JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Correlating Protein Hot Spot Surface Analysis Using ProBiS with Simulated Free Energies of Protein−Protein Interfacial Residues

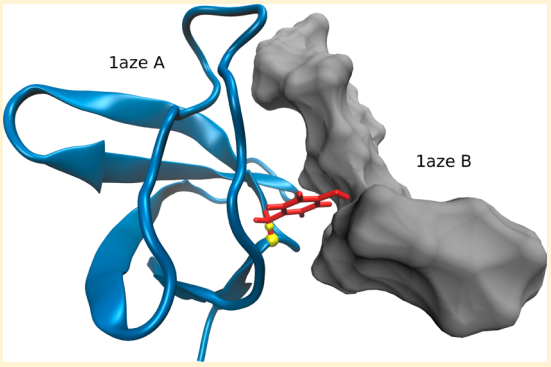
Nejc Carl,[†] Milan Hodošček,[†] Blaž Vehar,[†] Janez Konc,[†] Bernard R. Brooks,[‡] and Dušanka Janežič*[,†]

[†]National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

[‡]National Institutes of Health, National Heart, Lung and Blood Institute, Laboratory of Computational Biology, 5635 Fishers Lane, T-900 Suite, Rockville, Maryland 20852, United States

Supporting Information

ABSTRACT: A protocol was developed for the computational determination of the contribution of interfacial amino acid residues to the free energy of protein−protein binding. Thermodynamic integration, based on molecular dynamics simulation in CHARMM, was used to determine the free energy associated with single point mutations to glycine in a protein−protein interface. The hot spot amino acids found in this way were then correlated to structural similarity scores detected by the ProBiS algorithm for local structural alignment. We find that amino acids with high structural similarity scores contribute on average −3.19 kcal/mol to the free energy of protein−protein binding and are thus correlated with hot spot residues, while residues with low similarity scores contribute on average only −0.43 kcal/mol. This suggests that the local structural alignment method provides a good approximation of the contribution of a residue to the free energy of binding and is particularly useful for detection of hot spots in proteins with known structures but undetermined protein−protein complexes.



INTRODUCTION

When binding to one another, proteins exhibit a diverse array of physical and functional interactions and determine cellular functions that are fundamental to living organisms.[1,2] Typically, a few amino acid residues, the so-called hot spots, account for the majority of the binding free energy in a protein−protein binding site.[3] Protein−protein interactions are affected by point mutations, whose effect is most pronounced at hot spots.

We tested the validity of local structural alignment, conducted with the Protein Binding Sites (ProBiS) algorithm which identifies protein binding sites through local structural alignment,[4−8] for identification of hot spots. To assess the relationship between local structural similarities and the contribution of specific amino acids to the free energy of binding we employed molecular dynamics free energy calculations.[9]

Molecular dynamics free energy calculations were first applied to proteins some 25 years ago,[9,10] but the limited computational capabilities available at that time gave rise to problems related to adequate sampling of the conformational space. This field is subject to continuous methodological development,[11−23] and by 2002, advances in the field were considered sufficient to make molecular dynamics free energy calculations more widely applicable in the field of biological macromolecules.[9] Since then, they have been frequently applied to the determination of the free energy of ligand-protein binding.[24] Examples include the determination of binding free energy of the antibiotic sparsomycin with the 50S ribosomal subunit[25] and discrimination between true ligands and a set of putative binders identified by docking.[26] These methods can predict unknown binding affinities fairly accurately,[27] but they have limitations.[28] Molecular dynamics free energy calculations have been applied to enzymatic reactions, with the reactive center described by quantum mechanics and the environment by molecular mechanics[29] and were also used to assess the difference in interaction energies of a protein with two different membrane lipids in a cell membrane.[30] In addition, there are many rapid free energy methods like Linear Interaction Energy (LIE)[31,32] and Linear Response Approximation (LRA).[33,34] The insights to protein−protein binding derived from molecular dynamics free energy calculations can provide a necessary augmentation to the developments in the field of drug design.[35−37]

Molecular dynamics free energy simulations of point mutations in proteins were shown to be in quantitative agreement with relevant experimental data.[38,27,21] For example, the effect of amino acid mutations on the thermostability of proteins was successfully simulated in a study that also exposed the need for methods capable of performing molecular dynamics free energy calculations inside protein−protein interfaces.[38]

This article examines the relationship between hot spot amino acid residues identified by the ProBiS algorithm[5] and the

Received: July 11, 2012
Published: September 25, 2012

dx.doi.org/10.1021/ci3003254 | J. Chem. Inf. Model. 2012, 52, 2541−2549

interaction free energy of these amino acid residues determined by molecular dynamics via free energy methods. Local structural alignment of protein surfaces is shown to be a useful means of understanding the role in protein−protein interactions of individual amino acids and can be used for identification of hot spots in known proteins with undetermined protein−protein complex structures.

COMPUTATIONAL METHODS

Protein−Protein Interactions; Identification of Important Residues. Many methods for prediction of protein−protein binding sites[39−41] are based on amino acid sequence,[42,43] spatial position of surface atoms,[44] functional groups[5,45] or amino acid residues,[46] or energy scoring functions.[47,48] Few methods are dedicated to prediction of hot spots.[49−51]

To predict hot spot amino acids, we used the ProBiS algorithm[5,8] which detects structurally similar sites on protein surfaces by local surface structure alignment. ProBiS compares protein surfaces by examining their physicochemical properties and detects local similarities in their spatial arrangement. It is independent of amino acid sequence and compares query proteins against a database of protein structures regardless of their fold similarities. It calculates local structural similarity scores, maps them to surface amino acids of the query protein, and thus identifies regions whose structures have been conserved during evolution. Evolutionarily conserved regions are often involved in protein−protein binding.[52] Local surface structures which are dissimilar, i.e., their physicochemical chemical properties cannot be aligned with rmsd <2.0 Å, are discarded.[5] To compute local structural similarity scores of the query protein surface, the ProBiS Web server[8] searches a database of approximately 30,000 nonredundant protein structures from the PDB. [Single chain protein structures from PDB that have >95% sequence identical structures are clustered together, and a representative of each cluster is chosen.]

