PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129

Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking†

Fu Chen,[a] Hui Liu,[a] Huiyong Sun,[a] Peichen Pan,[a] Youyong Li,[c] Dan Li[a] and Tingjun Hou*[ab]

Received 27th May 2016,
Accepted 12th July 2016

DOI: 10.1039/c6cp03670h

www.rsc.org/pccp

Understanding protein–protein interactions (PPIs) is quite important to elucidate crucial biological processes and even design compounds that interfere with PPIs with pharmaceutical significance. Protein–protein docking can afford the atomic structural details of protein–protein complexes, but the accurate prediction of the three-dimensional structures for protein–protein systems is still notoriously difficult due in part to the lack of an ideal scoring function for protein–protein docking. Compared with most scoring functions used in protein–protein docking, the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) methodologies are more theoretically rigorous, but their overall performance for the predictions of binding affinities and binding poses for protein–protein systems has not been systematically evaluated. In this study, we first evaluated the performance of MM/PBSA and MM/GBSA to predict the binding affinities for 46 protein–protein complexes. On the whole, different force fields, solvation models, and interior dielectric constants have obvious impacts on the prediction accuracy of MM/GBSA and MM/PBSA. The MM/GBSA calculations based on the ff02 force field, the GB model developed by Onufriev *et al.* and a low interior dielectric constant (ε_{in} = 1) yield the best correlation between the predicted binding affinities and the experimental data (r_p = −0.647), which is better than MM/PBSA (r_p = −0.523) and a number of empirical scoring functions used in protein–protein docking (r_p = −0.141 to −0.529). Then, we examined the capability of MM/GBSA to identify the possible near-native binding structures from the decoys generated by ZDOCK for 43 protein–protein systems. The results illustrate that the MM/GBSA rescoring has better capability to distinguish the correct binding structures from the decoys than the ZDOCK scoring. Besides, the optimal interior dielectric constant of MM/GBSA for re-ranking docking poses may be determined by analyzing the characteristics of protein–protein binding interfaces. Considering the relatively high prediction accuracy and low computational cost, MM/GBSA may be a good choice for predicting the binding affinities and identifying correct binding structures for protein–protein systems.

Introduction

Protein–protein interactions (PPIs) play important roles in most crucial biological processes in living cells,[1–3] and many PPIs have even been regarded as potential drug targets as well.[4–6] A variety of experimental methods have been developed to determine whether a protein can interact with another, but most of the biophysical and/or biochemical associated techniques cannot afford the structural details to understand how two proteins interact.[7–11] A complete picture of how and where two proteins interact can only be provided by the three-dimensional structure of the protein–protein complex. Experimentally, X-ray crystallography and solution nuclear magnetic resonance (NMR)

[a] *College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China. E-mail: tingjunhou@zju.edu.cn, tingjunhou@hotmail.com; Tel: +86-571-8820-8412*
[b] *State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang 310058, P. R. China*
[c] *Institute of Functional Nano & Soft Materials (FUNSOM), Soochow University, Suzhou, Jiangsu 215123, China*

† Electronic supplementary information (ESI) available: Fig. S1. The scatter plots of the experimental binding affinities ($-\log(K_d)$) *versus* the scores predicted by four scoring functions; Table S1. The interaction features for the interfaces of 43 protein–protein complexes. See DOI: 10.1039/c6cp03670h

can be used to determine the structures of a protein–protein complex at the atomic level,[12] but it will be very difficult or even impossible to solve the high-resolution structures for all PPIs.[13] Alternatively, computational approaches have been developed to investigate how two proteins interact to form a three-dimensional complex.[14,15]

Protein–protein docking, which predicts the binding mode and binding affinity between individual protein structures, is a hopeful method for the prediction of PPIs.[16] In principle, the stable structure for a protein–protein complex can be determined computationally as the energy minimum on the potential surface.[12] Extensive efforts have been dedicated to develop feasible protein–protein docking approaches to detect and analyze PPIs *in silico*. However, it is an arduous task to determine the energy minima from the huge conformational space using molecular docking techniques.[12] Most protein–protein docking approaches consist of two stages: the docking stage and the ranking stage. In the docking stage, a number of structures are generated and the potential binding structures (or referred to as decoys) are sampled from numerous structures; and in the ranking stage, the decoys sampled from the first stage are scored by various scoring functions.[16]

The scoring function in protein–protein docking is supposed to have the capability to identify the most favorable binding poses from decoys, but in consideration of computational efficiency, approximations were usually introduced into the scoring functions, which unfortunately, often impair the prediction accuracy.[14,17] The scoring functions of protein–protein docking can be roughly categorized into two classes: force field-based scoring functions and knowledge-based scoring functions. A force field-based scoring consists of a few potential terms,[18,19] such as van der Waals interactions, electrostatic interactions, hydrophobic effects, desolvation energies, and entropic effects,[16,20] and the total energy of a conformation is calculated by summing up the contributions of all energy terms.[14] Knowledge-based scoring functions do not focus on the optimization of specified terms, but instead, they employ some optimization methods to get the best weights for each scoring term based on the training sets, and therefore these methods are also referred to as informatics-driven methods, such as IFACE,[21] DARS,[22] SPA-PP,[23] DrugScore^{PPI},[24] TS,[25] *etc.* In the past decade, more theoretically rigorous free energy calculation methods, such as free energy perturbation (FEP),[26] energy representation (ER),[12] Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA)[27] and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA),[28] have been employed to estimate protein–protein binding free energies. These methodologies are much more time-consuming than most scoring functions, whereas, with the rapid development of computer hardware, it is promising to use these methods in protein–protein docking in the near future.

Compared with thermodynamic integration (TI) and free energy perturbation (FEP) approaches, the MM/PBSA and MM/GBSA approaches are more computationally efficient and allow for the decomposition into different interaction terms.[27,29,30] In the spirit of MM/PBSA or MM/GBSA, the binding free energy for a protein–protein complex can be calculated using the following equations:

$$\Delta G_{\text{bind}} = G_{\text{com}} - (G_{\text{rec}} + G_{\text{lig}}) \quad (1)$$

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

$$\Delta E_{\text{MM}} = \Delta E_{\text{internal}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdw}} \quad (3)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}} \quad (4)$$

where ΔG_{bind} represents the total binding free energy upon protein–protein binding; ΔE_{MM} is the total gas phase energy (sum of $\Delta E_{\text{internal}}$, $\Delta E_{\text{electrostatic}}$, and ΔE_{vdw}); ΔG_{sol} is the sum of polar ($\Delta G_{\text{PB/GB}}$) and non-polar (ΔG_{SA}) contributions to solvation; and $-T\Delta S$ is the conformational entropy upon binding (usually calculated by normal-mode analysis). $\Delta E_{\text{internal}}$ is the internal energy arising from the bond, angle, and dihedral terms in the molecular mechanics (MM) force field (this term always amounts to zero in the MM/PBSA and MM/GBSA calculations based on the single trajectory of a complex). $\Delta E_{\text{electrostatic}}$ and ΔE_{vdw} are the electrostatic and van der Waals energies from MM calculations. $\Delta G_{\text{PB/GB}}$ is the polar contribution to the solvation free energy (calculated *via* the Poisson–Boltzmann (PB) or generalized Born (GB) method). ΔG_{SA} is the nonpolar solvation free energy, usually computed with a linear function of the solvent-accessible surface area (SASA). MM/GBSA and MM/PBSA have been successfully applied to predict the binding free energies for various protein–protein/peptide complexes[31–34] or identify the hot-spots at protein–protein binding interfaces,[35–37] but the previous studies mostly focused on certain specific systems and the prediction results cannot afford the overall accuracy of MM/PBSA and MM/GBSA for protein–protein systems.

