: February 12, 2015

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

tracing protein dynamics in solution and at full atomistic level on the femto- to microsecond scale.[14] These time scales cover vibrations as well as side-chain rotations in proteins.[15] Recently, large scale microsecond trajectories obtained on standard computer clusters were reported,[16] as well as even millisecond simulations through the use of dedicated simulation hardware.[17] Acquired MD simulation data are important in the analysis of protein dynamics and function.[18]

Degrees of freedom within a polypeptide chain are commonly split into contributions of the peptide backbone and the side-chains. Occurring torsional states can be depicted in Ramachandran plots[19] and Janin plots,[20] respectively. Side-chains may adopt several rotamer states at room temperature,[21] which is crucial e.g. for ligand binding.[22,23] Accessible rotamer states are stored in residue-wise rotamer libraries[24−26] and frequently used for structure validation.[27] Recently, discrete rotamer libraries were extended to continuous probabilistic models for side-chain orientations.[28]

Both experimental and computational approaches provide access to structural ensembles and can be utilized for quantification of global and local flexibility. A standard metric for flexibility are root-mean-square fluctuations and related B-factors available from X-ray diffraction as well as MD simulations.[29] Alternatively, order parameters of internal vectors within a protein can be assessed by NMR experiments and simulation.[30,31] B-factors and order parameters are directly available from most simulation packages and their related analysis and visualization tools.[32−37]

Another alternative is the analysis of torsion angle distributions within protein backbone and side-chains. T-analyst is a user-friendly tool for the analysis of torsional degrees of freedom.[38] Still, the current implementation is limited to a fixed bin width of 1° and one-dimensional analysis, thereby neglecting crucial couplings between degrees of freedom.[39] Recently, parameter-free kernel density estimation has been successfully used to generate continuous probability functions for entropy estimation while avoiding to choose an arbitrary histogram bin width.[40] In the context of protein flexibility, resulting alignment-independent conformational entropies were applied to study thermal stabilities.[41]

Biomolecular flexibility can also be assessed by counting of constraining interaction networks reducing macromolecular structures through graph theory.[42,43] In general, caution has to be exercised when attempting to quantify a system's local flexibility, as time scales and hence sampling might vary over several orders of magnitude.[44]

Except for torsional analysis, all described metrics for flexibility depend on alignment and residue size. We introduce a new metric that bypasses size dependence by normalization to the isolated amino acids and splits individual degrees of freedom within a system by performing multiple local alignments. Thereby, global and local B-factors that are comparable between different amino acids are generated for both backbone and side-chains. We demonstrate their validity by investigating molecular flexibility within molecular dynamics derived structural ensembles of bovine pancreatic trypsin inhibitor (BPTI) and explore the impact of ligand binding on side-chain flexibility of three serine proteases. As dynamics are a key factor in protease recognition,[45] we expect metrics for local backbone and side-chain flexibility helpful in rationalizing differences in binding specificity.

METHODS

Molecular Dynamics Simulations of Unrestrained Amino Acids. We performed 22 independent molecular dynamics simulations of isolated amino acids using the simulation package Amber12.[34] Histidine was included in two tautomeric states with hydrogen on δ- or ε-position, respectively (as Amber types HID and HIE) as well as in the protonated form (HIP). In order to represent a peptide-bound state and avoid perturbation by terminal charges, we capped amino acids with an N-terminal acetyl group (ACE) and a C-terminal *N*-methylamine (NME) group. Peptide topologies in extended conformation were generated using tleap of AmberTools[34] and solvated with a truncated octahedral TIP3P[46] water box with minimum wall distance of 12 Å. Peptide parameters were taken from the AMBER force field 99SB-ILDN.[47]

Simulations were performed in the NpT ensemble employing a 8 Å van der Waals cutoff at a temperature of 300 K maintained by the Langevin thermostat.[48] Long range electrostatics were included through Particle Mesh Ewald.[49] We employed an extensive equilibration protocol[50] followed by 1 μs unrestrained sampling using the CUDA implementation of pmemd[51] achieving a performance of approximately 0.25 μs per day on a single GeForce GTX 680 GPU. Bonds involving hydrogens were constrained using the SHAKE algorithm[52] to allow a time step of 2 fs. We saved 10000 evenly spaced snapshots per trajectory.

Molecular Dynamics Simulations of Restrained Amino Acids. Starting from freely equilibrated systems, we performed restrained molecular dynamics simulations, where phi and psi backbone dihedrals of the model peptide were restrained independently to maintain a helical or extended conformation. We applied flat bottom restraint potentials to allow free fluctuations within these allowed regions. Around the flat bottom we employed a steep repulsive harmonic potential with force constant 200 kcal/(mol*deg). Following definitions of Hovmöller et al.,[53] we restrained the phi angle of helical peptides in the range of −180° to 0° and the psi angle within −100° and 45°. Model peptides in extended (β-sheet) conformation were restrained to phi between −180° and −45° and psi between 45° and 225°. Other simulation parameters were identical to unrestrained simulations.

Molecular Dynamics Simulation of BPTI. We reanalyzed a molecular dynamics trajectory of BPTI covering a sampling time of 1.03 ms stored as 4140000 snapshots kindly provided by D. E. Shaw Research.[17] The simulation was performed at 300 K based on the joint neutron/X-ray refined structure of BPTI with PDB ID 5PTI.[54]

Molecular Dynamics Simulation of Serine Proteases. We performed MD simulations of factor Xa, thrombin, and trypsin based on X-ray structures of Nar et al. (PDB: 1G2L, 1G30, 1OYQ).[55] In addition to simulations of the three serine proteases in complex with the ligand BIBT0871, we created apo systems by deletion of the ligand resulting in six topologies.

After removal of bound sulfate ions and sulfate modifications at tyrosine residues, we prepared the respective heavy protease chains in MOE[56] using protonate3D.[57] The ligand was parametrized using the Generalized Amber Force Field (GAFF).[58] Automated atom typing was overridden for the benzamidine moiety, where aromatic atoms were set to enforce planarity. Point charges were derived by RESP fitting[59] at the HF-6/31G* level using Gaussian03.[60] Ca^{2+} ions were included

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

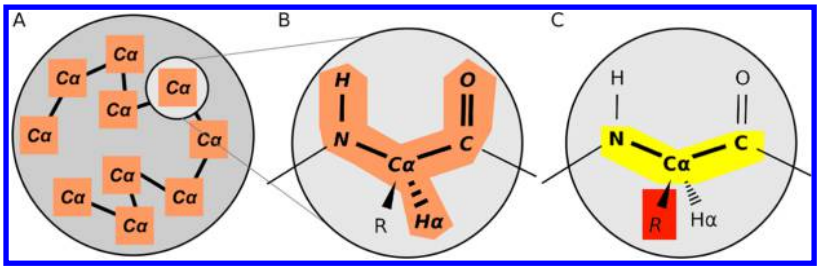


Figure 1. Schematic diagram for different flexibility metrics: Alignment is performed on atoms in yellow and italics; B-factors are calculated on atoms in red and bold highlighting. Orange indicates atoms that are included in both alignment and B-factor calculation. Part A schematically shows a global alignment based on all Cα atoms in the system. Parts B and C both depict residue-wise local alignments. Part B schematically shows the calculation of local backbone B-factors based on multiple residue-wise alignments. Part C schematically shows the calculation of B-factors for side-chains that are calculated after alignment to the backbone atoms N, Cα, and C (in IUPAB nomenclature).

using standard parameters for the AMBER force field.[61] Following careful equilibration[50] 1 μs unrestrained simulations for all six systems were performed using settings and parameters identical to simulations of capped amino acids.