We hypothesize that the amino acids with high similarity scores predicted by ProBiS are correlated with hot spot residues and play a major role in protein−protein binding. These residues are not necessarily contiguous in protein sequence but are typically found in well-defined local 3D arrangements.[5,49,50] Examples of important residues identified by ProBiS are given in Figure 1, where residues have a local structural similarity score ranging from 0 (blue, dissimilar) to 9 (red, similar). In this study, we defined high-scoring residues as those with local structural similarity scores of ≥7 and low-scoring residues as those with scores ≤3. These values were selected because intermediate similarity scores (4−6) represent residues whose contribution to structural conservation or lack thereof cannot be reliably assessed.[5] If local structural similarities of protein surfaces are correlated to the free energy of binding, then high-scoring residues should contribute significantly more to the free energy of binding than low-scoring residues.

The contributions of high- and low-scoring amino acid residues to the free energy of binding were calculated by thermodynamic integration as implemented in CHARMM. If the high-scoring residues predicted by ProBiS indeed play significant roles in binding and are candidates for the role of hot spot amino acids, then examination by thermodynamic integration can be restricted to just these residues with a concomitant reduction in the necessary computational resources. A schematic representation of this approach is given in Figure 2.

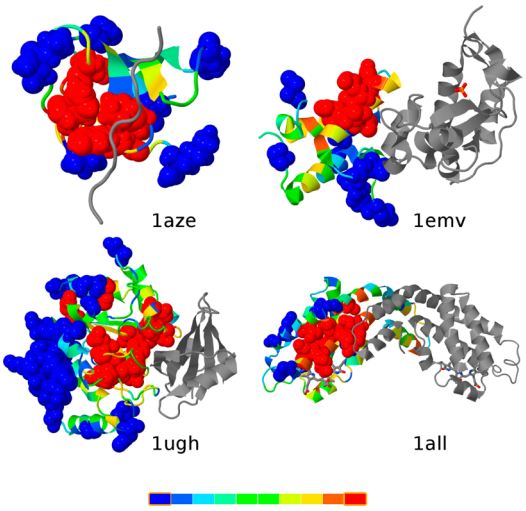


Figure 1. ProBiS results for four complexes used in this study to examine the correlation between local structural similarity score and contribution to the free energy of binding. Proteins that form complexes with the investigated proteins are shown in gray. ProBiS scores are presented in a color scheme (bottom) with the most significant residues with a local structural similarity score of 9 colored red, and the least significant with a score of 0 colored blue.

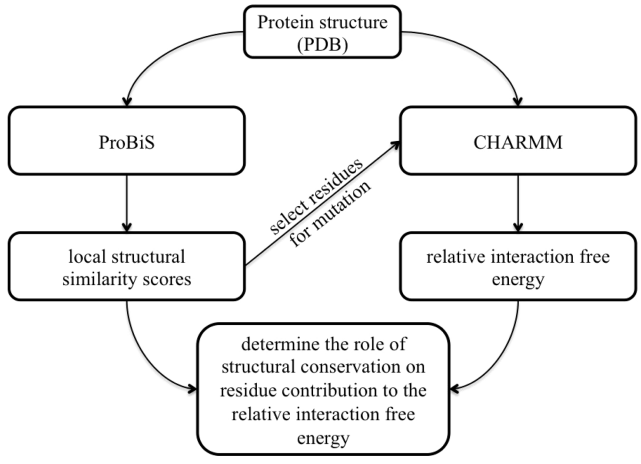


Figure 2. The process of determining the role of structural conservation on amino acid's contribution to the free energy of binding as determined by molecular dynamics free energy simulation.

Protein Simulation Systems. Protein−protein complexes 1aze, 1emv, 1ugh, and 1all from the RCSB Protein Data Bank (www.pdb.org)[53] were prepared for molecular dynamics simulations. Hydrogens were added to all protein chains by the HBUILD routine in CHARMM.[54] Each protein system was explicitly solvated in water and neutralized with Na^+ and Cl^- ions. All systems were simulated using an academic version of CHARMM[55−57] with the CHARMM27 force field.[58−60] Arginine and lysine side chains were simulated in a protonated form, while glutamic and aspartic acids were simulated in deprotonated form; histidines were left uncharged. No disulfide patches were necessary because none of the studied proteins contain a disulfide bridge.

dx.doi.org/10.1021/ci3003254 | J. Chem. Inf. Model. 2012, 52, 2541−2549

Table 1. Simulation Systems for Examined Proteins with Parameters Indicating the Computational Intensity[a]

protein PDB ID	protein chain identifier	number of protein atoms	number of sodium ions	number of chloride ions	number of atoms in simulation system	size of simulation cube [Å, (λ=0)]
1aze	**A**	**904**	**89**	**86**	**35646**	**70.1**
1aze	**A, B**	**1081**	**91**	**91**	**35654**	**70.0**
1emv	A	1269	89	81	23864	61.5
1emv	A, B	3340	93	82	59864	83.5
1ugh	**E**	**3573**	**75**	**80**	**55124**	**81.3**
1ugh	**E, I**	**4863**	**79**	**72**	**73441**	**89.4**
1all	A	2409	87	84	51744	79.6
1all	A, B	4837	88	85	95574	97.6

[a]Each simulation system is composed of the protein, water molecules, sodium and chloride ions.

Each protein complex was placed in a cube of water, so that the protein was at least 10 Å from the edge of the cube, and periodic boundary conditions were applied. Details of different simulation systems are summarized in Table 1.