Recently, based on a benchmark consisting of 81 protein–protein complexes[14] (46 in the revised paper[38]) with experimental binding affinity data, Kastritis and coworkers evaluated the performance of nine scoring functions implemented in well-established protein–protein docking programs along with a binding affinity prediction algorithm for protein–protein complexes.[14,38] Their results illustrated that the correlations between the experimental binding affinities and predicted scores were poor for all algorithms tested. When the complexes were categorized into the low, medium and high affinity classes, improved linear correlations could be observed. Apparently, for a specific group, the performance of a scoring function may be acceptable, but for the whole dataset no scoring function can give reliable predictions. Unfortunately, the MM/PBSA and MM/GBSA approaches were not included in the evaluation study reported by Kastritis *et al.*[14,38] Our group has comprehensively evaluated the performance of MM/PBSA and MM/GBSA to predict the binding affinities for extensive protein–ligand datasets and examined the rescoring capability of MM/PBSA and MM/GBSA in molecular docking.[29,39–42] However, to date, no work has been done to predict the binding affinities and identify the correct binding poses for protein–protein systems based on MM/PBSA and MM/GBSA approaches with a benchmark dataset. Therefore, in this study, based on Kastritis' dataset,[38]

we investigated the capability of MM/GBSA and MM/PBSA to predict the binding affinities for 46 protein–protein complexes. Then, we evaluated the rescoring capability of MM/GBSA to identify the near-native structures from the protein–protein decoys generated by ZDOCK.

Materials and methods

Benchmark datasets

The benchmark dataset (dataset I) to assess the ΔG_{bind} prediction capability of MM/PBSA and MM/GBSA was collected from the previous study reported by Kastritis *et al.*[14,38] The original dataset consists of 81 protein–protein complexes with experimentally determined binding constants (K_d).[14] However, after carefully examining the original experimental data, only 46 of the 81 reported experimental binding affinities could be considered fully accurate.[38] Then, the 46 protein–protein complexes listed in Table 1 were used in this study. Within this dataset, the Colicin E7 nuclease-Im7 immunity protein complex (PDB entry: 7CEI) has the highest binding affinity ($K_d = 5.0 \times 10^{-15}$ mol L^{-1}),[43] and the UEV domain–Ubiquitin complex (PDB entry: 1S1Q) has the lowest binding affinity ($K_d = 6.46 \times 10^{-4}$ mol L^{-1}).[44] The experimental binding affinities cover an extremely broad range of 11 orders of magnitude. The corresponding coordinates of the studied complexes were downloaded from the Protein Data Bank (PDB entries: 1E6J,[45] 1JPS,[46] 2FD6,[47] 2I25,[48] 2JEL,[49] 1BJ1,[50] 1IQD,[51] 1KXQ,[52] 1BVN,[53] 1DFJ,[54] 1E6E,[55] 1EZU,[56] 1IJK,[57] 1M10,[58] 1MAH,[59] 1ML0,[60] 1OPH,[61] 1R0R,[62] 2B42,[63] 2MTA,[64] 7CEI,[43] 1A2K,[65] 1AK4,[66] 1AKJ,[67] 1B6C,[68] 1E96,[69] 1GCQ,[70] 1GHQ,[71] 1H1V,[72] 1HE8,[73] 1I4D,[74] 1IBR,[75] 1J2J,[76] 1KXP,[77] 1QA9,[78] 1RLB,[79] 1S1Q,[44] 1T6B,[80] 1WQ1,[81] 1Z0K,[82] 2AJF,[83] 2BTF,[84] 2C0L,[85] 2HLE,[86] 2OOB[87] and 1GRN[88]).

The decoy dataset (dataset II) to assess the re-ranking capability of MM/GBSA was obtained from the decoy set generated by ZDOCK 3.0.2, one of the most popular programs for protein–protein docking.[89,90] ZDOCK employs a Fast Fourier Transform (FFT) algorithm to enable an efficient global docking search on a 3D grid, and utilizes a combination of shape complementarity, electrostatics and statistical potential terms for scoring.[12,91] The original dataset consists of the decoys for 176 protein–protein complexes. The decoys for each complex were generated by the Perl script offered by ZDOCK. For each system, the angular step size for the rotational sampling of the smaller protein (docked protein) was set to 15°, and then a total of 3600 decoy structures sorted by the ZDOCK scores were generated. Hits are defined as the predicted structures with the $RMSD_i$ (root mean square deviation of interface) value less than 2.5 Å from the crystal complex, where the $RMSD_i$ is defined as the interface C_a RMSD after a backbone-only least square fit of the predicted structure onto the target interfacial residues. The near-native structure is the hit with the lowest $RMSD_i$. In consideration of computational requirements and hardware capability, the systems whose near-native structure could not be found in top 100 structures were excluded from the original decoy set. Finally, 43 complexes (Table 2) were used to evaluate the re-ranking capability of MM/GBSA, including 1AZS,[92] 1BJ1,[50] 1BUH,[93] 1BVN,[53] 1CGI,[94] 1CLV,[95] 1DFJ,[54] 1E6E,[55] 1F51,[96] 1FSK,[97] 1GL1,[98] 1GPW,[99] 1I9R,[100] 1IQD,[51] 1J2J,[76] 1JTG,[101] 1JWH,[102] 1KXP,[77] 1KXQ,[52] 1MAH,[59] 1ML0,[60] 1N8O, 1NCA,[103] 1OYV,[104] 1PPE,[105] 1RV6,[106] 1TMQ,[107] 1UDI,[108] 1WDW,[109] 1WEJ,[110] 1XQS,[111] 1YVB,[112] 2AYO,[113] 2B42,[63] 2CFH,[114] 2G77,[115] 2I25,[48] 2NZ8,[116] 2OUL,[117] 2QFW,[118] 2SIC,[119] 4CPA[120] and 7CEI.[43]

Table 1 The pK_d values and types for 46 protein–protein complexes

PDB ID	R:L	pK_d	I-RMSD[a] (Å)	Type[b]	PDB ID	R:L	pK_d	I-RMSD (Å)	Type[a]
1E6J	HL:P	7.53	1.05	A	1AKJ	AB:DE	3.90	1.14	O
1JPS	HL:T	10.0	0.51	A	1B6C	A:B	5.55	1.96	O
2FD6	HL:U	9.00	1.07	A	1E96	A:B	5.22	0.71	O
2I25	N:L	9.00	1.21	A	1GCQ	B:C	4.77	0.92	O
2JEL	HL:P	8.55	0.17	AB	1GHQ	A:B	5.37	0.34	O
1BJ1	HL:VW	8.47	0.5	AB	1H1V	A:G	7.64	1.05	O
1IQD	AB:C	10.85	0.48	AB	1HE8	B:A	5.60	0.92	O
1KXQ	H:A	8.46	0.72	AB	1I4D	D:AB	5.52	1.41	O
1BVN	P:T	11.05	0.87	E	1IBR	A:B	9.52	2.54	O
1DFJ	E:I	13.23	1.02	E	1J2J	A:B	5.96	0.63	O
1E6E	A:B	6.07	0.96	E	1KXP	A:D	9.00	1.12	O
1EZU	C:AB	10.1	1.21	E	1QA9	A:B	5.05	0.73	O
1IJK	A:BC	7.64	0.68	E	1RLB	ABCD:E	6.87	0.66	O
1M10	A:B	8.24	2.1	E	1S1Q	A:B	3.19	0.98	O
1MAH	A:F	10.6	0.61	E	1T6B	X:Y	9.40	0.62	O
1ML0	ABC:H	8.51	1.02	E	1WQ1	R:G	4.77	1.16	O
1OPH	A:B	8.30	1.2	E	1Z0K	A:B	5.14	0.53	O
1R0R	A:C	9.47	0.45	E	2AJF	A:E	7.79	0.65	O
2B42	A:B	8.97	0.72	E	2BTF	A:P	5.70	0.75	O
2MTA	HL:A	5.35	0.41	E	2C0L	A:B	6.96	2.62	O
7CEI	A:B	14.3	0.70	E	2HLE	A:B	7.40	1.4	O
1A2K	C:AB	7.00	1.11	O	2OOB	A:B	4.22	0.85	O
1AK4	A:D	4.80	1.33	O	1GRN	A:B	6.41	1.22	O

[a] I-RMSD is the root-mean-square displacement of the C_a atoms of interface residues between the bound and unbound structures. [b] E represents the enzyme/inhibitor or the enzyme/substrate, A represents the antibody/antigen, AB represents the bound antibody/antigen, and O represents others.