Calculation of Flexibility Metrics. We used ptraj and cpptraj from AmberTools for analysis of simulation trajectories.[33] In order to quantify global movements within BPTI, we calculated B-factors for Cα-atoms after one single global alignment to all Cα-atoms of the protein starting structure. Similarly, we calculated all-atom B-factors for the model peptides after a global alignment to all atoms of the capped amino acids. All calculated B-factors were derived from the observed mean square fluctuation (MSF) of the respective atom groups. The MSF is the mean squared distance of an atom x_i from a reference position $\bar{x}$ across all N frames as given in eq 1.

$$\mathrm{MSF} = \frac{1}{N}\sum_{i=t_0}^{N}(x_i - \bar{x})^2 \tag{1}$$

The B-factor is then calculated from MSF by multiplication with a constant factor according to eq 2.

$$\text{B-factor} = \frac{8\pi^2}{3}\mathrm{MSF}\ [\text{Å}^2] \tag{2}$$

A local metric for backbone flexibility was generated via multiple local alignments to all backbone atoms including hydrogens of the residue of interest. Mass-weighted B-factors were calculated for those atom groups after alignment to the starting structure of the single residue, thereby yielding a metric for local backbone flexibility.

In order to trace side-chain movements, we locally aligned the three heavy atoms in the peptide main chain (N, Cα, C in IUPAB nomenclature) of the residue of interest of all snapshots to the starting structure. Based on the minimal set of three atoms required for alignment in three-dimensional space, this approach minimizes the influence of backbone motions on the split dynamics of the side-chain. We recorded side-chain flexibility as mass-weighted B-factor of all side-chain atoms including hydrogens. Thus, for glycine a small B-factor of the side-chain could also be observed. Figure 1 shows a schematic diagram of the approaches followed within the three flexibility metrics (global Cα, local backbone, local side-chain). Statistical metrics for B-factors of all model peptides were calculated as average and standard deviation over 10 trajectory parts corresponding to 100 ns each. Our metric explicitly monitors fluctuations of chemically equivalent groups (e.g., terminal methyl groups in valine), including these movements as a contribution to molecular flexibility. This is thermodynamically valid, as internal rotations significantly contribute to the entropy of amino acid side-chains.[62]

We used the mass-weighted side-chain B-factors of unrestrained simulations of isolated amino acids as normalization factors to correct for different residue connectivities and sizes according to eq 3.

$$\begin{aligned}&\text{normalized side-chain flexibility}\ [\,]\\ &\quad = \text{side-chain B-factor}\ [\text{Å}^2]/\text{side-chain B-factor}_{\text{isolated}}\ [\text{Å}^2]\end{aligned} \tag{3}$$

After division by the side-chain B-factor of the isolated amino acid, a dimensionless relative local side-chain flexibility of 1.0 shows side-chain mobility equal to the isolated amino acid, whereas lower values imply conformational restriction by the side-chain's environment. Values exceeding 1.0 indicate that amino acid side-chains are either mobilized by the protein environment or show some degree of geometrical distortion compared to the isolated state.

For comparison with an alignment-free flexibility metric, we calculated dihedral entropies over distributions of phi and psi torsion angles in the backbone and chi torsion angles in the side-chain of amino acids using T-analyst.[37] Further geometric parameters were extracted from simulation trajectories, i.e., Ramachandran plots, two-dimensional RMSD (2D-RMSD) plots, and solvent-accessible surface areas using the LCPO algorithm.[63] These analyses were carried out using ptraj and cpptraj.[33]

RESULTS

Molecular Dynamics Simulations of Unrestrained Amino Acids. After equilibration we obtained stable molecular dynamics trajectories for all 22 solvated and capped amino acids. A sampling time of 1 μs split into 10 parts shows convergence of all calculated flexibility metrics (see Table 1). For the case of the flexible lysine residue we demonstrate convergence through 2D-RMSD plots and time-resolved Ramachandran plots as well as through extension of the sampling time to 10 μs (see the Supporting Information for more details). B-factors calculated from the extended trajectory fall within the error bars of the 1 μs simulation, thus 1 μs simulations are considered sufficient to achieve convergence of unrestrained B-factors.

Calculated B-factors for isolated amino acids show how the overall flexibility of the amino acid calculated as all-atom B-factor can be split into the degrees of freedom of the backbone and the side-chain. Backbone contributions are almost constant

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

Table 1. Flexibility of Unrestrained Isolated Amino Acids[a]

amino acid	B-factor all-atom [Å^2]	B-factor backbone [Å^2]	B-factor side-chain [Å^2]
ALA	19.3 (0.7)	18.2 (2.3)	6.17 (0.03)
ARG	37.5 (0.9)	18.1 (1.1)	193.6 (9.2)
ASN	34.3 (0.8)	17.1 (2.2)	69.3 (2.3)
ASP	37.2 (1.3)	17.7 (3.2)	73.5 (4.9)
CYS	27.5 (1.2)	17.6 (2.3)	42.9 (4.1)
GLN	32.6 (1.0)	13.4 (3.3)	97.0 (5.4)
GLU	37.0 (1.5)	17.9 (2.0)	94.0 (4.4)
GLY	25.7 (0.5)	22.6 (3.8)	0.41 (0.01)
HID	40.3 (1.0)	18.8 (6.1)	154.1 (4.2)
HIE	34.5 (2.4)	15.8 (2.4)	134 (13)
HIP	37.0 (1.1)	15.2 (3.4)	153.1 (6.3)
ILE	33.6 (2.1)	15.6 (3.2)	60.2 (2.3)
LEU	25.3 (1.2)	18.5 (1.5)	40.1 (4.1)
LYS	27.4 (0.6)	18.0 (2.3)	121.7 (6.4)
MET	35.7 (1.9)	15.2 (2.4)	124.5 (5.3)
PHE	43.2 (0.9)	16.1 (1.2)	194.8 (5.1)
PRO	10.3 (0.5)	11.7 (2.4)	6.31 (0.10)
SER	22.0 (0.5)	19.0 (3.0)	21.2 (0.7)
THR	22.9 (1.5)	19.7 (2.4)	21.0 (3.6)
TRP	58.8 (3.1)	18.8 (2.8)	295 (11)
TYR	44.0 (3.0)	15.9 (3.9)	246 (17)
VAL	24.3 (2.1)	16.7 (2.2)	25.6 (4.3)

[a]B-factors for all atoms, backbone and side-chain atoms extracted from 1 μs unrestrained simulations of capped amino acids. Averages and standard deviations (shown in brackets) were calculated over 10 trajectory parts (100 ns each). Side-chain B-factors in the last column are used for normalization of side-chain B-factors of protein simulations to yield dimensionless relative side-chain flexibilities.

for all amino acids, only glycine (more flexible) and proline (more rigid) follow expected trends.[64] These data agree with data from T-analyst where glycine shows highest phi and psi entropies and proline lowest among all amino acids.