The protocol for equilibration of all systems was as follows. The protein atoms were first harmonically restrained, and only water molecules were minimized. Steepest descent minimization (500 steps) was followed by 1,000 steps of adopted basis Newton–Raphson minimization. Water was subsequently heated from 0 to 300 K in 2,000 steps with the leapfrog Verlet algorithm and then equilibrated with constant pressure and temperature (NPT) dynamics for 200 ps with an integration time step of 1 fs. For consistency with the CHARMM force field,[61] water molecules were modeled explicitly with the TIP3P model.[62] The cutoff for van der Waals interactions was set to 13.5 Å.[63] The temperature was held constant at 300 K with the Nose-Hoover thermostat.[64] Pressure was held constant with the internal barostat. Electrostatics were calculated with the particle mesh Ewald method. To ensure that the system was adequately relaxed, the harmonic restraint on protein atoms was removed, and molecular dynamics were simulated for further 1.5 ns, with an integration time step of 1 fs. To ensure an adequate ionic strength, 100 Na^+ and 100 Cl^- ions were added and distributed randomly inside the simulation box. Ions that overlapped with the protein were deleted; water molecules that overlapped with ions were also deleted. Next, an appropriate number of either cations or anions was deleted to leave a charge-neutral solution with salt content in the 200–400 mM range. The solution was allowed to relax for another 100 ps with an integration time step of 1 fs, and the final coordinates were used to define the initial system in calculations of free energies of point mutations.

To prepare structures of free proteins not bound to their binding partners, the coordinates of the binding partner from the complex were deleted, and the protein was solvated and relaxed during a molecular dynamics run as described above for protein complexes. The protein structures prepared in this way were used in the following point mutation protocol.

Molecular Dynamics Free Energy Calculations. Molecular dynamics free energy simulations were used to determine the contributions of high- and low-scoring residues to the binding free energy of investigated proteins. For each examined residue, the free energy change associated with its mutation to glycine was calculated for the protein in its complexed and unbound forms.

The basic principle of our binding analysis is illustrated in Figure 3. The binding free energy change caused by a single point mutation in the protein–protein interface can be calculated from the processes represented by the vertical legs

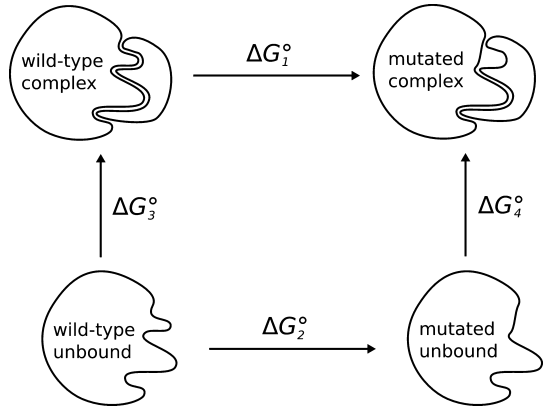


Figure 3. The thermodynamic cycle describing the binding of the wild-type and the mutated protein to its binding partner. The binding free energy change $\Delta\Delta G^\circ$ can be calculated as either $\Delta G^\circ_2 - \Delta G^\circ_1$ or as $\Delta G^\circ_3 - \Delta G^\circ_4$.

in Figure 3 as $\Delta\Delta G^\circ = \Delta G^\circ_3 - \Delta G^\circ_4$. Since free energy is a property of state, it can also be calculated from nonphysical processes represented by the horizontal legs in Figure 3 as $\Delta\Delta G^\circ = \Delta G^\circ_2 - \Delta G^\circ_1$, which can be more readily obtained by thermodynamic integration.

We calculated $\Delta\Delta G^\circ$ for the horizontal legs of the thermodynamic scheme shown in Figure 3, where ΔG°_2 is the difference in free energy between the unbound wild-type protein and its mutated counterpart, ΔG°_1 is the difference in free energy between the complexed wild-type protein and its mutated counterpart, and $\Delta\Delta G^\circ$ is the difference between these ΔG° values. The calculations were conducted with CHARMM's PERT module. To avoid van der Waals end point problems,[65] a soft core potential (PSSP) was used as implemented in PERT.[66–68]

A negative value of $\Delta\Delta G^\circ$ means that the mutation of an amino acid into glycine is more favorable in the free protein than in the protein complex. The implication is that amino acid interactions are more favorable in the wild-type protein complex compared to the mutated one. A negative $\Delta\Delta G^\circ$ therefore signifies a favorable contribution of the amino acid to the interaction free energy in the wild-type protein.

To calculate the free energy difference associated with a point mutation we employed thermodynamic integration in which gradual transition from one state into the other is achieved by slowly changing the coupling parameter λ. Thus the potential energy function of the system is gradually changed from the starting state U_A to the end state U_B as in the following equation[69]

$$U(r^N; \lambda) = (1 - \lambda)U_A(r^N; \lambda) + \lambda U_B(r^N; \lambda) \quad (1)$$

where U is the current potential energy function at coordinates r of N atoms.

Free energy difference between $\lambda = 0$ and $\lambda = 1$ is

$$\Delta G = \int_0^1 \frac{\partial G(\lambda)}{\partial \lambda} d\lambda \quad (2)$$

which gives us the equation for thermodynamic integration

$$\Delta G^{TI} = \int_0^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3)$$

The point mutation is presented in a single topology, and so all the atoms from states U_A and U_B are present at all values of λ. When $\lambda = 0$, the investigated protein is in the wild-type state, and when $\lambda = 1$, the target residue has been transformed to glycine by patching a hydrogen atom at the α carbon of the amino acid that is to be mutated. The introduction of the glycine hydrogen atom, coinciding with the disappearance of the side chain of the original amino acid, is implemented by scaling the charge and the van der Waals radius of the glycine hydrogen from 0 to 1, while scaling the same parameters of the original amino acid side chain from 1 to 0. By means of this process, the glycine hydrogen at $\lambda = 0$ and the side chain of the original amino acid at $\lambda = 1$ effectively become dummy atoms.