Table 2 The ranks of the near-native structures predicted by ZDOCK and MM/GBSA based on different dielectric constants for the 43 tested complexes

ID	RMSD$_i$[a]	ZDOCK rank	ε_{in} = 1 rank	ε_{in} = 2 rank	ε_{in} = 4 rank	ε_{in} = 6 rank
1AZS	1.18	88	36	80	84	85
1BJ1	1.07	17	1	4	12	14
1BUH	1.53	37	18	56	63	64
1BVN	1.39	13	1	1	2	2
1CGI	2.27	86	21	21	22	23
1CLV	1.38	4	17	5	2	2
1DFJ	1.37	16	5	4	5	5
1E6E	1.42	44	11	24	43	49
1F51	1.13	31	51	24	24	24
1FSK	1.11	20	6	15	23	27
1GL1	1.46	54	15	5	6	5
1GPW	1.41	48	2	1	2	3
1I9R	2.28	31	38	45	53	55
1IQD	0.79	15	1	2	6	10
1J2J	2.18	42	31	32	42	39
1JTG	1.33	1	18	30	35	36
1JWH	1.9	31	28	44	52	51
1KXP	1.92	4	34	30	30	30
1KXQ	1.16	6	4	10	13	15
1MAH	1.02	1	1	1	3	5
1ML0	1.23	6	2	7	13	15
1N8O	1.27	13	9	17	23	26
1NCA	1.04	56	1	8	13	16
1OYV	1.3	23	3	4	6	7
1PPE	0.77	10	1	4	5	5
1RV6	1.6	48	32	51	62	64
1TMQ	1.78	25	24	28	27	29
1UDI	1.46	61	3	1	1	1
1WDW	1.54	1	1	1	1	1
1WEJ	1.16	20	2	6	10	12
1XQS	2.25	34	6	15	35	43
1YVB	1.07	6	27	30	38	37
2AYO	1.63	82	95	18	7	7
2B42	1.07	8	4	5	8	9
2CFH	1.82	1	5	3	3	3
2G77	1.79	36	55	18	16	16
2I25	1.74	52	40	72	81	82
2NZ8	2.45	76	89	33	26	25
2OUL	1.24	2	6	6	7	7
2QFW	1.58	71	17	50	54	53
2SIC	0.64	73	2	5	11	14
4CPA	1.2	1	11	7	8	8
7CEI	0.95	86	9	23	38	47

[a] RMSD$_i$ is the root mean square deviation of the interface between the near-native structure and crystal complex.

Preparation and minimization of the tested structures

For each complex in dataset I, the structure was downloaded from the RCSB Protein Data Bank, and then non-peptide molecules containing heteroatoms were removed and the missing residues were built by using MODELLER v9.15.[121] For each structure in dataset II, 'TER' was added into the location of the missing residues in order to satisfy the requirement of Amber.[122]

Before the MM/PBSA and MM/GBSA calculations, each protein–protein complex was prepared using the standard procedure by using the *tleap* module in Amber14.[122] Four different force fields, including ff02, ff03, ff99, and ff14SB, were assigned to the proteins, respectively.

The minimizations were performed by using the *sander* module in Amber14. Each system was immersed in a TIP3P water box with the water molecules extended 8 Å from any solute atom in each direction. Counterions, Na^+ or Cl^-, were added to neutralize the unbalanced charge in each complex. A cutoff distance of 12 Å was used for van der Waals interactions. Each complex was minimized by three steps: first, all the backbone heavy atoms of the protein were restrained with an elastic constant of 50 kcal mol^{-1} Å^{-2} (2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations); next, the elastic constant was weakened to 10 kcal mol^{-1} Å^{-2} (2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations); finally, the whole system was minimized for 5000 steps without any restraints (2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations).

Moreover, in order to compare the impacts of implicit and explicit solvent models, each complex structure was prepared using the *in vacuo* model and the modified GB model developed by Onufriev *et al.* was used in the minimizations.[123] The whole system was minimized for 5000 steps without any restraints (2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations).

Molecular dynamics simulations

In the molecular dynamics (MD) simulation process, the particle mesh Ewald (PME) algorithm was employed to handle the long-range electrostatic interactions.[124] All of the covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm,[125] and the time step was set to 2 fs. Each system was gradually heated from 0 to 300 K during a period of 50 ps in the *NVT* ensemble. Then, 1 ns equilibration simulation was performed in the *NTP* (T = 300 K and P = 1 atm) ensemble. The snapshots were generated at an interval of 10 ps, namely, 100 frames for each system. At last, 100 snapshots were collected for the following MM/PBSA and MM/GBSA calculations.

MM/PBSA and MM/GBSA calculations

The binding free energy for each system was calculated by the MM/PBSA and MM/GBSA approaches based on the minimized structure (the final structure derived from the minimization stage), or the 100 snapshots extracted from the 1 ns MD trajectory.

In the MM/PBSA calculations, the polar component of desolvation was calculated by the PB models. The grid size was set to 0.5 Å. The atomic radii optimized by Tan and Luo with respect to the reaction field energies computed in the TIP3P explicit solvents were used.[126] The partial charges of proteins used in the PB calculations were taken from the force field parameters.

In the MM/GBSA calculations, the polar component of desolvation was calculated by three GB models, including the GB model developed by Hawkins *et al.* (GB^{HCT}),[127] and two modified GB models developed by Onufriev *et al.* (referred to as GB^{OBC1} and GB^{OBC2}).[123] According to our previous study, the interior (solute) dielectric constant has a significant effect on the predictions of binding affinities.[42] Thus, four interior dielectric constants, 1, 2, 4 or 6, were used for both of the MM/GBSA and MM/PBSA calculations. The exterior (solvent) dielectric constant was set to 80. The non-polar component of desolvation was estimated by using the LCPO algorithm,[128]

where γ and b were set to 0.0072 and 0, respectively. Due to the expensive computational demand and low prediction accuracy in most cases, entropies were not considered here.[29,41]

Estimation methods

For dataset I, the Pearson correlation coefficient (r_p) was employed to evaluate the linear correlation between the predicted binding free energies (ΔG_{bind}) and the experimental data (K_d), and the Spearman ranking coefficient (r_s) was employed to evaluate the capability of MM/GBSA or MM/PBSA to rank the binding affinities.[41] For dataset II, the success rate was employed to evaluate the re-ranking efficiency of MM/GBSA, *i.e.* the proportion of the success case, determined by the near-native structure could be found in the top N predictions, *e.g.* if in the top 10 predictions, MM/GBSA can recognize the correct protein–protein binding poses for 5 out of 43 tested cases, the success rate of the top 10 predictions is 5/43 = 11.63%.

Results and discussion

Performance of MM/PBSA and MM/GBSA in predicting binding affinities with an implicit solvent model

First, the 46 complexes in dataset I were minimized by using four different Amber force fields (ff03, ff99, ff02 or ff14SB) with an implicit solvent model, and then their binding affinities were predicted by MM/PBSA or MM/GBSA. As shown in Table 3, when ε_{in} was set to 1 and the GB^{OBC1} model was used, the ff02 force field yields the best prediction with $r_p = -0.647$ and $r_s = -0.680$ and the other three force fields show similar ranking capabilities. Interestingly, although the force field recommended by Amber14 for proteins and nucleic acid simulation is ff14SB, it appears that ff14SB is not the best choice for the binding affinity prediction based on this benchmark dataset.

Similar to our previous studies,[29,39–42] here, we also compared the MM/GBSA predictions based on three different GB models. The three GB models include the pairwise GB model developed

Table 3 The correlation coefficients between the experimentally determined dissociation constants (pK_d) and the binding affinities predicted by MM/GBSA or MM/PBSA