Within the side-chains we observe a strong correlation of flexibility with residue size in the B-factor metric over all side-chain atoms. Therefore, we use the B-factors for unrestrained amino acids given in Table 1 to normalize side-chain B-factors in protein simulations. Furthermore, side-chain B-factors strongly correlate to entropies over chi1 torsions (Spearman rank correlation r = 0.91). As our metric is based on an alignment optimizing the exit vector into the side-chain this is expected for geometric reasons. The number of torsions as well as the sum of all chi entropies show lower correlation (both r = 0.76, with an internal correlation of both metrics of r = 0.95). Here, aromatic residues with larger side-chains and side-chain B-factors but lower amount of torsions decrease correlation of the metrics. Interestingly, the number of atoms itself shows a stronger correlation (r = 0.80) than these torsion-dependent approaches.

Molecular Dynamics Simulations of Restrained Amino Acids. We trapped the simulated model peptides in helical or extended state by imposing a restraint on phi and psi backbone dihedrals (see Table 2 and Table 3). Exemplary Ramachandran plots for simulations of lysine illustrate the restriction of sampled conformational space (see the Supporting Information).

As expected, all-atom as well as backbone B-factors are lower with these restricted degrees of freedom. Backbone B-factors are constant for all amino acids with around 2 Å^2 for helical and 3 Å^2 for extended (β-sheet) state, respectively. This is due to

Table 2. Flexibility of Isolated Amino Acids Restrained to α-Helical Backbone Conformation[a]

amino acid	B-factor all-atom [Å^2]	B-factor backbone [Å^2]	B-factor side-chain [Å^2]
ALA	4.47 (0.13)	2.05 (0.07)	6.16 (0.02)
ARG	33.2 (1.4)	2.05 (0.09)	202 (6.6)
ASN	24.8 (1.0)	2.17 (0.09)	68.1 (3.6)
ASP	21.7 (1.5)	1.79 (0.06)	57.4 (4.5)
CYS	16.4 (1.0)	1.92 (0.08)	43.8 (2.5)
GLN	25.3 (0.8)	2.05 (0.05)	95.6 (6.3)
GLU	26.6 (1.1)	1.76 (0.06)	92.4 (2.8)
GLY	3.97 (0.20)	2.33 (0.10)	0.42 (0.01)
HID	33.8 (1.1)	2.02 (0.10)	156 (2.1)
HIE	30.9 (1.3)	2.08 (0.09)	155 (3.8)
HIP	31.6 (1.0)	2.20 (0.09)	148 (7.2)
ILE	25.4 (2.3)	1.89 (0.09)	60.4 (4.7)
LEU	17.3 (0.9)	1.93 (0.07)	43.4 (4.2)
LYS	21.3 (0.9)	2.07 (0.09)	124 (7.1)
MET	27.3 (0.7)	2.02 (0.06)	125 (5.8)
PHE	39.0 (2.6)	2.03 (0.06)	197 (4.7)
PRO	2.92 (0.07)	1.00 (0.05)	6.61 (0.18)
SER	8.24 (0.17)	1.88 (0.07)	20.4 (0.42)
THR	6.45 (0.59)	1.64 (0.07)	11.4 (1.3)
TRP	52.4 (5.8)	2.02 (0.42)	262 (47)
TYR	42.5 (1.9)	2.01 (0.05)	271 (4.5)
VAL	8.91 (1.90)	1.51 (0.05)	18.1 (3.6)

[a]B-factors were calculated for all atoms, backbone and side-chains based on 10 trajectory parts (100 ns each) with standard deviation in brackets.

Table 3. Flexibility of Isolated Amino Acids Restrained to Extended (β-Sheet) Conformation[a]

amino acid	B-factor all-atom [Å^2]	B-factor backbone [Å^2]	B-factor side-chain [Å^2]
ALA	4.94 (0.15)	3.16 (0.10)	6.15 (0.02)
ARG	30.3 (1.0)	3.15 (0.16)	185 (7.5)
ASN	24.7 (0.7)	3.21 (0.27)	68.6 (2.0)
ASP	26.8 (2.8)	3.07 (0.19)	75.5 (8.0)
CYS	17.0 (1.1)	3.29 (0.17)	45.5 (3.1)
GLN	24.5 (0.7)	2.89 (0.21)	97.5 (4.0)
GLU	27.3 (1.0)	2.99 (0.16)	96.6 (4.9)
GLY	7.28 (0.46)	4.90 (0.21)	0.40 (0.01)
HID	31.9 (1.0)	2.90 (0.16)	151 (6.5)
HIE	22.7 (1.8)	2.94 (0.21)	107 (13)
HIP	30.3 (0.7)	3.44 (0.22)	147 (5.4)
ILE	22.1 (1.4)	2.16 (0.14)	55.6 (2.9)
LEU	14.9 (1.0)	2.78 (0.13)	38.1 (4.5)
LYS	18.9 (0.7)	2.98 (0.27)	118 (4.3)
MET	26.5 (0.9)	3.05 (0.16)	123 (6.1)
PHE	34.9 (0.9)	2.76 (0.15)	189 (5.4)
PRO	3.01 (0.07)	0.92 (0.10)	6.16 (0.10)
SER	9.17 (0.31)	3.61 (0.22)	21.5 (0.8)
THR	14.9 (1.3)	2.99 (0.32)	29.4 (2.7)
TRP	52.7 (1.6)	2.68 (0.17)	296 (5.8)
TYR	36.7 (1.8)	2.72 (0.12)	238 (18.6)
VAL	14.2 (1.9)	2.34 (0.15)	32.4 (4.3)

[a]B-factors were calculated for all atoms, backbone and side-chains based on 10 trajectory parts (100 ns each) with standard deviation in brackets.

the broader conformational space accessible for β-sheets as imposed by the dihedral definitions.[53] Side-chain flexibility as

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

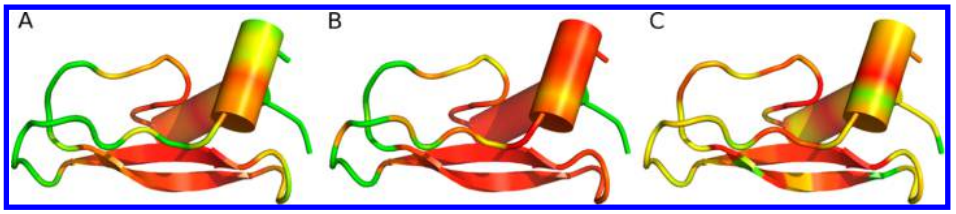


Figure 2. Global and local flexibility in BPTI: Residue-wise B-factors are color-coded on a cartoon representation of BPTI on a color ramp from red (rigid) over yellow to green (flexible). Part A shows global movements via calculation of B-factors for Cα atoms after a global alignment (red: 4 Å^2, green: ≥50 Å^2). Parts B and C show local movements in BPTI: backbone flexibility is shown in Part B (red: 0 Å^2, green: ≥2 Å^2), normalized side-chain flexibility in part C (red: 0, green: ≥2).

