ELSEVIER

Biophysical Chemistry 115 (2005) 255–261

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Novel knowledge-based potentials for ligand docking: variational approach to the old problem

A.M. Ruvinsky*, A.V. Kozintsev

Force Field Laboratory, Algodign, LLC, B. Sadovaya, 8, 103379, Moscow, Russia

Received 20 May 2004; received in revised form 1 November 2004; accepted 10 December 2004
Available online 23 December 2004

Abstract

The variational approach of evaluation for knowledge-based potentials is considered for the first time. In this approach, the problem to derive knowledge-based potentials is solved as the optimization task in the multiparametric model of atom types, reference states and interaction cutoff radii. Using analogy to liquid state theory we offered four new reference states and derived corresponding knowledge-based potentials. The cutoff radii and atom types are optimized to minimize averaged root-mean square deviations (RMSD) of the ligand docked positions regarding to the experimentally determined poses. The number of atom types is varied on the developed atom type tree with 6 root (C, N, O, S, P and the halogen type) and 49 apical atom types. We showed a pronounced effect of atom type choice on docking accuracy and proved that splitting of elements C, N and O of the periodic system up to the 18 optimal atom types essentially improves docking accuracy.

Keywords: Knowledge-based potentials; Protein–ligand binding; Docking

1. Introduction

The development of accurate and fast methods of predicting ligand docked positions on the protein surface and computing free energy difference of protein–ligand binding is a field of significant practical importance in biology, physics and chemistry. However at present binding free energy cannot not be obtained exactly from ab initio calculations and statistical mechanics, and therefore various approximate methods [1,2] have been used to calculate protein–ligand binding free energy change ΔG. These methods fall into three main categories: force-field-based approaches [3–8], knowledge-based approaches [9–14] and regression-based (empirical) approaches [15–18]. De novo drug design and virtual database screening [19] demand fast estimates of binding affinity of millions of generated ligands to the protein targets. In comparison with knowledge-based and regression-based approaches physically more grounded force-field methods require more computational time but do not achieve essentially better accuracy. Knowledge-based potentials and empirical scoring functions are popular due to the more simple procedures of their derivation and application. The continuous increasing of the quantity of known protein–ligand structures in Protein Data Bank (PDB) [20] and known values of protein–ligand binding constants inspire following development of both methods.

The knowledge-based method have attracted a great interest after the paper of Tanaka and Scheraga [21], first proposed to derive statistical potentials from amino acid pairing frequencies observed in protein structures. Their method was developed in [22,23] and applied to the folding problem and protein structure recognition [24–26], mutant activity analysis [27], to the problems of docking/scoring of small-molecule ligands into protein binding sites [9–14].

* Corresponding author. Tel.: +7 95 209 14 35; fax: +7 95 209 14 15.
E-mail addresses: anatoly.ruvinsky@algodign.com (A.M. Ruvinsky), alexander.kozintsev@algodign.com (A.V. Kozintsev).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.12.029

The common knowledge-based approach grounds on the debated idea of the atom pair-wise form of protein–ligand binding free energy [28]

$$\Delta G = \sum_{i,j} U_{ij}(r_{ij}), \tag{1}$$

where $U_{ij}(r_{ij})$ characterizes the interaction between ligand and protein atoms i and j located at the distance r_{ij}. The pair-wise scoring functions (1) allow to realize high-speed calculations of binding affinity (e.g. in virtual screening experiments), but at the same time a number of entropic effects, which are not pair-wise functions of atom-to-atom relative positions per se, fall out of the consideration.

Force-field-based models [3–8,29] utilize a decomposition of the protein–ligand binding free energy in the form of

$$\Delta G = \Delta H + \Delta G_{\mathrm{np}} + \Delta G_{\mathrm{tr}} + \Delta G_{\mathrm{v}} + \Delta G_{\mathrm{H}}, \tag{2}$$

where ΔH is the enthalpy change; ΔG_{np} is the nonpolar solvation contribution, accounting for the hydrophobic effect; ΔG_{tr} is the entropy contribution of rotational and translational overall and relative motions [7,30–33]; ΔG_{v} arises due to the changes of molecular eigenfrequencies during binding [8]; ΔG_{H} determines entropy contributions of protonation/deprotonation events [34–36]. Empirical approaches [15–18] use the similar to the Exp. (2) form of the free energy decomposition. Contrary to the knowledge-based approach (1), only ΔH and ΔG_{np} in the Exp. (2) can be considered as atom pair-wise terms.

Entropy terms ΔG_{tr}, ΔG_{H} and ΔG_{v} are essentially not pair-wise functions and thus cannot be considered with traditionally derived knowledge-based potentials. The estimates of the lost effect only due to the ΔG_{tr} are as follows: 3.6–4.8 kcal/mol [33], 4.5–6 kcal/mol [7], 7 kcal/mol [6], 15 kcal/mol [30]. The conformational entropy changes do not accounted in the common knowledge-based scheme, also should be included additionally especially for large flexible ligands [14,37]. Calorimetric investigations [34–36] show pH dependence of the molecular charged states and protein–ligand binding affinity, but the knowledge-based free energy change (1) does not depend on pH.