		MM/GBSA GB^{HCT}		MM/GBSA GB^{OBC1}		MM/GBSA GB^{OBC2}		MM/PBSA	
Force field	ε_{in}	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
ff02[a]	1	−0.582	−0.643	−0.647	−0.680	0.123	0.158	−0.396	−0.438
	2	−0.524	−0.569	−0.575	−0.637	0.100	0.105	−0.511	−0.577
	4	−0.483	−0.537	−0.502	−0.548	0.024	0.003	−0.496	−0.545
	6	−0.469	−0.517	−0.480	−0.540	−0.079	−0.081	−0.481	−0.515
ff03[a]	1	−0.547	−0.571	−0.603	−0.646	0.137	0.182	−0.434	−0.391
	2	−0.515	−0.552	−0.544	−0.596	0.097	0.129	−0.523	−0.558
	4	−0.489	−0.539	−0.500	−0.542	−0.008	−0.036	−0.506	−0.547
	6	−0.480	−0.533	−0.487	−0.547	−0.155	−0.167	−0.495	−0.530
ff99[a]	1	−0.538	−0.601	−0.582	−0.633	0.123	0.145	−0.380	−0.363
	2	−0.507	−0.556	−0.538	−0.601	0.090	0.104	−0.498	−0.566
	4	−0.474	−0.522	−0.486	−0.524	0.007	−0.004	−0.487	−0.534
	6	−0.462	−0.512	−0.469	−0.525	−0.107	−0.122	−0.474	−0.506
ff14SB[a]	1	−0.536	−0.587	−0.578	−0.632	0.115	0.143	−0.375	−0.379
	2	−0.509	−0.552	−0.531	−0.598	0.073	0.088	−0.492	−0.549
	4	−0.472	−0.516	−0.482	−0.526	−0.044	−0.073	−0.483	−0.520
	6	−0.460	−0.510	−0.466	−0.522	−0.212	−0.208	−0.472	−0.497
ff02[b]	1	−0.463	−0.529	−0.303	−0.306	0.110	0.163	−0.211	−0.178
	2	−0.482	−0.550	−0.503	−0.591	0.094	0.145	−0.430	−0.483
	4	−0.456	−0.521	−0.482	−0.537	0.058	0.101	−0.465	−0.538
	6	−0.444	−0.508	−0.461	−0.507	0.015	0.051	−0.457	−0.505
ff14SB[b]	1	−0.429	−0.476	−0.321	−0.359	0.105	0.152	−0.168	−0.122
	2	−0.469	−0.546	−0.475	−0.567	0.083	0.138	−0.404	−0.462
	4	−0.451	−0.510	−0.464	−0.522	0.033	0.066	−0.435	−0.504
	6	−0.442	−0.508	−0.450	−0.523	−0.028	−0.003	−0.452	−0.502
ff02[c]	1	−0.632	−0.670	−0.625	−0.652	0.405	0.512	−0.638	−0.662
	2	−0.623	−0.665	−0.621	−0.651	0.285	0.368	−0.653	−0.683
	4	−0.613	−0.661	−0.612	−0.657	−0.147	−0.098	−0.633	−0.666
	6	−0.610	−0.659	−0.609	−0.661	−0.449	−0.455	−0.625	−0.669
ff02[d]	1	−0.502	−0.521	−0.569	−0.578	0.175	0.229	−0.501	−0.502
	2	−0.477	−0.513	−0.527	−0.560	0.156	0.192	−0.480	−0.497
	4	−0.454	−0.494	−0.475	−0.500	0.108	0.130	−0.459	−0.486
	6	−0.446	−0.499	−0.463	−0.503	0.043	0.020	−0.451	−0.500

[a] Minimization in the implicit solvent model. [b] Minimization in the explicit water model. [c] MD in the implicit solvent model. [d] MD in the explicit water model.

by Hawkins *et al.* (GB^{HCT}; *igb* = 1 in Amber's terminology)[127] and two modified GB models developed by Onufriev *et al.* (GB^{OBC1} and GB^{OBC2}; *igb* = 2 and 5, respectively).[123] According to the Pearson correlation coefficients and Spearman ranking coefficients, GB^{OBC1} performs the best and GB^{OBC2} performs the worst no matter which force field is used. According to the previous study, for the dataset with 59 small molecules interacting with six different proteins, GB^{OBC1} was the most successful model in ranking the binding affinities.[39] Therefore, for either protein–small molecule systems or protein–protein systems, the GB^{OBC1} model is the best choice for the MM/GBSA calculations.

It is well known that the MM/GBSA and MM/PBSA approaches are quite sensitive to the interior dielectric constant, which is quantitatively correlated with the contributions of solvent penetration to dielectric effects in the protein interior.[129–131] As shown in Table 3, taking the GB^{OBC1} model as an example, based on the interior dielectric constant of 6, the accuracies of MM/GBSA (r_p = −0.466 to −0.487) are even worse than those of some docking scoring functions (the best r_p = −0.529).[38] Fortunately, this situation can be substantially improved by using a relatively lower interior dielectric constant (1 or 2). For dataset I, ε_{in} = 1 yields the highest correlations for all four force fields (r_p = −0.578 to −0.647).

As shown in Fig. 1, the prediction of 1IBR is more accurate by using a lower ε_{in} (ε_{in} = 1), but it drifts further away from the regression line with a higher ε_{in}. On the basis of our previous analysis, when the binding pocket is hydrophobic and no ion–ion interactions are formed between the protein and ligand, the interior dielectric constant of 1 is the best choice.[39] We then calculated the properties of the protein–protein binding interface by using the COCOMAPS online service.[132] For 1IBR, the ratio of the polar buried area to non-polar buried area is relatively low (1.20), suggesting that ε_{in} = 1 may be preferred to 1IBR.

As shown in Table 3, the correlations (r_p) between the predicted ΔGs by MM/PBSA and the experimental pK_d values are −0.375 to −0.523. The correlations based on ε_{in} = 2 are the best (r_p = −0.492 to −0.523), which are slightly better than those based on ε_{in} = 4 (r_p = −0.483 to −0.506) or 6 (r_p = −0.472 to −0.495) and much better than those based on ε_{in} = 1 (r_p = −0.375 to −0.434). Apparently, MM/PBSA performs much worse than MM/GBSA, and even worse than some docking scoring functions. This observation is also consistent with our previous study.[40]

In summary, for the studied systems, based on the minimized structures with the ff02 force field, MM/GBSA based on the GB^{OBC1} model with a relatively lower interior dielectric constant (ε_{in} = 1) can provide the best predictions (r_p = −0.647 and r_s = −0.680). According to the predictions reported by Kastritis *et al.*,[38] for the same database, the correlation (r_p) between the experimental pK_d and the scores given by nine scoring functions ranges from −0.141 to −0.529. A more meaningful comparison was performed by using more recent scoring functions. The structures of 46 complexes were submitted to the CCharPPI server (http://life.bsc.es/pid/ccharppi), and the binding capabilities were estimated by ZRANK, ZRANK2, ROSETTADOCK, PYDOCK, FIREDOCK, SIPPER, AP_PISA and CP_PIE, including four scoring functions that were not used in the previous study.[38] As shown in Fig. S1 (ESI†), AP_PISA[133] yields the best predictions with r_p = −0.517 and r_s = −0.547, but it still performs worse than MM/GBSA (r_p = −0.647 and r_s = −0.680). Apparently, compared with the tested scoring functions, MM/GBSA has better capability to rank the binding affinities for different protein–protein systems.

Are minimizations based on the explicit water model necessary?

We then calculated the binding free energies based on the minimized structures with the TIP3P explicit solvent model and two different force fields (ff02 and ff14SB). As shown in Table 3, the predictions based on the ff02 force field and GB^{OBC1} with ε_{in} = 2 yield the best correlation (r_p = −0.503 and r_s = −0.591). The ff02 force field performs slightly better than the ff14SB force field. Similar to the predictions using minimized structures with the implicit solvent model, the predictions of GB^{HCT} and GB^{OBC1} are much better than those of GB^{OBC2} (r_p = −0.028 to 0.110 and r_s = −0.003 to 0.163). On the whole, the prediction accuracy of MM/GBSA or MM/PBSA using the minimized structures with the explicit water model is worse than that using the minimized structures with the implicit solvent model. The results show that the explicit water model is not quite necessary to improve the prediction capability of MM/GBSA or MM/PBSA.

Can relatively short MD simulations improve the predictions?