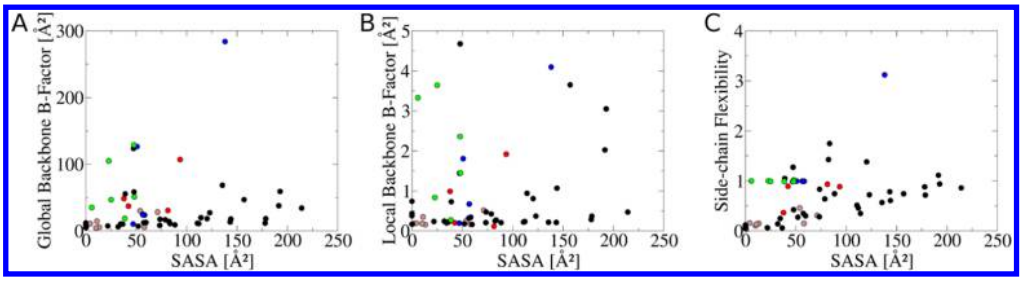


Figure 3. Correlation of flexibility metrics and SASA in BPTI: Residue-wise B-factors are correlated to the trajectory average of the solvent-accessible surface area (SASA). Aromatic residues (Phe, Tyr) are highlighted in brown, proline residues in red, alanine in blue, glycine in green. Part A shows global flexibility as global Cα B-factors; Parts B and C depict local flexibility as B-factors for backbone (B) and side-chain flexibility (C).

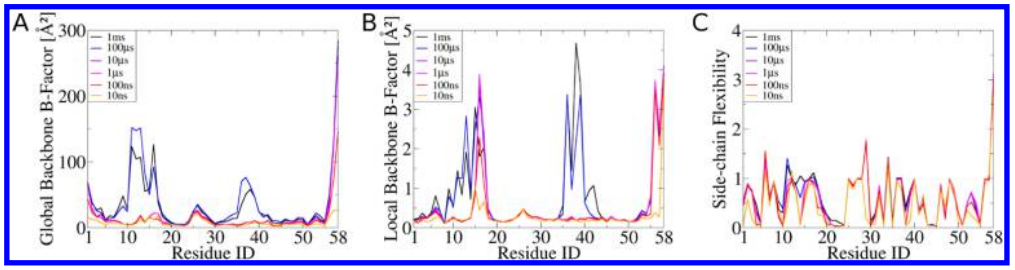


Figure 4. Time scales of flexibility in BPTI: Flexibility metrics are calculated from parts of the molecular dynamics trajectory, allowing for the estimation of required sampling time. Covered time scales range from 10 ns (orange) via several orders of magnitude in red to blue shading to the full trajectory length of 1 ms (black). Part A shows convergence of global Cα B-factors; parts B and C depict local flexibility as B-factors for backbone (B) and side-chain flexibility (C), respectively. Flexibility of side-chains converges earlier than movements of the protein backbone in BPTI.

an independent degree of freedom is preserved upon confinement of the backbone conformation. Thus, the side-chain B-factors observed in these simulations are very similar to side-chain B-factors of isolated amino acids.

Small deviations from this behavior can be observed for threonine. Extension of sampling time to 10 μs does not alleviate the observed deviation. Thus, we presume it is not an artifact due to insufficient sampling (B-factor$_{\text{All-atom}}$ = 22.0 Å^2, B-factor$_{\text{Backbone}}$ = 18.5 Å^2, B-factor$_{\text{Side-chain}}$ = 23.2 Å^2). This amino acid shows reduced side-chain mobility in helical state and increased mobility in extended conformation. As we observe no major change in intramolecular hydrogen bonding of the hydroxy function, we attribute this finding to the larger solvent exposure of the side-chain in extended conformation. Here, rotation of the side-chain is facilitated leading to a broader distribution of side-chain torsions and thus to an increase of the side-chain B-factor. As the capped peptide samples helix as well as extended conformations in the unrestrained simulation, the observed B-factors average between both states. A similar trend is observed for valine. Our hypothesis of surface accessibility causing mobilization is consistent with this finding as intramolecular hydrogen bonding is chemically impossible for valine.

Flexibility in Molecular Dynamics Simulation of BPTI.

Reanalysis of a large scale molecular dynamics simulation of BPTI exceeding 1 ms in length[17] demonstrates the applicability of our metrics to protein simulations. We analyzed global backbone movements within the protein via a global alignment on all Cα-atoms and calculation of B-factors. Local movements were captured via multiple residue-wise local alignments (58 for all residues in BPTI) and calculation of B-factors for backbone and side-chain atoms separately. Color-coding of flexibility hot spots shows very different flexibility patterns of local and global movements in the analyzed BPTI trajectory (see Figure 2). Side-chain flexibilities measured by normalized B-factors of all eight aromatic amino acids in BPTI are smaller than 0.5, showing a restriction of side-chain movements in comparison to the isolated amino acid. Side-chain flexibility data capture the alternating flexibility patterns of exposed and buried residues within the small β-sheet in BPTI. Exposed residues adopt more diverse conformations in solution compared to buried and rigidified internal side-chains.

The Spearman rank correlation r of global Cα B-factors and local backbone movements is 0.74 indicating a strong interdependence. Correlation between Cα and side-chain flexibility is lower with 0.68 but still larger than the correlation

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

between local backbone and side-chain flexibility with $r = 0.48$. If B-factors of side-chains are not normalized to flexibility of isolated amino acids, correlation with backbone movements is reduced to almost random $r = 0.02$.

A statistical analysis highlights positive correlations between solvent-accessible surface area (SASA) and all three flexibility measures (see Figure 3 and Supporting Table 1 for absolute values). Spearman rank correlation between SASA and side-chain flexibility measured by normalized B-factors is highest with 0.43 indicating pronounced side-chain movements in exposed loop regions. Correlation is reduced for global Cα B-factors ($r = 0.33$) as well as for local backbone B-factors ($r = 0.20$). Side-chain B-factors without normalization to the isolated amino acids correlate to SASA with $r = 0.72$ showing the inherent correlation of residue size to both SASA and side-chain flexibility.