Thus, it makes an evidence of the necessity to reformulate the knowledge-based approach. We think that following more physical methods an application of knowledge-based potentials should also use free energy decomposition like the Exp. (2) with additionally introduced entropy terms. We suppose that the physically more-grounded form of knowledge-based scoring functions can improve their predictive ability. In this paper we consider the pair-wise term $\Delta H+\Delta G_{\mathrm{np}}$ significant for ligand docking. Non pair-wise terms in the Exp. (2) will be derived elsewhere.

2. Methods and materials

2.1. Construction of training and test sets

To construct training and test sets we choose from the PDB entries with protein–ligand structures with resolution better than 2.5 Å. PDB entries with theoretical models, with missing atoms, residues or with unknown atoms (marked as UNX in the PDB ent-format) in the active site, with only main chain coordinates were not selected. Also we excluded PDB entries with metals, with membrane and fibrillar structures, complexes with DNA or RNA. All selected structures contain noncovalently bound ligand molecules with 6–60 heavy atoms. This yielded 1937 entries. Then 130 entries with only one protein chain and one ligand molecule, containing no more than 30 heavy atoms, were selected for the test set. Remaining 1807 entries were used as the training set to derive potentials.

2.2. Atom types

In the knowledge-based approach atom typification aims to describe interatomic and intermolecular interactions averaged over electron density and hydrogen distribution, if hydrogen atom types are not introduced. We developed an atom type tree with 6 root (C, N, O, S, P and one halogen type HAL) and 49 apical atom types (C-14, O-9, N-16, S-5, P-1, F-1, Cl-1, Br-1, I-1).[1] Rising the tree the atom types gradually split (see, for example, Fig. 1) accordingly with the following criteria: the element name; the atom hybridization and inclusion in aromatic rings; the number of bonds with polar atoms (for carbon types); the number of bound heavy atoms; donor–acceptor properties; the atom ability to be in a charged state; the inclusion in mostly prevailing functional groups (e.g. amidines, amides, carboxyles) for N and O.

For the atom typification of PDB protein–ligand structures we used a PDB het-dictionary in the mmCIF format [20]. Every ligand in PDB can be subdivided into fragments (het-groups) described in the PDB het-dictionary. The het-dictionary in the mmCIF format contains data on bond orders for all atom pairs in the het-groups. The atom typification was realized in two stages. At first we typified the PDB het-dictionary. Then atom types were transferred to the PDB entries and contact atoms between different het-groups and their neighbors were additionally typified.

2.3. Knowledge-based potentials

The derivation of knowledge-based potentials is similar to the introduction of potentials of mean force (PMF) in liquid state theory. PMF are measures of interatomic or intermolecular interactions that are derived from the corresponding

[1] The full description of the atom types is available upon request to authors.

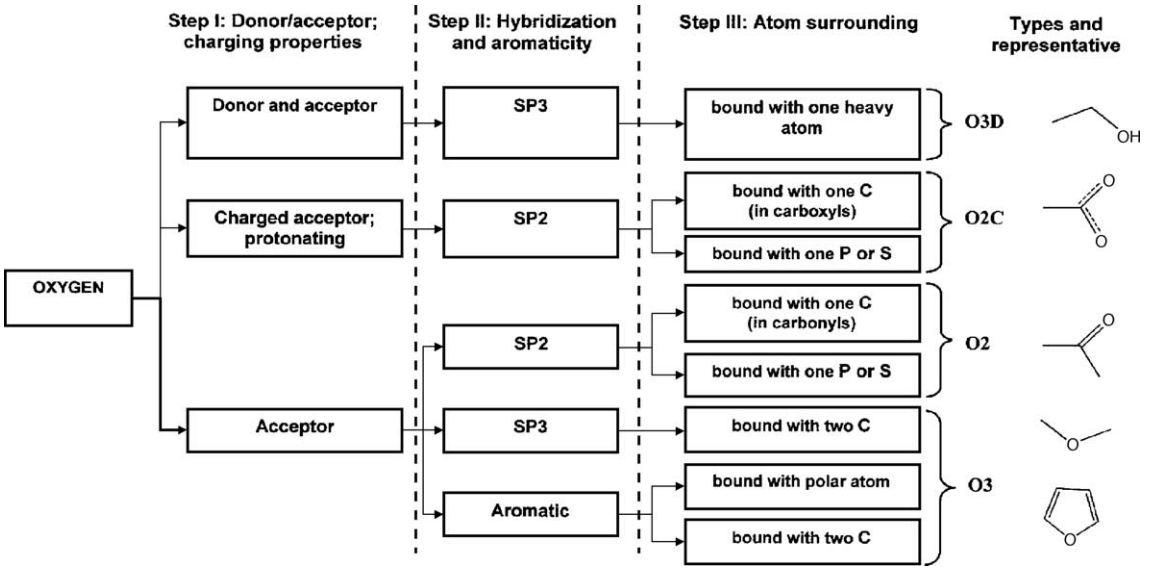

Fig. 1. The atom type tree for oxygen.