We then conducted 1 ns MD simulations for all the complexes and calculated the binding free energies based on the MD trajectories with the ff02 force field and two different solvent models (TIP3P explicit solvent and GB implicit solvent). As shown in Table 3, for MM/GBSA, the predictions based on the 1 ns MD trajectories with the implicit solvent are slightly better than those based on the minimized structures, with all $|r_p|$ for GB^{HCT} or GB^{OBC1} higher than 0.6 and $|r_s|$ higher than 0.65. The predictions of MM/GBSA based on GB^{HCT} or GB^{OBC1} are much better than those based on GB^{OBC2} ($|r_p|$ = 0.147–0.449 and $|r_s|$ = 0.098–0.512). Similarly, ε_{in} = 1 yields the best correlation no matter which GB model was used. On the other hand, for MM/PBSA, all $|r_p|$ given by MM/PBSA with the ff02 force field are higher than 0.62 and $|r_s|$ are higher than 0.66. The predictions based on the 1 ns MD trajectories can be significantly improved compared with those based on the minimized structures, indicating that MM/PBSA designed with a more physically rigorous model may be more sensitive to the conformation of the binding poses. Our previous studies on protein–small molecule systems also reached similar conclusions.[29]

As shown in Table 3, based on the 1 ns MD trajectories with the TIP3P explicit solvent, the predictions are better than those based on the minimized structures. For MM/GBSA, the GB^{OBC1} model and ε_{in} = 1 can provide the best predictions (r_p = −0.569 and r_s = −0.578). For MM/PBSA, the correlations of four ε_{in} are quite similar (r_p = −0.451 to −0.501 and r_s = −0.486 to −0.502). However, on the whole, the prediction accuracy is worse than those using the minimized structures or 1 ns MD trajectories

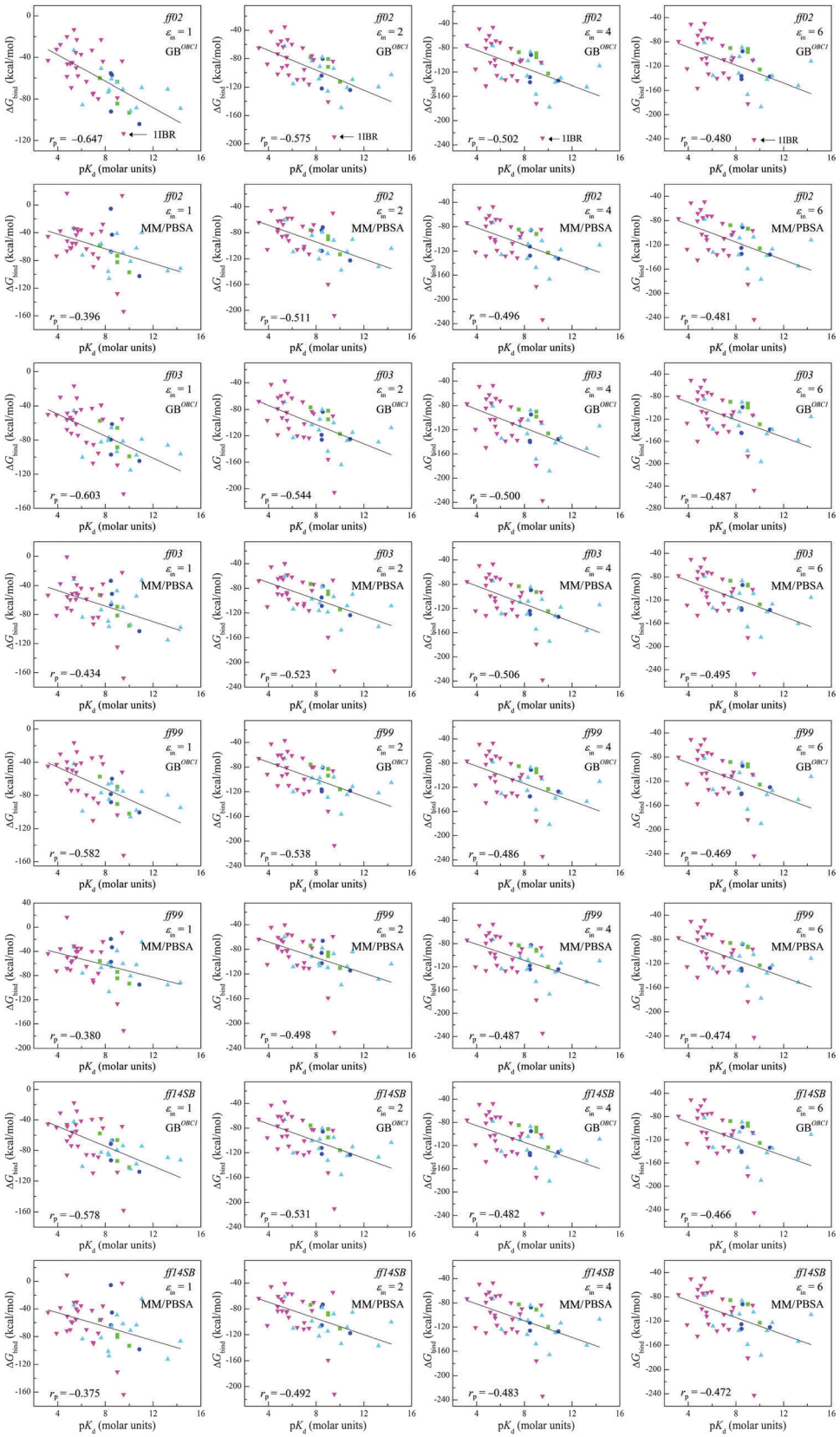

Fig. 1 Scatter plots of the experimental binding affinities ($-\log(K_d)$) (*x*-axis) *versus* binding free energies predicted by MM/GBSA or MM/PBSA (*y*-axis) based on the minimized structures for the 46 complexes. Different colors and shapes of the data points represent different categories. ▲ (cyan) is E (enzyme/inhibitor or enzyme/substrate), ■ (green) is A (antibody/antigen), ● (blue) is AB (bound antibody/antigen), ▼ (magenta) is O (others).

based on the implicit solvent model. The results show that, as for crystal structures, short MD simulations (1 ns) with the GB implicit solvent can improve the prediction accuracy, but in consideration of high computational cost, MD simulations may be not the best choice for some large systems.

Capability of MM/GBSA to recognize the correct protein–protein binding poses

In terms of the relatively high prediction accuracy and low computational cost, MM/GBSA based on the minimized structures with implicit solvent is the best choice for predicting the binding affinities of protein–protein interactions, suggesting that MM/GBSA has a great potential to identify the correct binding structures from the decoys generated by protein–protein docking. Therefore, we examined the re-ranking capability of MM/GBSA to the 43 protein–protein complexes in dataset II.

For each complex, the top 100 structures ranked by the ZDOCK scores were minimized with the implicit solvent model and rescored by MM/GBSA. The re-ranking capabilities of MM/GBSA based on four different interior dielectric constants (1, 2, 4 and 6) were compared as well. The rank levels are summarized in Table 2 and the success rates are shown in Fig. 2. As shown in Fig. 2, MM/GBSA rescoring can improve the capability to distinguish the near-native binding structures from the decoys given by the ZDOCK scores. For 43 protein–protein complexes, in terms of the top-10 level, the success rate is 30.23% for ZDOCK, while the success rate is promoted to 60.46% by the MM/GBSA rescoring. At a relatively lower interior dielectric constant (ε_{in} = 1 or 2) MM/GBSA can improve the ranking results given by ZDOCK for 33 protein–protein systems; while at a higher interior dielectric constant (ε_{in} = 4 or 6) MM/GBSA can improve the ranking results for 23 protein–protein systems. Apparently, at ε_{in} = 1 MM/GBSA has the best re-ranking capability (ε_{in} = 1 works well in 29 out of the 43 studied systems), and when ε_{in} becomes higher, MM/GBSA performs worse (ε_{in} = 2 works well in 25 systems, and ε_{in} = 4 or 6 works well in 22 systems).

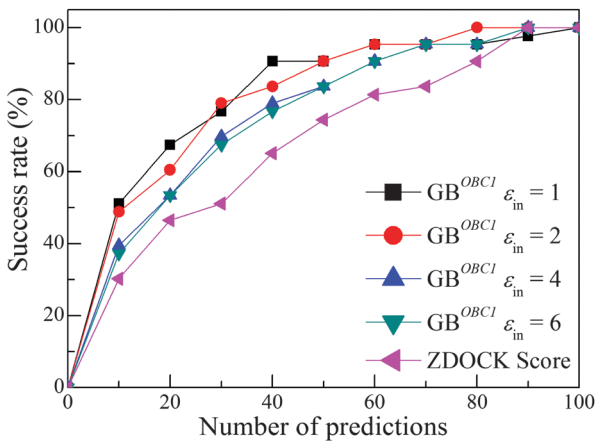

Fig. 2 Success rates (the proportion of the case in which the near-native structure could be found in the top N predictions) for MM/GBSA and ZDOCK scoring.