The long simulation time exceeding 1 ms allows investigating the time dependence of the different flexibility metrics. In order to analyze time scales of different metrics, we compare global and local B-factors computed from different trajectory lengths. We start out with the first 10 ns and increased the trajectory length over several orders of magnitude up to the full trajectory of 1 ms (see Figure 4). A trajectory length of 10 ns does not provide the full picture of flexibility characteristics available from the complete 1 ms trajectory. In particular, major local and global backbone movements of residues 10−20 and residues 35−45 as well as the C-terminal region are not captured. Positional fluctuations of the C-terminus are sufficiently sampled starting from 100 ns, whereas statistical sampling for flexibility in the region of residue 35−45 requires at least 100 μs. Similarly, global and local backbone flexibility is significantly altered over time scales from ns to μs for residues 10−20.

In contrast, side-chain flexibility shows faster convergence in BPTI. 100 ns simulation time appear to capture most side-chain movements. Only single residues show pronounced changes in side-chain flexibility values at larger time scales. Most prominently, they include residues involved in larger backbone rearrangements propagating to side-chain reorientations (e.g., residues 10−15).

In an effort to decouple side-chain flexibility from the sequence of major backbone rearrangements in the BPTI trajectory, we extracted snapshots according to the kinetic clustering described by Shaw et al.[17] We found residues to show distinct behavior: Most amino acids are not affected by backbone movements and thus show constant side-chain flexibility. Others react to changes in backbone conformation with altered side-chain movements as reflected by the ratio of maximum and minimum side-chain flexibility within five clusters (see Figure 5). The latter group includes several aromatic residues (e.g., Tyr-10) as well as cysteine residues (e.g., Cys-14) involved in disulfide isomerization.[17] A similar clustering approach could be followed to cluster trajectories according to distinct side-chain states and thereby allow for investigating potential changes of side-chain flexibilities of surrounding amino acids.

Flexibility in Molecular Dynamics Simulation of Serine Proteases. To examine the influence of small molecule binding, we compared side-chain flexibility of apo as well as ligand bound serine proteases. We found major changes in side-chain movements depending on the systems' complexation states in our simulations. Ligand binding in general rigidifies side-chains within the binding sites of the three serine proteases. Especially, degrees of freedom of amino acids involved in protein−ligand hydrogen bonding (especially in the S1 pocket) as well as those buried upon ligand binding are frozen (see Figure 6). Although local backbone flexibility is virtually unaffected, side-chain flexibility is frozen by 80 to 90% for the Asp residue involved in salt bridge formation in the deep S1 pocket. Similarly, degrees of freedom in side-chains of conserved Cys residues in the S1 pocket are considerably frozen upon ligand binding in all three examples (85 to 95%).

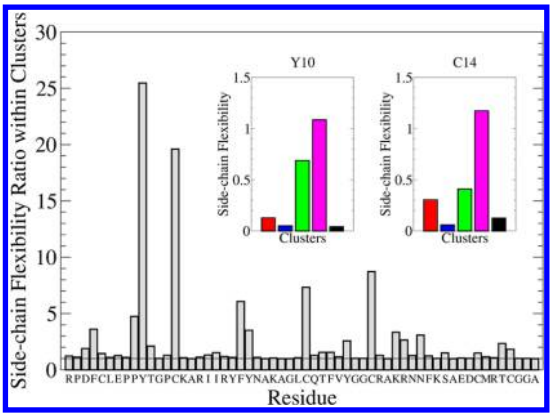


Figure 5. Side-chain flexibility in different conformational states of BPTI: The ratio between maximum and minimum side-chain flexibilities among five conformational states (according to the kinetic clustering scheme of Shaw et al.[17]) is shown for all residues in BPTI. Thereby, a value of 1 depicts invariant side-chain flexibility, whereas peaks indicate major changes in side-chain mobility depending on the protein conformation. Side-chain flexibilities for major peaks Y10 and C14 are split to respective clusters following the color scheme of Shaw et al.[17] C14 and Y10 show a simultaneous gain in mobility in one particular backbone conformation.

DISCUSSION

Our study introduces novel metrics to quantify global and local backbone and side-chain movements within protein simulations independently. We have shown that side-chain movements of isolated amino acids are independent of backbone conformation. Restriction of the conformational space of the protein backbone to α-helix or extended (β-sheet) conformation does not alter side-chain B-factors as individual degree of freedom for free amino acids. We thus conclude that the flexibility of isolated amino acid side-chains is independent from the local backbone conformation. For proteins, we observe only modest correlations for backbone and side-chain flexibility, proving the independence of these degrees of freedom and therefore the need for metrics that quantitatively capture side-chain flexibility. State-of-the-art flexibility analyses mainly focus on global movements in proteins, discarding critical local dynamic components. Additionally, local metrics facilitate decoupling of movements, thereby opening new ways for identification of hinge regions.

We introduce a metric for side-chain plasticity that calculates B-factors for the residual flexibility of the side-chain by individually aligning all snapshots of each amino acid to the corresponding backbone. As expected, we observe strong correlation of calculated B-factors with side-chain size and connectivity. For example aspartate and asparagine show comparable side-chain B-factors that increase for the chemically homologous glutamate and glutamine. In general, residues of similar size show similar B-factors demonstrating the intuitive