With each change in λ a perturbation is introduced into the system, to which the system is allowed to adapt with molecular dynamics during which energy values are not accumulated, followed by an accumulation run in which energies that will contribute to appropriate averages are gathered. The total run time is a sum of relaxation dynamics and accumulation dynamics. To minimize the time span between the start of our simulations and the gathering of the results, we simulated each λ point on a separate computer. To provide a starting configuration for each λ point we first simulated the transition from $\lambda = 0$ to $\lambda = 1$ in 1 ns. In this simulation, the system visits all the values of λ that will be later simulated at length. The last configuration that the system visits for a particular value of λ is then used as a starting configuration for lengthier simulations at that λ point, and consequently each λ point is simulated on a separate computer. To ensure sufficient sampling we varied the length of simulations from 3 to 23 ns. We tested several protocols for variation of the parameter λ and selected a protocol that slows down when λ approaches 1 since at the end point the side chain of the original amino acid is switched off and its volume in the simulation cube becomes available to water molecules.

The Molecular Dynamics Free Energy Calculation Protocol. For accurate predictions of free energy differences the calculations must proceed in a slow, reversible manner. Consequently, energy fluctuations at individual λ points should not exceed thermal noise RT, which is approximately 0.6 kcal/mol at 300 K. Point mutation introduces a perturbation into the system that would lead to energy fluctuations far in excess of RT. When λ is gradually changed from 0 to 1, the system goes through a series of nonphysical states, but since free energy is a function of state, only the start and end configurations need be chemically or biologically relevant.

Adoption of a slow and reversible manner requires different dynamics schedules for different amino acids. A schedule was selected which ensures consistent results, i.e. similar results of simulations started with the same set of starting coordinates, the same velocity distribution, but different initial values for velocities of individual atoms. To check that energy fluctuations between neighboring λ points are below RT, perturbations of different amino acids into glycine were carried out with tripeptides AXA, where X represents the amino acid that will be mutated. The terminal alanines were both capped, and the structure of AXA is shown in Figure 4.

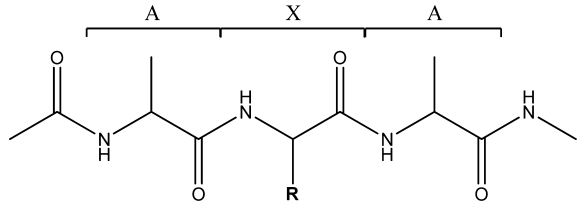


Figure 4. Tripeptide in the form of AXA, where the central amino acid is surrounded by capped alanines. R denotes the side chain of the central amino acid X.

In Figure 5, we show how energy fluctuations vary in size with various amino acid residues in a dynamics schedule.

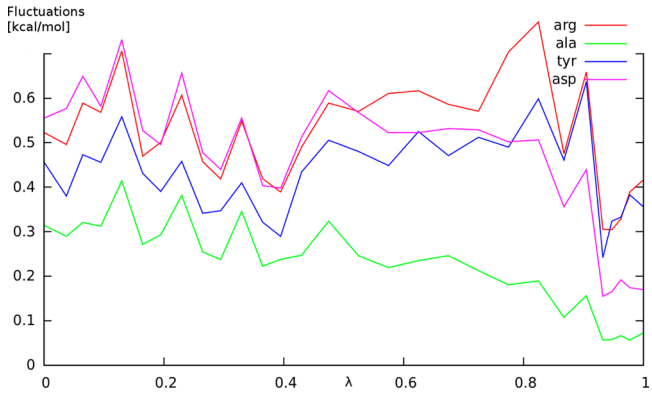


Figure 5. Energy fluctuations for different amino acids in AXA tripeptide in a dynamics schedule designed to keep fluctuations close to 0.6 kcal/mol, the value of thermal noise RT at 300 K. The λ values that the system visits are not spaced equidistantly but are sampled more frequently where perturbations are largest (in the case of the glycine point mutation, when λ approaches 1). Bulkier and charged amino acids perturb the system more than amino acids with small and neutral side chains. For simplicity, we present data for only four amino acid residues.

Perturbations to the system are always largest at the two end points, where the relative perturbation of the system is the largest (i.e., at the start of the simulation a new atom is introduced into the system, and at the end of the simulation the original side chain 'disappears'). The dynamics schedule was constructed so that energy fluctuations were close to thermal noise, which was achieved by adjusting the step size between two consecutive lambda points.

When the introduced amino acid chain is a hydrogen atom, the perturbation at the end point is greater than at the starting point because the change in the volume available to water molecules does not change significantly upon the introduction of a hydrogen atom, whereas upon the removal of the original amino acid side chain a large gap appears that can be filled with water molecules. In anticipation of this development the simulations were performed with additional λ values at the end of the simulation. The rationale for this was that the smaller the

dx.doi.org/10.1021/ci3003254 | J. Chem. Inf. Model. 2012, 52, 2541−2549

Table 2. $\Delta\Delta G^\circ$ for Tyr52 and Asn42 of Protein 1aze for Different Simulation Time Periods[a]