It is obvious that the prediction accuracy is sensitive to the interior dielectric constant. If the ranking level is taken as a criterion, the performance of MM/GBSA based on ε_{in} = 1 is the best and that based on ε_{in} = 6 is poor for 1AZS, 1BJ1, 1BUH, 1E6E, 1FSK, 1IQD, 1J2J, 1JWH, 1KXQ, 1MAH, 1ML0, 1N8O, 1NCA, 1OYV, 1PPE, 1RV6, 1TMQ, 1WEJ, 1XQS, 2B42, 2I25, 2QFW, 2SIC and 7CEI; in contrast, MM/GBSA based on ε_{in} = 6 performs the best and that based on ε_{in} = 1 performs the worst for 1CLV, 1F51, 1GL1, 2AYO, 2G77 and 2NZ8. For six systems, including 1BVN, 1CGI, 1DFJ, 1GPW, 1UDI and 1WDW, MM/GBSA performs well on all tested ε_{in}. For 2CFH, 2OUL and 4CPA, the rank levels given by MM/GBSA are slightly lower than those given by ZDOCK but still within the list of the top 10. Whereas, MM/GBSA performs worse on all tested ε_{in} for 1I9R, 1JTG, 1KXP and 1YVB. We thus investigated the protein–protein binding interfaces in order to find some clues to understand the performance of MM/GBSA with different interior dielectric constants for some systems. The properties of the protein–protein binding interfaces were calculated by using the COCOMAPS online service. The properties are listed in Table S1 (ESI†) and the data were ordered from low to high by the parameter of the polar buried area (PBA) upon complex formation. It is quite interesting that good correlation between the PBAs and the optimal interior dielectric constants can be observed. When complexes have relatively smaller PBAs ($<$1251.9 Å^2) a lower interior dielectric constant has better capability to identify possible near-native structures from the decoys. In contrast, in the case of the complexes with relatively larger PBAs ($>$1637.4 Å^2), ε_{in} = 6 has better performance. Thereafter, we suggest that the parameter of the PBA upon complex formation may be used as a useful measurement to help choose the optimal interior dielectric constant in the MM/GBSA calculations.

Conclusion

In this study, we examined the capability of MM/PBSA and MM/GBSA to predict the binding free energies and identify the correct binding poses for protein–protein complexes. Our conclusions are as follows:

(1) We evaluated the performance of MM/GBSA and MM/PBSA to estimate the binding free energies for 46 protein–protein complexes. The best MM/GBSA predictions yield the Pearson correlation coefficient of −0.647 for the experimentally determined pK_d, which is better than MM/PBSA (r_p = −0.523) and all the tested scoring functions used in protein–protein docking (r = −0.529 for the best one).

(2) A low interior dielectric constant is necessary to improve the rescoring accuracy for MM/GBSA. When the ff02 force field and the GB^{OBC1} model were used, MM/GBSA based on ε_{in} = 1 yields the best predictions with r_p = −0.647 and r_s = −0.680. The prediction accuracy of MM/GBSA decreases with the increase of ε_{in}.

(3) The polar buried area (PBA) of the protein–protein binding interface of a complex may be used as a criterion to choose the optimal interior dielectric constant for MM/GBSA to identify the correct binding structures from the decoys generated

by protein–protein docking. A lower interior dielectric constant is preferred to the complexes with relatively smaller polar buried areas, and a higher interior dielectric is preferred to the complexes with relatively larger polar buried areas.

Compared with most scoring functions used in protein–protein docking tools, MM/GBSA can achieve better accuracy for predicting the binding affinities and identifying correct binding structures for protein–protein systems. Therefore, using a rapid scoring scheme followed by the MM/GBSA rescoring may be an efficient protocol to improve the predictions of PPIs.

Acknowledgements

This study was supported by the National Science Foundation of China (21575128; 81302679), and the National Major Basic Research Program of China (2016YFA0501701 and 2016YFB0201700).

References

1 T. Vreven, H. Hwang, B. G. Pierce and Z. P. Weng, *Protein Sci.*, 2012, **21**, 396–404.
2 J. Andreani and R. Guerois, *Arch. Biochem. Biophys.*, 2014, **554**, 65–75.
3 G. Ramakrishnan, N. R. Chandra and N. Srinivasan, *IUBMB Life*, 2014, **66**, 759–774.
4 A. Metz, E. Ciglia and H. Gohlke, *Curr. Pharm. Des.*, 2012, **18**, 4630–4647.
5 D. Gonzalez-Ruiz and H. Gohlke, *Curr. Med. Chem.*, 2006, **13**, 2607–2625.
6 B. Nisius, F. Sha and H. Gohlke, *J. Biotechnol.*, 2012, **159**, 123–134.
7 S. A. I. Seidel, P. M. Dijkman, W. A. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, J. S. Joseph, P. Srinivasan, P. Baaske, A. Simeonov, I. Katritch, F. A. Melo, J. E. Ladbury, G. Schreiber, A. Watts, D. Braun and S. Duhr, *Methods*, 2013, **59**, 301–315.
8 Syafrizayanti, C. Betzen, J. D. Hoheisel and D. Kastelic, *Expert Rev. Proteomics*, 2014, **11**, 107–120.
9 A. Winter, A. P. Higueruelo, M. Marsh, A. Sigurdardottir, W. R. Pitt and T. L. Blundell, *Q. Rev. Biophys.*, 2012, **45**, 383–426.
10 R. Ghai, R. J. Falconer and B. M. Collins, *J. Mol. Recognit.*, 2012, **25**, 32–52.
11 A. G. N. Wetie, I. Sokolowska, A. G. Woods, U. Roy, K. Deinhardt and C. C. Darie, *Cell. Mol. Life Sci.*, 2014, **71**, 205–228.
12 K. Takemura, H. Guo, S. Sakuraba, N. Matubayasi and A. Kitao, *J. Chem. Phys.*, 2012, **137**, 215105.
13 I. Anishchenko, P. J. Kundrotas, A. V. Tuzikov and I. A. Vakser, *Proteins: Struct., Funct., Bioinf.*, 2014, **82**, 278–287.
14 P. L. Kastritis and A. Bonvin, *J. Proteome Res.*, 2010, **9**, 2216–2225.
15 I. A. Vakser and P. Kundrotas, *Curr. Pharm. Biotechnol.*, 2008, **9**, 57–66.
16 S. Y. Huang, *Drug Discovery Today*, 2014, **19**, 1081–1096.
17 G. Sowmya and S. Ranganathan, *Protein Pept. Lett.*, 2014, **21**, 779–789.
18 N. Basdevant, D. Borgis and T. Ha-Duong, *J. Phys. Chem. B*, 2007, **111**, 9390–9399.
19 J. Fernandez-Recio, M. Totrov and R. Abagyan, *Proteins: Struct., Funct., Genet.*, 2003, **52**, 113–117.
20 S. Y. Huang, S. Z. Grinter and X. Q. Zou, *Phys. Chem. Chem. Phys.*, 2010, **12**, 12899–12908.
21 J. Mintseris, B. Pierce, K. Wiehe, R. Anderson, R. Chen and Z. P. Weng, *Proteins: Struct., Funct., Bioinf.*, 2007, **69**, 511–520.
22 G. Y. Chuang, D. Kozakov, R. Brenke, S. R. Comeau and S. Vajda, *Biophys. J.*, 2008, **95**, 4217–4227.
23 Z. Q. Yan, L. Y. Guo, L. Hu and J. Wang, *Bioinformatics*, 2013, **29**, 1127–1133.
24 D. M. Kruger, J. I. Garzon, P. Chacon and H. Gohlke, *PLoS One*, 2014, **9**, e89466.
25 D. Tobi, *BMC Struct. Biol.*, 2010, **10**, 40.
26 B. O. Brandsdal and A. O. Smalas, *Protein Eng.*, 2000, **13**, 239–245.
27 N. Homeyer and H. Gohlke, *Mol. Inf.*, 2012, **31**, 114–122.
28 R. Chowdhury, M. Rasheed, D. Keidel, M. Moussalem, A. Olson, M. Sanner and C. Bajaj, *PLoS One*, 2013, **8**, e51307.
29 H. y. Sun, Y. y. Li, S. Tian, L. Xu and T. j. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.
30 J. M. Wang, T. J. Hou and X. J. Xu, *Curr. Comput.-Aided Drug Des.*, 2006, **2**, 287–306.
31 W. Wang and P. A. Kollman, *J. Mol. Biol.*, 2000, **303**, 567–582.
32 T. J. Hou, K. Chen, W. A. McLaughlin, B. Z. Lu and W. Wang, *PLoS Comput. Biol.*, 2006, **2**, 46–55.
33 H. Gohlke and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 238–250.
34 M. Ylilauri and O. T. Pentikainen, *J. Chem. Inf. Model.*, 2013, **53**, 2626–2633.
35 R. T. Bradshaw, B. H. Patel, E. W. Tate, R. J. Leatherbarrow and I. R. Gould, *Protein Eng., Des. Sel.*, 2011, **24**, 197–207.
36 J. Sorensen, D. S. Palmer and B. Schiott, *J. Agric. Food Chem.*, 2013, **61**, 7949–7959.
37 S. Fulle, C. Withers-Martinez, M. J. Blackman, G. M. Morris and P. W. Finn, *J. Chem. Inf. Model.*, 2013, **53**, 573–583.
38 P. L. Kastritis and A. Bonvin, *J. Proteome Res.*, 2011, **10**, 921–922.
39 T. J. Hou, J. M. Wang, Y. Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
40 T. j. Hou, J. m. Wang, Y. y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
41 L. Xu, H. y. Sun, Y. y. Li, J. m. Wang and T. j. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.
42 H. Y. Sun, Y. Y. Li, M. Y. Shen, S. Tian, L. Xu, P. C. Pan, Y. Guan and T. J. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22035–22045.
43 T. P. Ko, C. C. Liao, W. Y. Ku, K. F. Chak and H. S. Yuan, *Structure*, 1999, **7**, 91–102.
44 W. I. Sundquist, H. L. Schubert, B. N. Kelly, G. C. Hill, J. M. Holton and C. P. Hill, *Mol. Cell*, 2004, **13**, 783–789.