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

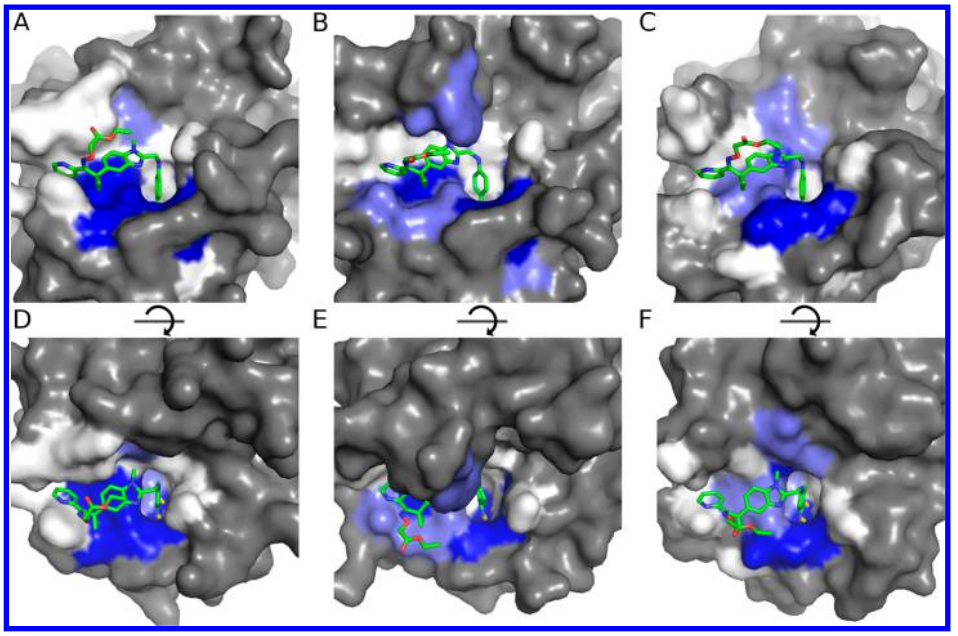


Figure 6. Freezing of side-chain movements upon ligand binding: The serine proteases factor Xa (A, D), thrombin (B, E), and trypsin (C, F) are shown in surface representation. Residues in a radius of 5 Å around the ligand are colored white if freezing is <10% of initial flexibility, light blue if between 10% and 50%, blue if flexibility is reduced by more than 50%. Regions buried by the ligand are considerably frozen, especially in the S1 and S4 region of the proteases. Lower pictures show identical systems in a top view to visualize freezing on the bottom of the deep S1 pocket, where salt bridges between ligand and protein freeze side-chain mobility of an aspartate residue (yellow dashed lines).

behavior of our approach. Proline disobeys this trend, since the cyclic linkage of the side-chain to the backbone restricts the conformational space of this side-chain.

We remove the influence of residue size via normalization to the flexibility of isolated amino acids. This results in a size-independent metric for side-chain flexibility for protein simulations derived from B-factors. We derived scaling factors based on simulations in the AMBER force field 99SB-ILDN.[47] Scaling factors for other potentials might differ but can easily be derived by unrestrained simulations of isolated amino acids with the respective potentials as outlined in the Methods section. Amino acids with small scaling factors such as glycine, proline, or alanine result in normalized side-chain flexibilities that are difficult to rationalize. For the same reason, side-chains of these residues are not included in torsional analysis via T-analyst.[38] Still, our approach has two major advantages: normalized side-chain flexibilities are comparable between residues and simulated systems and provide an intuitive scaling relative to the flexibility of isolated amino acids.

Exemplary reanalysis of a 1 ms molecular dynamics trajectory of BPTI provides global and local measures of backbone and side-chain flexibilities within the protein. While loop rearrangements dominate flexibility on the global scale, side-chain flexibility is very localized. Our metric allows for tracing side-chain flexibility within the short β-sheet of BPTI, thereby pinpointing alternating rigid and flexible side-chains which correlate with solvent exposure of the corresponding residues.

A constant relative side-chain flexibility of 1.0 for glycine (see Figure 3) proves the successful decoupling of backbone from side-chain movements showing the validity of our side-chain exit vector definition via alignment to three neighboring backbone atoms. Similarly, we observe side-chain flexibilities for alanine residues around 1.0. Only the C-terminal Ala-58 shows differences in flexibility compared to the isolated amino acid: Side-chain flexibility is increased to 3.1 due to expected elevated flexibility around the fluctuating exposed protein terminus (see also Figure 4C). In regions with extreme backbone flexibility, local backbone alignment becomes fuzzy, and backbone and side-chain flexibility appear to be coupled.

Mobility of proline residues in BPTI form two groups: Three prolines show side-chain flexibilities comparable to the isolated state with side-chain flexibilities between 0.89 and 0.93. The fourth proline with a side-chain flexibility of 0.37 forms part of a rigid hydrophobic core formed by Pro-9, Phe-22, and Phe-33 implying a rigid local environment lowering the mobility of all residues involved. In general, peripheral regions show higher side-chain flexibility than core regions of the protein. This behavior is also shown by the correlation analysis of solvent accessibility and side-chain flexibilities in Figure 3. Side-chain freezing upon internalization into the protein has been extensively studied, e.g. by Dill and co-workers.[65,66]

Kinetics of aromatic ring rotations in BPTI are well characterized from NMR experiments.[67,68] Our metric allows for assessing flexibility in the side-chains of these residues, thereby identifying the three slowest flipping aromatic rings (Tyr-23, Tyr-35, Phe-45) with the lowest B-factors among the eight aromatic amino acids in BPTI. Dihedral entropies over chi1 side-chain torsions yield similar but not identical results: Here, five aromatic amino acids (Phe-22, Tyr-23, Phe-33, Tyr-35, and Phe-45) show lower entropy and thus higher ordering than the other three. In general, dihedral entropies of aromatic residues are comparably low in agreement with normalized side-chain flexibilities but not absolute side-chain B-factors, which correlate to residue size. As for the simulations of the individual amino acids, we find local side-chain flexibility derived from B-factors in good agreement with alignment-free torsional entropy measures. Normalized side-chain flexibility has the benefit of producing a single value independent from residue site in contrast to multiple entropy values for side-chain torsions depending on residue type.

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

Additionally, isomerization of the disulfide bond formed by Cys-14 and Cys-38 has been observed by NMR studies.[69] Shaw et al. described a similar behavior within the simulation trajectory of BPTI.[17] Our metric for side-chain flexibility identifies the two critical cysteines Cys-14 and Cys-38 with side-chain flexibilities exceeding 0.4 in contrast to values below 0.15 for other cysteine residues. The distinct behavior of Cys-14 is also reflected by major differences in side-chain flexibilities depending on clustered backbone conformations (see also Figure 5). Similarly, side-chain mobility of Tyr-10 depends on the backbone conformation. Flexibility in one particular conformational state is comparable to the residue's mobility in the isolated state.

Further amino acids show interesting behavior in BPTI (see also Figure 5): On the one hand, Leu-6, Thr-11, Arg-17, Leu-29, and Val-34 show normalized side-chain flexibilities exceeding 1.0 in BPTI. Thus, these amino acids which all occupy solvent-exposed positions are more flexible than as isolated amino acid. We speculate that the protein environment fosters rotation of these side-chains, thereby increasing side-chain flexibilities. On the other hand, particular residues in loop regions show restricted side-chain movements due to distinct molecular interactions. Arg-42 in a loop region forms a salt bridge to Asp-3 within the N-terminal helix. Hence, the side-chain flexibility is reduced to 0.61 although the Arg-42 is readily accessible for the surrounding solvent molecules.

Convergence of flexibility metrics over simulation time agrees with the expected behavior. Sampling of global and local movements of the backbone requires the longest simulation trajectories spanning all time scales up to hundreds of microseconds, which is in agreement with NMR relaxation dispersion experiments.[70] In contrast, our metrics highlight that side-chain flexibilities converge earlier (see Figure 4). Spearman rank correlation between side-chain flexibilities calculated at different time scales longer than 100 ns consistently exceed 0.9. Thus, no major changes in side-chain flexibility are observed at time scales that exceed state-of-the-art molecular dynamics simulations except for single amino acids where major backbone movements allow new degrees of freedom in the side-chain. Global as well as local backbone B-factors require 100 μs sampling time respectively to reach the same level of convergence in BPTI. In their original publication, Shaw et al. described the dynamic effects involved in the slower convergence of backbone B-factors (e.g., aromatic ring rotation, disulfide bridge isomerization).[17] Our novel local flexibility metrics for side-chain flexibility show earlier convergence, thus requiring shorter simulation times than global backbone analyses.