PDB ID	amino acid	ps/λ	parallel no.	ΔG° [kcal/mol]	no. of λ points	average fluctuation [kcal/mol]	$\Delta\Delta G^\circ$[kcal/mol]
1aze	**Asn42**	**1000**	**1**	**65.00**	**23**	**0.49**	**−0.11**
1azeA				**64.89**		**0.50**	
1aze	Asn42	500	1	64.75	23	0.47	0.38
1azeA				65.13		0.51	
1aze	**Asn42**	**250**	**1**	**64.88**	**23**	**0.47**	**−0.26**
1azeA				**64.62**		**0.52**	
1aze	Asn42	125	1	65.32	23	0.46	−0.64
1azeA				64.68		0.48	
1aze	**Asn42**	**125**	**2**	**65.44**	**23**	**0.45**	**−0.94**
1azeA				**64.50**		**0.49**	
1aze	Asn42	125	3	65.67	23	0.45	−0.80
1azeA				64.87		0.49	
1aze	**Tyr52**	**1000**	**1**	**5.66**	**28**	**0.42**	**−5.25**
1azeA				**0.41**		**0.45**	
1aze	Tyr52	500	1	5.44	28	0.42	−4.98
1azeA				0.46		0.45	
1aze	**Tyr52**	**250**	**1**	**6.26**	**28**	**0.40**	**−5.64**
1azeA				**0.62**		**0.45**	
1aze	Tyr52	125	1	7.37	28	0.40	−5.93
1azeA				1.44		0.43	
1aze	**Tyr52**	**125**	**2**	**7.32**	**28**	**0.39**	**−5.73**
1azeA				**1.59**		**0.43**	
1aze	Tyr52	125	3	7.55	28	0.40	−5.67
1azeA				1.88		0.43	

[a]The third column presents picoseconds of accumulation dynamics for each λ point. The 125 ps calculations were done in three parallels where the parallels differ only in initial velocities of the atoms, with overall velocity distribution remaining the same for all parallels. Free energy calculations for the protein complex and the free protein that were part of the same thermodynamic cycle are shown in consecutive rows in boldface.

difference between consecutive λ values, the smaller the perturbation that each succeeding state induces. As expected, the smallest perturbations were observed in amino acids that differ little from glycine, such as alanine, and the largest perturbations occurred in large residues such as arginine.

RESULTS AND DISCUSSION

ProBiS Analysis of Selected Protein Complexes. Protein complexes 1aze, 1emv, 1ugh, and 1all were analyzed with the ProBiS Web server. High- and low-scoring residues, with local structural similarity scores ≥ 7 and ≤ 3, respectively, were separately recorded for chains 1azeA, 1emvA, 1ughE, and 1allA.

Several residues were selected for molecular dynamics free energy calculations: high-scoring Asn51 and Tyr52 from 1azeA; Gln144 and Ser169 from 1ughE; Asp13, Val38, Thr45, Asp89, Arg93, Tyr97, and Val100 from 1allA; low-scoring Val7 from 1azeA; Cys23, Asn24, Thr27, Asp51, and Asp62 from 1emvA; Lys28 and Val31 from 1allA. Each of the selected amino acids is part of a protein−protein interface and is projected from the query protein surface toward its binding partner.

Molecular Dynamics Free Energy Calculation Test Case. To observe the change in free energy and establish a simulation protocol associated with a point mutation we investigated the complex between N-SH3 domain of Grb2 protein and a peptide derived from Sos (1aze)[70] a small (20 kDa) protein−protein complex. This complex was chosen due to its small size which enabled us to perform many test simulations. All mutations in the test case were done on chain A of the complex (1azeA).

We investigated the free energy change in relation to the simulation duration and the assignment of initial atom velocities. Longer simulations yield more accurate results but at a higher computational cost. For example, 125 ps of simulation time for chains A and B of protein 1all (see Table 1 for details) required 56 h on a workstation with two Intel Xeon 2.27 GHz processors running Gentoo Linux. A smaller system (chains A and B of protein 1aze) required 18 h for 125 ps of simulation time and scaled linearly with increasing length, with 1 ns of simulation time requiring 150 h of computation time on the same computer system. The velocity distribution (temperature) in individual simulations was constant; the initial velocities assigned to individual atoms however were not. Since initial velocities should not influence the results, multiple parallels of the same calculation were performed with different initial velocities. The resulting free energy differences were checked for conformity.

In Table 2, the data for point mutations of various amino acids into glycine in both the free protein (1azeA) and the complex (1aze) are presented. The focus was on two amino acids of chain A, Tyr52 and Asn42, which have different roles in the formation of the protein−protein complex, as seen from the NMR structures in the PDB entry 1aze;[70] Tyr52 is a part of the protein−protein binding site, whereas Asn42 is not. Tyr52 forms a strong hydrogen bond with the binding partner, and consequently it is expected that the $\Delta\Delta G^\circ$ associated with this point mutation should be less than −3 kcal/mol. The position of Tyr52 relative to the binding partner is depicted in Figure 6. During the simulation, Tyr52 is mutated into Gly52. The side chain of Gly52 (a single hydrogen atom) is located along the bond between the α and β carbon atoms of Tyr52.

From Table 2 it can be seen that the absolute value of $\Delta\Delta G^\circ$ for Asn42 is close to −0.6 kcal/mol (RT), whereas $\Delta\Delta G^\circ$ for Tyr52 is approximately −5 kcal/mol, equivalent to the energy

dx.doi.org/10.1021/ci3003254 | J. Chem. Inf. Model. 2012, 52, 2541−2549

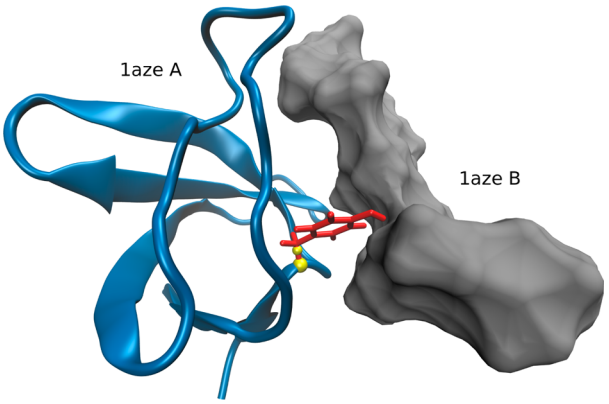


Figure 6. Protein complex 1azeA in blue and its binding partner 1azeB in gray. The side chain of the hot spot amino acid Tyr52 is shown as red sticks. Point mutation of Tyr52 results in glycine, whose hydrogen and α-carbon atoms are shown as yellow spheres.

of a strong hydrogen bond. An appropriate simulation time appears to be between 20 and 30 ns depending on the amino acid, where the simulation time is calculated as the number of λ points multiplied by time/λ point, which is also consistent with published data.[38] However, we suggest that the size of $\Delta\Delta G°$ could be approximately determined using much shorter simulations, about 3 ns long, which can also be seen in Table 2. Greater precision is achieved with longer simulations, but shorter simulations still give information about the $\Delta\Delta G°$ trend.