45 S. Monaco-Malbet, C. Berthet-Colominas, A. Novelli, N. Battai, N. Piga, V. Cheynet, F. Mallet and S. Cusack, *Structure*, 2000, **8**, 1069–1077.
46 K. Faelber, D. Kirchhofer, L. Presta, R. F. Kelley and Y. A. Muller, *J. Mol. Biol.*, 2001, **313**, 83–97.
47 Q. Huai, A. P. Mazar, A. Kuo, G. C. Parry, D. E. Shaw, J. Callahan, Y. D. Li, C. Yuan, C. B. Bian, L. Q. Chen, B. Furie, B. C. Furie, D. B. Cines and M. D. Huang, *Science*, 2006, **311**, 656–659.
48 R. L. Stanfield, H. Dooley, P. Verdino, M. F. Flajnik and I. A. Wilson, *J. Mol. Biol.*, 2007, **367**, 358–372.
49 L. Prasad, E. B. Waygood, J. S. Lee and L. T. J. Delbaere, *J. Mol. Biol.*, 1998, **280**, 829–845.
50 Y. A. Muller, Y. Chen, H. W. Christinger, B. Li, B. C. Cunningham, H. B. Lowman and A. M. de Vos, *Structure*, 1998, **6**, 1153–1167.
51 P. C. Spiegel, M. Jacquemin, J. M. R. Saint-Remy, B. L. Stoddard and K. P. Pratt, *Blood*, 2001, **98**, 13–19.
52 A. Desmyter, S. Spinelli, F. Payan, M. Lauwereys, L. Wyns, S. Muyldermans and C. Cambillau, *J. Biol. Chem.*, 2002, **277**, 23645–23650.
53 G. Wiegand, O. Epp and R. Huber, *J. Mol. Biol.*, 1995, **247**, 99–110.
54 B. Kobe and J. Deisenhofer, *Nature*, 1995, **374**, 183–186.
55 J. J. Muller, A. Lapko, G. Bourenkov, K. Ruckpaul and U. Heinemann, *J. Biol. Chem.*, 2001, **276**, 2786–2789.
56 S. A. Gillmor, T. Takeuchi, S. Q. Yang, C. S. Craik and R. J. Fletterick, *J. Mol. Biol.*, 2000, **299**, 993–1003.
57 K. Fukuda, T. A. Doggett, L. A. Bankston, M. A. Cruz, T. G. Diacovo and R. C. Liddington, *Structure*, 2002, **10**, 943–950.
58 E. G. Huizinga, S. Tsuji, R. A. P. Romijn, M. E. Schiphorst, P. G. de Groot, J. J. Sixma and P. Gros, *Science*, 2002, **297**, 1176–1179.
59 Y. Bourne, P. Taylor and P. Marchot, *Cell*, 1995, **83**, 503–512.
60 J. M. Alexander, C. A. Nelson, V. van Berkel, E. K. Lau, J. M. Studts, T. J. Brett, S. H. Speck, T. M. Handel, H. W. Virgin and D. H. Fremont, *Cell*, 2002, **111**, 343–356.
61 A. Dementiev, M. Simonovic, K. Volz and P. G. W. Gettins, *J. Biol. Chem.*, 2003, **278**, 37881–37887.
62 J. R. Horn, S. Ramaswamy and K. P. Murphy, *J. Mol. Biol.*, 2003, **331**, 497–508.
63 A. Pollet, S. Sansen, G. Raedschelders, K. Gebruers, A. Rabijns, J. A. Delcour and C. M. Courtin, *FEBS J.*, 2009, **276**, 4340–4351.
64 L. Y. Chen, R. C. E. Durley, F. S. Mathews and V. L. Davidson, *Science*, 1994, **264**, 86–90.
65 M. Stewart, H. M. Kent and A. J. McCoy, *J. Mol. Biol.*, 1998, **277**, 635–646.
66 T. R. Gamble, F. F. Vajdos, S. H. Yoo, D. K. Worthylake, M. Houseweart, W. I. Sundquist and C. P. Hill, *Cell*, 1996, **87**, 1285–1294.
67 G. F. Gao, J. Tormo, U. C. Gerth, J. R. Wyer, A. J. McMichael, D. I. Stuart, J. I. Bell, E. Y. Jones and B. K. Jakobsen, *Nature*, 1997, **387**, 630–634.
68 M. Huse, Y. G. Chen, J. Massague and J. Kuriyan, *Cell*, 1999, **96**, 425–436.
69 K. Lapouge, S. J. M. Smith, P. A. Walker, S. J. Gamblin, S. J. Smerdon and K. Rittinger, *Mol. Cell*, 2000, **6**, 899–907.
70 M. Nishida, K. Nagata, Y. Hachimori, M. Horiuchi, K. Ogura, V. Mandiyan, J. Schlessinger and F. Inagaki, *EMBO J.*, 2001, **20**, 2995–3007.
71 G. Szakonyi, J. M. Guthridge, D. W. Li, K. Young, V. M. Holers and X. J. S. Chen, *Science*, 2001, **292**, 1725–1728.
72 H. Choe, L. D. Burtnick, M. Mejillano, H. L. Yin, R. C. Robinson and S. Choe, *J. Mol. Biol.*, 2002, **324**, 691–702.
73 M. E. Pacold, S. Suire, O. Perisic, S. Lara-Gonzalez, C. T. Davis, E. H. Walker, P. T. Hawkins, L. Stephens, J. F. Eccleston and R. L. Williams, *Cell*, 2000, **103**, 931–943.
74 C. Tarricone, B. Xiao, N. Justin, P. A. Walker, K. Rittinger, S. J. Gamblin and S. J. Smerdon, *Nature*, 2001, **411**, 215–219.
75 I. R. Vetter, A. Arndt, U. Kutay, D. Gorlich and A. Wittinghofer, *Cell*, 1999, **97**, 635–646.
76 T. Shiba, M. Kawasaki, H. Takatsu, T. Nogi, N. Matsugaki, N. Igarashi, M. Suzuki, R. Kato, K. Nakayama and S. Wakatsuki, *Nat. Struct. Biol.*, 2003, **10**, 386–393.
77 L. R. Otterbein, C. Cosio, P. Graceffa and R. Dominguez, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8003–8008.
78 J. Wang, A. Smolyar, K. M. Tan, J. Liu, M. Y. Kim, Z. J. Sun, G. Wagner and E. L. Reinherz, *Cell*, 1999, **97**, 791–803.
79 H. L. Monaco, M. Rizzi and A. Coda, *Science*, 1995, **268**, 1039–1041.
80 E. Santelli, L. A. Bankston, S. H. Leppla and R. C. Liddington, *Nature*, 2004, **430**, 905–908.
81 K. Scheffzek, M. R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz and A. Wittinghofer, *Science*, 1997, **277**, 333–338.
82 S. Eathiraj, X. J. Pan, C. Ritacco and D. G. Lambright, *Nature*, 2005, **436**, 415–419.
83 F. Li, W. H. Li, M. Farzan and S. C. Harrison, *Science*, 2005, **309**, 1864–1868.
84 C. E. Schutt, J. C. Myslik, M. D. Rozycki, N. C. W. Goonesekere and U. Lindberg, *Nature*, 1993, **365**, 810–816.
85 W. A. Stanley, F. V. Filipp, P. Kursula, N. Schuller, R. Erdmann, W. Schliebs, M. Sattler and M. Wilmanns, *Mol. Cell*, 2006, **24**, 653–663.
86 J. E. Chrencik, A. Brooun, M. L. Kraus, M. I. Recht, A. R. Kolatkar, G. W. Han, J. M. Seifert, H. Widmer, M. Auer and P. Kuhn, *J. Biol. Chem.*, 2006, **281**, 28185–28192.
87 P. Peschard, G. Kozlov, T. Lin, A. Mirza, A. M. Berghuis, S. Lipkowitz, M. Park and K. Gehring, *Mol. Cell*, 2007, **27**, 474–485.
88 N. Nassar, G. R. Hoffman, D. Manor, J. C. Clardy and R. A. Cerione, *Nat. Struct. Biol.*, 1998, **5**, 1047–1052.
89 B. G. Pierce, Y. Hourai and Z. P. Weng, *PLoS One*, 2011, **6**, e24657.
90 H. Hwang, T. Vreven, J. Janin and Z. P. Weng, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 3111–3114.
91 B. G. Pierce, K. Wiehe, H. Hwang, B. H. Kim, T. Vreven and Z. P. Weng, *Bioinformatics*, 2014, **30**, 1771–1773.