A second application studied the effects of ligand binding on side-chain flexibility in the binding site of the three serine proteases factor Xa, thrombin, and trypsin. Considerable parts of the binding sites are frozen in their rotation in the presence of the ligand (see also Figure 6). Our metric allows for quantifying these changes, thus providing access to entropic factors in ligand recognition. Depending on the shape of the binding pocket we see differences in side-chain freezing: The rather open and shallow pocket of trypsin is rigidified by the ligand to a smaller extent than the occluded site of factor Xa. These factors might play a crucial role in the specificity of protease-ligand recognition.[71]

CONCLUSIONS

Our study provides novel tools for decomposing dynamic contributions in protein structural ensembles into global and local degrees of freedom of backbone and side-chains. By normalization to the flexibility of isolated amino acids we obtain dimensionless side-chain flexibilities. They are valuable in tracing conformational plasticity of side-chains independent of residue size and connectivity and, thus, comparable between residue types. By reanalysis of a long simulation trajectory of BPTI, we showed that the new flexibility metrics cover diverse motions and time scales. Side-chain dynamics in BPTI appear to converge around 100 ns, a time scale accessible with state-of-the-art MD simulations. Additionally, we showed an analysis of binding site flexibility in three serine proteases and found major changes upon ligand binding. An extension of the analysis toward nonstandard amino acids or nucleotides is straightforward and accessible using standard molecular dynamics simulation and analysis packages. Since NMR-derived dynamics data has recently even been used in fitting of novel force fields with improved reproduction of dynamics time scales,[72] we are convinced that intuitive metrics for flexibility quantification are of utmost importance.

Since side-chain plasticity is a critical factor in protein folding,[73] protein−protein interfaces,[74] protein−ligand binding,[75] and protein design,[76] novel methods for quantifying side-chain flexibility are crucial for structure-based modeling. With modern binding models including receptor flexibility such as the conformational selection paradigm,[77,78] conformational dynamics and their link to protein folding[79] and binding specificity[80] are more and more at the center of interest for understanding biological interactions and functions. Our metric allows for quantifying subtle changes in local receptor flexibility upon ligand binding or complex formation, allowing for the estimation of individual thermodynamic contributions[81] associated with the change in flexibility upon binding.

To ensure experimental feedback for the metrics introduced here, we encourage the community to investigate side-chain flexibility via NMR experiments in more detail. Since currently applied methods are mostly limited to backbone dynamics,[82] flexibility within predefined vectors,[83] or methyl rotations in side-chains,[84−86] a quantitative comparison of full side-chain flexibility is hardly possible. We further encourage the development and application of new methodologies, e.g. based on full isotope labeling techniques[87] or selective labeling,[88] to assess full protein dynamics by theoretical and experimental means. We expect our novel metric for side-chain flexibility and the included local alignment frames helpful in the interpretation of results from NMR experiments. We hope that a combination of these techniques will provide unique insights into biological processes at atomistic resolution.

ASSOCIATED CONTENT

Supporting Information

A 2D-RMSD plot of the simulation of capped lysine as well as time-resolved Ramachandran plots for free and restraint simulations are available as Supporting Figure 1. Supporting Figure 2 shows results from an elongated simulation of capped lysine (length 10 μs) indicating no changes in flexibility beyond the time scale of 1 μs. Supporting Table 1 lists residue-wise SASA and flexibility metrics calculated from the BPTI trajectory used for generation of Figure 3. This material is available free of charge via the Internet at http://pubs.acs.org.

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

AUTHOR INFORMATION

Corresponding Author

*E-mail: Klaus.Liedl@uibk.ac.at.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

DOC-fellowships of the Austrian Academy of Sciences to Julian E. Fuchs and Roland G. Huber, Austrian Science Fund FWF: project "Targeting Influenza Neuraminidase" (P23051).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Julian E. Fuchs and Roland G. Huber thank the Austrian Academy of Sciences for a DOC fellowship at the Institute of General, Inorganic and Theoretical Chemistry at University of Innsbruck. This work was supported by the Austrian Science Fund FWF via the project "Targeting Influenza Neuraminidase" (P23051) and the platform High Performance Computing at University of Innsbruck. The authors thank Ursula Kahler and Benjamin Hye for code testing and critical discussions and John Klepeis from D. E. Shaw Research for kindly providing access to the BPTI trajectory. Valuable feedback on NMR techniques from Martin Tollinger, University of Innsbruck, is gratefully acknowledged.

ABBREVIATIONS

BPTI, bovine pancreatic trypsin inhibitor; MD, molecular dynamics; MSF, mean squared fluctuation; NMR, nuclear magnetic resonance; SASA, solvent-accessible surface area