Comparison of Local Structural Similarity Score and the Interaction Free Energy of Amino Acids. We sought correlation between local structural similarity scores[71,72] and calculated interaction free energies. Accordingly, we calculated $\Delta\Delta G°$ values for the amino acids belonging to a number of protein−protein binding sites in different protein−protein

Table 3. Data for Various Point Mutations with Local Structural Similarity Scores Calculated with ProBiS[a]

PDB ID	amino acid	$\Delta G°$ [kcal/mol]	local structural similarity score (0−9)	$\Delta\Delta G°$ [kcal/mol]	HotPoint prediction
1aze	**Val7**	**−7.61**	**0**	**−0.84**	**nonhot**
1azeA		**−8.45**			
1aze	Asn51	72.21	7	−0.79	nonhot
1azeA		71.42			
1aze	**Tyr52**	**7.41**	**9**	**−5.77**	**nonhot**
1azaA		**1.64**			
1emv	Cys23	−2.91	1	−1.93	hot
1emvA		−4.83			
1emv	**Asn24**	**70.81**	**0**	**−0.25**	**nonhot**
1emvA		**70.56**			
1emv	Thr27	12.52	1	−2.98	nonhot
1emvA		9.57			
1emv	**Asp51**	**121.58**	**0**	**3.15**	**nonhot**
1emvA		**124.73**			
1emv	Asp62	122.41	0	1.72	nonhot
1emvA		124.14			
1ugh	**Gln144**	**47.35**	**7**	**−0.20**	**hot**
1ughE		**47.15**			
1ugh	Ser169	−12.57	9	−2.03	hot
1ughE		−14.60			
1all	**Asp13**	**129.92**	**8**	**−3.95**	**hot**
1allA		**125.97**			
1all	Lys28	36.35	2	−0.42	nonhot
1allA		35.93			
1all	**Val31**	**−6.91**	**2**	**−1.86**	**nonhot**
1allA		**−8.77**			
1all	Val38	−9.17	8	−1.51	hot
1allA		−10.67			
1all	**Thr45**	**11.46**	**8**	**−0.78**	**nonhot**
1allA		**10.68**			
1all	Asp89	129.45	8	−1.72	nonhot
1allA		127.73			
1all	**Arg93**	**263.64**	**9**	**−7.81**	**nonhot**
1allA		**255.83**			
1all	Tyr97	11.82	9	−7.15	hot
1allA		4.68			
1all	**Val100**	**−3.21**	**8**	**−3.33**	**hot**
1allA		**−6.54**			

[a]The interaction free energy was calculated with molecular dynamics free energy simulations. Data for each point mutation is shown in consecutive rows in boldface where the first row carries the information about the protein complex and the second row carries the information about the protein in free form. HotPoint predictions of hot spots are shown in the last column.

complexes, high- or low-scoring according to their ProBiS structural similarity scores.

In Table 3 are presented, side by side, the interaction free energies calculated by molecular dynamics free energy simulations and the structural similarity scores calculated by ProBiS for the investigated amino acids. The high-scoring residues contribute on average −3.19 kcal/mol to the interaction free energy, while low-scoring residues contribute on average −0.43 kcal/mol.

For comparison, we also calculated hot spots using HotPoint,[73] a method for prediction of computational hot spots based on conservation, solvent accessibility, and statistical pairwise residue potentials of the interface residues. The results are presented in Table 3. Residues predicted as hot spots using this method contribute on average −2.87 kcal/mol, whereas predicted nonhot spot residues contribute −1.53 kcal/mol. In addition, residues predicted as hot spots have average local structural similarity scores of 7.14, whereas those predicted as nonhot spots have the average score of 3.83. We observe some differences between ProBiS, molecular-dynamics, and HotPoint predictions. In 1aze, ProBiS and molecular dynamics free energy calculations predict Tyr52 as a hot spot, whereas HotPoint does not. The reason for this might be that Tyr52 is very solvent accessible, which decreases the HotPoint score that relies on solvent accessibility. In contrast, molecular dynamics shows that Tyr52 forms a hydrogen bond with its binding partner, which is not seen in the 1aze crystal structure, and thus predicts this residue as a hot spot.

Even though the results in Table 3 represent a relatively small sample, we observe a correlation between evolutionary conserved residues (high structural similarity scores) and free energies of amino acid residues involved in protein−protein binding. In contrast, there are residues such as Gln144 from 1ughE which have high structural similarity scores but low interaction free energies. This suggests that although the ProBiS algorithm can identify hot spots in most cases, the role of a residue in protein−protein interactions cannot always be explained by the similarity scores alone, or, alternatively, the residues predicted by ProBiS and with poor interaction energies could be structurally conserved due to their importance for other reasons.

CONCLUSION

In this article we present a protocol for determining a contribution of an amino acid to the free energy of formation of a protein−protein complex. To achieve this, we employed thermodynamic integration as implemented in CHARMM's PERT module. To keep the simulation systems close to equilibrium during perturbations, we adjusted the dynamics schedule to keep energy fluctuations close to thermal noise RT. We established an appropriate sampling simulation time at 20−30 ns, with shorter simulations (around 3 ns) sufficing for an estimation of the $\Delta\Delta G^{\circ}$ value.