92 J. J. G. Tesmer, R. K. Sunahara, A. G. Gilman and S. R. Sprang, *Science*, 1997, **278**, 1907–1916.
93 Y. Bourne, M. H. Watson, M. J. Hickey, W. Holmes, W. Rocque, S. I. Reed and J. A. Tainer, *Cell*, 1996, **84**, 863–874.
94 H. J. Hecht, M. Szardenings, J. Collins and D. Schomburg, *J. Mol. Biol.*, 1991, **220**, 711–722.
95 P. J. B. Pereira, V. Lozanov, A. Patthy, R. Huber, W. Bode, S. Pongor and S. Strobl, *Structure*, 1999, **7**, 1079–1088.
96 J. Zapf, U. Sen, Madhusudan, J. A. Hoch and K. I. Varughese, *Structure*, 2000, **8**, 851–862.
97 O. Mirza, A. Henriksen, H. Ipsen, J. N. Larsen, M. Wissenbach, M. D. Spangfort and M. Gajhede, *J. Immunol.*, 2000, **165**, 331–338.
98 A. Roussel, M. Mathieu, A. Dobbs, B. Luu, C. Cambillau and C. Kellenberger, *J. Biol. Chem.*, 2001, **276**, 38893–38898.
99 A. Douangamath, M. Walker, S. Beismann-Driemeyer, M. C. Vega-Fernandez, R. Sterner and M. Wilmanns, *Structure*, 2002, **10**, 185–193.
100 M. Karpusas, J. Lucci, J. Ferrant, C. Benjamin, F. R. Taylor, K. Strauch, E. Garber and Y. M. Hsu, *Structure*, 2001, **9**, 321–329.
101 D. Lim, H. U. Park, L. De Castro, S. G. Kang, H. S. Lee, S. Jensen, K. J. Lee and N. C. J. Strynadka, *Nat. Struct. Biol.*, 2001, **8**, 848–852.
102 K. Niefind, B. Guerra, I. Ermakowa and O. G. Issinger, *EMBO J.*, 2001, **20**, 5320–5331.
103 W. R. Tulip, J. N. Varghese, W. G. Laver, R. G. Webster and P. M. Colman, *J. Mol. Biol.*, 1992, **227**, 122–148.
104 I. H. Barrette-Ng, K. K. S. Ng, M. M. Cherney, G. Pearce, C. A. Ryan and M. N. G. James, *J. Biol. Chem.*, 2003, **278**, 24062–24071.
105 W. Bode, H. J. Greyling, R. Huber, J. Otlewski and T. Wilusz, *FEBS Lett.*, 1989, **242**, 285–292.
106 H. W. Christinger, G. Fuh, A. M. de Vos and C. Wiesmann, *J. Biol. Chem.*, 2004, **279**, 10382–10388.
107 S. Strobl, K. Maskos, G. Wiegand, R. Huber, F. X. Gomis-Ruth and R. Glockshuber, *Structure*, 1998, **6**, 911–921.
108 R. Savva and L. H. Pearl, *Nat. Struct. Biol.*, 1995, **2**, 752–757.
109 S. J. Lee, K. Ogasahara, J. C. Ma, K. Nishio, M. Ishida, Y. Yamagata, T. Tsukihara and K. Yutani, *Biochemistry*, 2005, **44**, 11417–11427.
110 S. E. Mylvaganam, Y. Paterson and E. D. Getzoff, *J. Mol. Biol.*, 1998, **281**, 301–322.
111 Y. Shomura, Z. Dragovic, H. C. Chang, N. Tzvetkov, J. C. Young, J. L. Brodsky, V. Guerriero, F. U. Hartl and A. Bracher, *Mol. Cell*, 2005, **17**, 367–379.
112 S. X. Wang, K. C. Pandey, J. R. Somoza, P. S. Sijwali, T. Kortemme, L. S. Brinen, R. J. Fletterick, P. J. Rosenthal and J. H. McKerrow, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11503–11508.
113 M. Hu, P. W. Li, L. Song, P. D. Jeffrey, T. A. Chernova, K. D. Wilkinson, R. E. Cohen and Y. G. Shi, *EMBO J.*, 2005, **24**, 3747–3756.
114 D. Kummel, J. J. Muller, Y. Roskel, N. Henke and U. Heinemann, *J. Mol. Biol.*, 2006, **361**, 22–32.
115 X. J. Pan, S. Eathiraj, M. Munson and D. G. Lambright, *Nature*, 2006, **442**, 303–306.
116 M. K. Chhatriwala, L. Bow, D. K. Worthylake and J. Sondek, *J. Mol. Biol.*, 2007, **368**, 1307–1320.
117 S. X. Wang, K. C. Pandey, J. Scharfstein, J. Whisstock, R. K. Huang, J. Jacobelli, R. J. Fletterick, P. J. Rosenthal, M. Abrahamson, L. S. Brinen, A. Rossi, A. Sali and J. H. McKerrow, *Structure*, 2007, **15**, 535–543.
118 Y. J. Peng, C. Zhong, W. Huang and J. P. Ding, *Protein Sci.*, 2008, **17**, 1542–1554.
119 Y. Takeuchi, Y. Satow, K. T. Nakamura and Y. Mitsui, *J. Mol. Biol.*, 1991, **221**, 309–325.
120 D. C. Rees and W. N. Lipscomb, *J. Mol. Biol.*, 1982, **160**, 475–498.
121 M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 291–325.
122 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
123 A. Onufriev, D. Bashford and D. A. Case, *J. Phys. Chem. B*, 2000, **104**, 3712–3720.
124 T. Darden, D. York and L. Pedersen, *J. Phys. Chem.*, 1993, **98**, 10089–10092.
125 J. P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
126 C. H. Tan, L. J. Yang and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 18680–18687.
127 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.*, 1996, **100**, 19824–19839.
128 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
129 P. L. Kastritis and A. Bonvin, *J. R. Soc., Interface*, 2013, **10**, 20120835.
130 W. L. DeLano, *Curr. Opin. Struct. Biol.*, 2002, **12**, 14–20.
131 J. J. Dwyer, A. G. Gittis, D. A. Karp, E. E. Lattman, D. S. Spencer, W. E. Stites and B. Garcia-Moreno, *Biophys. J.*, 2000, **79**, 1610–1620.
132 A. Vangone, R. Spinelli, V. Scarano, L. Cavallo and R. Oliva, *Bioinformatics*, 2011, **27**, 2915–2916.
133 S. Viswanath, D. V. S. Ravikant and R. Elber, *Proteins: Struct., Funct., Bioinf.*, 2013, **81**, 592–606.