REFERENCES

(1) Falke, J. J. *Science* **2002**, *295*, 1480−1481.
(2) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789−796.
(3) Fenwick, R. B.; Esteban-Martin, S.; Salvatella, X. *Eur. Biophys. J.* **2011**, *40*, 1339−1355.
(4) Hensen, U.; Meyer, T.; Haas, J.; Rex, R.; Vriend, G.; Grubmüller, H. *PLoS One* **2012**, *7*, e33931.
(5) Mittermaier, A.; Key, L. E. *Science* **2006**, *312*, 224−228.
(6) Haustein, E.; Schwille, P. *Curr. Opin. Struct. Biol.* **2004**, *14*, 531−540.
(7) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158−170.
(8) Ringe, D.; Petsko, G. A. *Methods Enzymol.* **1986**, *131*, 389−433.
(9) Fontana, A.; Fassina, G.; Vita, C.; Dalzoppo, D.; Zamai, M.; Zambonin, M. *Biochemistry* **1986**, *25*, 1851−1857.
(10) Yang, L.-W.; Eyal, E.; Chennubhotla, C.; Jee, J. G.; Gronenborn, A. M.; Bahar, I. *Structure* **2007**, *15*, 741−749.
(11) Earnest, T.; Fauman, E.; Craik, C. S.; Stroud, R. *Proteins* **1991**, *10*, 171−187.
(12) Yuan, Z.; Zhao, J.; Wang, Z.-X. *Protein Eng.* **2003**, *16*, 109−114.
(13) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(14) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.
(15) Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2006**, *106*, 3055−3079.
(16) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. S. F.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 2997−3002.
(17) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−346.
(18) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(19) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95−99.
(20) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357−386.
(21) Schrauber, H.; Eisenhaber, F.; Argos, P. *J. Mol. Biol.* **1993**, *230*, 592−612.
(22) Zavodsky, M. I.; Kuhn, L. A. *Protein Sci.* **2005**, *14*, 1104−1114.
(23) Gaudreault, F.; Chartier, M.; Najmanovich, R. *Bioinformatics* **2012**, *28*, i423−i430.
(24) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1987**, *198*, 295−310.
(25) Dunbrack, R. L., Jr.; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543−574.
(26) Dunbrack, R. L., Jr.; Karplus, M. *J. Mol. Biol.* **1994**, *1*, 334−340.
(27) Doreleijers, J. F.; Sousa da Silva, A. W.; Krieger, E.; Nabuurs, S. B.; Spronk, C. A. E. M.; Stevens, T. J.; Vranken, W. F.; Vriend, G.; Vuister, G. W. *J. Biomol. NMR* **2012**, *54*, 267−283.
(28) Harder, T.; Boomsma, W.; Paluszewski, M.; Frellsen, J.; Johansson, K. E.; Hamelryck, T. *BMC Bioinf.* **2010**, *11*, 306.
(29) Kidera, A.; Go, N. *J. Mol. Biol.* **1992**, *225*, 457−475.
(30) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522−4534.
(31) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. O.; Xu, H.; Trbovic, N.; Friesner, R. A.; Palmer, A. G., III; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112*, 6155−6158.
(32) Case, D. A. et al. *AMBER 12*; University of California: San Francisco, 2012.
(33) Roe, D. R.; Cheatham, T. E., III *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(34) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.
(35) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(36) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(37) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719−1751.
(38) Ai, R.; Fatmi, M. Q.; Chang, C. A. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 819−827.
(39) Chang, C. A.; McLaughlin, W. A.; Baron, R.; Wang, W.; McCammon, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7456−7461.
(40) Huber, R. G.; Fuchs, J. E.; von Grafenstein, S.; Laner, M.; Wallnoefer, H. G.; Abdelkader, N.; Kroemer, R. T.; Liedl, K. R. *J. Phys. Chem. B* **2013**, *117*, 6466−6472.
(41) Huber, R. G.; Eibl, C.; Fuchs, J. E. *Protein Sci.* **2015**, *24*, 174−181.
(42) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. *Proteins* **2001**, *44*, 150−165.
(43) Fulle, S.; Gohlke, H. In *Computational Drug Discovery and Design, Methods in Molecular Biology*; Baron, R., Ed.; Wiley: 2012; Vol. *819*, pp 75−91.
(44) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *252*, 492−503.
(45) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Wallnoefer, H. G.; Liedl, K. R. *Proteins* **2014**, *82*, 546−555.
(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(47) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950−1958.
(48) Adelman, S. A.; Doll, J. D. *J. Chem. Phys.* **1976**, *64*, 2375−2388.
(49) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860

(50) Wallnoefer, H. G.; Handschuh, S.; Liedl, K. R.; Fox, T. *J. Phys. Chem. B* **2010**, *114*, 7405−7412.
(51) Le Grand, S.; Götz, A. W.; Walker, R. C. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
(52) Lippert, R. A.; Bowers, K. J.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Shaw, D. E. *J. Chem. Phys.* **2007**, *126*, 046101.
(53) Hovmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr.* **2002**, *D58*, 768−776.
(54) Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301−329.
(55) Nar, H.; Bauer, M.; Schmid, A.; Stassen, J. M.; Wienen, W.; Priepke, H. W.; Kauffmann, I. K.; Ries, U. J.; Hauel, N. H. *Structure* **2001**, *9*, 29−37.
(56) *Molecular Operating Environment (MOE), 2012.10*; Chemical Computing Group Inc.: Montreal, QC, Canada, 2012.
(57) Labute, P. *Proteins* **2009**, *75*, 187−205.
(58) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(59) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Chem. Phys.* **1993**, *97*, 10269−10280.
(60) Frisch, M. J. et al. *Gaussian 03, Revision C.02*; Gaussian Inc.: Wallingfort, CT, 2004.
(61) Bradbrook, G. M.; Gleichmann, T.; Harrop, S. J.; Habash, J.; Raftery, J.; Kalb, J.; Yariv, J.; Hillier, I. H.; Helliwell, J. R. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1603−1611.
(62) Doig, A. J. *Biophys. Chem.* **1996**, *61*, 131−141.
(63) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217−230.
(64) Ho, B. K.; Brasseur, R. *BMC Struct. Biol.* **2005**, *5*, 14.
(65) Bromberg, S.; Dill, K. A. *Protein Sci.* **1994**, *3*, 997−1009.
(66) Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561−602.
(67) Wagner, G.; DeMarco, A.; Wüthrich, K. *Biophys. Struct. Mech.* **1976**, *2*, 139−158.
(68) Wagner, G.; Brühwiler, D.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 227−231.
(69) Otting, G.; Liepinsh, E.; Wüthrich, K. *Biochemistry* **1993**, *32*, 3571−3582.
(70) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2012**, *134*, 2555−2562.
(71) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Margreiter, M. A.; Spitzer, G. M.; Wallnoefer, H. G.; Liedl, K. R. *PLoS Comput. Biol.* **2013**, *9*, e1003007.
(72) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. *Proteins* **2014**, *82*, 195−215.
(73) Zhang, J.; Liu, J. S. *PLOS Comput. Biol.* **2006**, *2*, e168.
(74) Cole, C.; Warwicker, J. *Protein Sci.* **2002**, *11*, 2860−2870.
(75) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. *Proteins* **2000**, *39*, 261−268.
(76) Fleishman, S. J.; Khare, S. D.; Koga, N.; Baker, D. *Protein Sci.* **2011**, *20*, 753−757.
(77) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **1999**, *12*, 713−720.
(78) Tsai, C. J.; Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9970−9972.
(79) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *J. Phys. Chem. B* **2013**, *117*, 12935−12942.
(80) Wallnoefer, H. G.; Lingott, T.; Gutierrez, J. M.; Merfort, I.; Liedl, K. R. *J. Am. Chem. Soc.* **2010**, *132*, 10330−10337.
(81) Pickett, S. D.; Sternberg, M. J. E. *J. Mol. Biol.* **1993**, *231*, 825−839.
(82) Palmer, A. G., III *Curr. Opin. Struct. Biol.* **1997**, *7*, 732−737.
(83) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(84) Lee, A. L.; Wand, A. J. *Nature* **2001**, *411*, 501−504.
(85) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959−8966.
(86) Hansen, D. F.; Kay, L. E. *J. Am. Chem. Soc.* **2011**, *133*, 8272−8281.
(87) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. *Nature* **2006**, *440*, 52−57.
(88) Hansen, A. L.; Lundstroem, P.; Velyvis, A.; Kay, L. E. *J. Am. Chem. Soc.* **2012**, *134*, 3178−3189.

DOI: 10.1021/ct500633u
J. Chem. Theory Comput. 2015, 11, 851−860