Furthermore, we tested the accuracy of the ProBiS algorithm in its identification of binding sites by juxtaposing its scores with the interaction free energies calculated by thermodynamic integration. Our results indicate that high-scoring amino acid residues contribute more to the free energy of binding. This suggests that the local structural similarity score accurately identifies residues involved in protein−protein binding. For rapid estimation of the contribution of individual amino acids to the free energy of protein−protein complex formation, local structural alignment methods can be used to provide a short list of residues that are likely to be critical for protein−protein binding. Since molecular dynamics free energy methods require structures of complexed proteins for an estimation of amino acid contribution to the binding free energy, local structural alignment methods can be particularly useful when the structures of complexes with interesting binding partners remain undetermined.

ASSOCIATED CONTENT

Supporting Information

rmsd and energy time series for protein system used in simulations; data from Table 2 depicted in graph form; data from Table 3 depicted in graph form. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dusa@cmm.ki.si.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors express their thanks to Stefan Boresch for helpful discussions that arose during the course of this work and Urban Bren for helpful discussions and manuscript review. The financial support through grant (P1-0002) of the Ministry of Higher Education, Science, and Technology of Slovenia and the Slovenian Research Agency is acknowledged.

REFERENCES

(1) Pawson, T.; Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **2003**, *300*, 445−452.

(2) Papin, J. A.; Hunter, T.; Palsson, B. O.; Subramaniam, S. Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 99−111.

(3) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1−9.

(4) Konc, J.; Janežič, D. ProBiS: a web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.* **2010**, *38*, W436−W440.

(5) Konc, J.; Janežič, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160−1168.

(6) Carl, N.; Konc, J.; Janežič, D. Protein surface conservation in binding sites. *J. Chem. Inf. Model.* **2008**, *48*, 1279−1286.

(7) Carl, N.; Konc, J.; Vehar, B.; Janežič, D. Protein-protein binding site prediction by local structural alignment. *J. Chem. Inf. Model.* **2010**, *50*, 1906−1913.

(8) Konc, J.; Janežič, D. ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. *Nucleic Acids Res.* **2012**, *40*, W214−W221.

(9) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430−437.

(10) Warshel, A.; Sussman, F.; King, G. Free-energy of charges in solvated proteins - microscopic calculations using a reversible charging process. *Biochemistry* **1986**, *25*, 8368−8372.

(11) *Free energy calculations*; Chipot, C., Pohorille, A., Eds.; Springer-Verlag: Berlin, Germany, 2007.

(12) Bennett, C. H. Efficient estimation of free-energy differences from Monte-Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(13) Knight, J. L.; Brooks, C. L. λ-Dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.

(14) Wu, X. W.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **2003**, *381*, 512−518.

(15) Koenig, G.; Boresch, S. Non-Boltzmann sampling and Bennett's acceptance ratio method: How to profit from bending the rules. *J. Comput. Chem.* **2011**, *32*, 1082−1090.
(16) Koenig, G.; Bruckner, S.; Boresch, S. Unorthodox uses of Bennett's acceptance ratio method. *J. Comput. Chem.* **2009**, *30*, 1712−1718.
(17) Bren, M.; Florian, J.; Mavri, J.; Bren, U. Do all pieces make a whole? Thiele cumulants and the free energy decomposition. *Theor. Chem. Acc.* **2007**, *117*, 535−540.
(18) Bren, U.; Martinek, V.; Florian, J. Decomposition of the solvation free energies of deoxyribonucleoside triphosphates using the free energy perturbation method. *J. Phys. Chem. B* **2006**, *110*, 12782−12788.
(19) Bren, U.; Martinek, V.; Florian, J. Free energy simulations of uncatalyzed DNA replication fidelity: Structure and stability of T-G and dTTP-G terminal DNA mismatches flanked by a single dangling nucleotide. *J. Phys. Chem. B* **2006**, *110*, 10557−10566.
(20) Christ, C. D.; van Gunsteren, W. F. Enveloping distribution sampling: a method to calculate free energy differences from a single simulation. *J. Chem. Phys.* **2007**, *126*, 184110.
(21) Massova, I.; Kollman, P. A. Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(22) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Protein-protein recognition: A computational mutagenesis study of the Mdm2-P53 complex. *Theor. Chem. Acc.* **2008**, *120*, 533−542.
(23) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot spots—A review of the protein−protein interface determinant amino-acid residues. *Proteins* **2007**, *68*, 803−812.
(24) Oostenbrink, C.; van Gunsteren, W. F. Free energies of ligand binding for structurally diverse compounds. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6750−6754.
(25) Ge, X.; Roux, B. Calculation of the standard binding free energy of sparsomycin to the ribosomal peptidyl-transferase P-site using molecular dynamics simulations with restraining potentials. *J. Mol. Recognit.* **2010**, *23*, 128−141.
(26) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118−1134.
(27) Koenig, G.; Brooks, B. R. Predicting binding affinities of host-guest systems in the SAMPL3 blind challenge: the performance of relative free energy calculations. *J. Comput. Aided Mol. Des.* **2012**, *26*, 543−550.
(28) Merz, K. M. Limits of free energy computation for protein-ligand interactions. *J. Chem. Theory Comput.* **2010**, *6*, 1769−1776.
(29) Kastner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: Application to an enzymatic reaction. *J. Chem. Theory Comput.* **2006**, *2*, 452−461.
(30) Rodriguez, Y.; Mezei, M.; Osman, R. The PT1-Ca^{2+} Gla domain binds to a membrane through two dipalmitoylphosphatidylserines. A computational study. *Biochemistry* **2008**, *47*, 13267−13278.
(31) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.
(32) Bren, U.; Oostenbrink, C. Cytochrome P450 3A4 inhibition by ketoconazole: Tackling the problem of ligand cooperativity using molecular dynamics simulations and free-energy calculations. *J. Chem. Inf. Model.* **2012**, *52*, 1573−1582.
(33) Lee, F. S.; Chu, Z.-T.; Bolger, M. B.; Warshel, A. Calculations of antibody-antigen interactions: Microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng.* **1992**, *5*, 215−228.
(34) Bren, U.; Lah, J.; Bren, M.; Martinek, V.; Florian, J. DNA duplex stability: The role of preorganized electrostatics. *J. Phys. Chem. B* **2010**, *114*, 2876−2885.
(35) Loving, K.; Alberts, I.; Sherman, W. Computational approaches for fragment based de novo design. *Curr. Top. Med. Chem.* **2010**, *10*, 14−32.
(36) Bienstock, R. J. Computational drug design targeting protein-protein interactions. *Curr. Pharm. Des.* **2012**, *18*, 1873−4286.
(37) Rechfeld, F.; Gruber, P.; Hofmann, J.; Kirchmair, J. Modulators of protein-protein interactions: novel approaches in targeting protein kinases and other pharmaceutically relevant biomolecules. *Curr. Top. Med. Chem.* **2011**, *11*, 1305−1319.
(38) Seeliger, D.; de Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* **2010**, *98*, 2309−2316.
(39) de Vries, S. J.; Bonvin, A. M. J. J. How proteins get in touch: Interface prediction in the study of biomolecular complexes. *Curr. Protein Pept. Sci.* **2008**, *9*, 394−406.
(40) Matthiesen, R. Methods, algorithms and tools in computational proteomics: A practical point of view. *Proteomics* **2007**, *7*, 2815−2832.
(41) Tuncbag, N.; Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Briefings Bioinf.* **2009**, *10*, 217−232.
(42) Ofran, Y.; Rost, B. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett.* **2003**, *544*, 236−239.
(43) Chen, X. W.; Jeong, J. C. Sequence-based prediction of protein interaction sites with an integrative method. *Bioinformatics* **2009**, *25*, 585−591.
(44) Petsalaki, E.; Stark, A.; Garcia-Urdiales, E.; Russell, R. B. Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput. Biol.* **2009**, *5*, 1−10.
(45) Konc, J.; Trykowska Konc, J.; Penca, M.; Janežič, D. Binding-sites prediction assisting protein-protein docking. *Acta Chim. Slov.* **2011**, *58*, 396−401.
(46) Shoemaker, B. A.; Zhang, D. C.; Thangudu, R. R.; Tyagi, M.; Fong, J. H.; Marchler-Bauer, A.; Bryant, S. H.; Madej, T.; Panchenko, A. R. Inferred biomolecular interaction server-a web server to analyze and predict protein interacting partners and binding sites. *Nucleic Acids Res.* **2010**, *38*, D518−D524.
(47) Liang, S. D.; Zhang, C.; Liu, S.; Zhou, Y. Q. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res.* **2006**, *34*, 3698−3707.
(48) Dosztanyi, Z.; Meszaros, B.; Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745−2746.
(49) Keskin, O.; Ma, B. Y.; Nussinov, R. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **2005**, *345*, 1281−1294.
(50) Guney, E.; Tuncbag, N.; Keskin, O.; Gursoy, A. HotSprint: database of computational hot spots in protein interfaces. *Nucleic Acids Res.* **2008**, *36*, D662−D666.
(51) Landon, M. R.; Lancia, D. R., Jr.; Yu, J.; Thiel, S. P.; Vajda, S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J. Med. Chem.* **2007**, *50*, 1231−1240.
(52) Konc, J.; Janežič, D. Protein-protein binding-sites prediction by protein surface structure conservation. *J. Chem. Inf. Model.* **2007**, *47*, 940−944.
(53) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(54) Brunger, A. T.; Karplus, M. Polar hydrogen positions in proteins - empirical energy placement and neutron-diffraction comparison. *Proteins* **1988**, *4*, 148−156.
(55) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(56) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.;

York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(57) Miller, B. T.; Singh, R. P.; Klauda, J. B.; Hodoscek, M.; Brooks, B. R.; Woodcock, H. L. CHARMMing: A new, flexible web portal for CHARMM. *J. Chem. Inf. Model.* **2008**, *48*, 1920−1929.
(58) Foloppe, N.; MacKerell, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86−104.
(59) MacKerell, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105−120.
(60) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(61) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; Garcia-Moreno, E. B. High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* **2000**, *79*, 1610−1620.
(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(63) Janežič, D.; Venable, R. M.; Brooks, B. R. Harmonic-analysis of large systems. 3. Comparison with molecular-dynamics. *J. Comput. Chem.* **1995**, *16*, 1554−1566.
(64) Hoover, W. G. Canonical dynamics - equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(65) Boresch, S.; Bruckner, S. Avoiding the van der Waals endpoint problem using serial atomic insertion. *J. Comput. Chem.* **2011**, *32*, 2449−2458.
(66) Zacharias, M.; Straatsma, T. P.; Mccammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(67) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett'S acceptance ratio method. *J. Comput. Chem.* **2011**, *32*, 1303−1319.
(68) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, *32*, 1320−1333.
(69) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(70) Vidal, M.; Goudreau, N.; Cornille, F.; Cussac, D.; Gincel, E.; Garbay, C. Molecular and cellular analysis of Grb2 SH3 domain mutants: Interaction with Sos and dynamin. *J. Mol. Biol.* **1999**, *290*, 717−730.
(71) Burgoyne, N. J.; Jackson, R. M. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006**, *22*, 1335−1342.
(72) Shulman-Peleg, A.; Shatsky, M.; Nussinov, R.; Wolfson, H. J. Spatial chemical conservation of hot spot interactions in protein-protein complexes. *BMC Biol.* **2007**, *5*.
(73) Tuncbag, N.; Gursoy, A.; Keskin, O. Identification of computational hotspots in protein interfaces: Combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics* **2009**, *25*, 1513−1520.