pairing distribution functions [38] $G(r)=\exp(-U_{12}(r)/T)$, where $G(r)$ defines the probability $\Delta W(r)$ to find a particle in the spherical layer Δr at the distance r from another fixed particle as

$$\Delta W(r) = G(r)\frac{4\pi r^2 \Delta r}{V}, \tag{3}$$

where V is the volume of homogeneous and isotropic liquid. Let us rewrite the Exp. (3) in the form of

$$\Delta W(r) = G(r)P_1P_2(r), \tag{4}$$

where $P_1=nV/N=1$ (n is the particle concentration, N is the total number of the particles) is the probability to find the fixed particle and, $P_2(r)=4n\pi r^2\Delta r/nV=4\pi r^2\Delta r/V$ is the probability to find another particle in the spherical shell with the radius r.

The Exp. (4) allows to draw a bridge of the analogy between liquid state theory and the knowledge-based approach, introducing conception of the reference state (see Refs. [9–14,39]) as P_1P_2 in Exp. (4), where $P_{1,2}$ are the probabilities to find a contact between the ligand (protein) atom type 1(2) and an arbitrary protein (ligand) atom type. Further we obtain structural potentials using four reference states, which differ from the known ones [9–14].

2.3.1. Reference state 1: the volume probability depends on ligand atom types

The correlation function $G_{ab}(r)$ is defined by

$$P_{ab}(r) = P_a(r)P_bG_{ab}(r), \tag{5}$$

where

$$G_{ab}(r) = \exp(-U_{ab}(r)/T); \tag{6}$$

$$P_{ab}(r) = \frac{N_{ab}(r)}{\sum_{abr} N_{ab}(r)} \tag{7}$$

is the probability to find a contact between the ligand atom type A and the protein atom type B; $N_{ab}(r)$ is a number of contacts between atoms A and B, located at the distance r in the training set. The summation in the Exp. (7) runs over all possible distances r. We do not introduce the radii for statistics gathering as were done in [9–14]. The number of contacts $N_{ab}(r)$ was calculated for each increment of 0.2 Å. In the Exp. (5) P_b is the probability to find the protein atom type B in a contact with an arbitrary ligand atom

$$P_b = \frac{\sum_{a,r} N_{ab}(r)}{\sum_{abr} N_{ab}(r)} = \frac{\sum_s N^{l(s)}N_b^{p(s)}}{\sum_s N^{l(s)}N^{p(s)}}, \tag{8}$$

where $N^{l,p(s)}$ is the number of ligand (protein) atoms in the complex with the number s in the training set, $N_{a,b}^{l,p(s)}$ is the number of ligand (protein) atoms A (B) in the complex s. The probability to find the ligand atom A in a contact with an arbitrary protein atom at the distance r is

$$P_a(r) = \frac{\sum_s \sum_{k=1}^{N_a^{l(s)}} \Delta V_k^{a(s)}(r)}{\sum_s N^{l(s)}V^{(s)}}, \tag{9}$$

where $V^{(s)}$ is a volume of the protein in the complex s, $\Delta V_k^{a(s)}(r)$ is a protein volume cut by the spherical shell with the radius r, thickness Δr and the shell center in the ligand atom k of the type A. $P_a(r)$ accounts for the volume depending part of the probability. In the Exp. (6) we use T=0.6 kcal/mol. Exp. (9) can be simplified under an assumption of equal atom volumes of protein atoms. In that case we have

$$V^{(s)} \approx N^{p(s)}v_p, \quad \Delta V_k^{a(s)}(r) \approx v_p \sum_i N_{ki}^{a(s)}(r), \tag{10}$$

where v_p is the average volume of a protein atom, $N_{ki}^{a(s)}(r)$=1 for the atom contact at the distance r between the ligand atom k (type A) and the protein atom i, and

$$P_a(r) = \frac{\sum_b N_{ab}(r)}{\sum_s N^{l(s)} N^{p(s)}} \quad (11)$$

2.3.2. Reference state 2: the volume probability does not depend on atom types

The correlation function $G_{ab}(r)$ is defined by

$$P_{ab}(r) = P_a P_b P(r) G_{ab}(r), \quad (12)$$

where $P(r) = \sum_{a,b} N_{ab}(r) / \left(\sum_s N^{l(s)} N^{p(s)} \right)$ is the probability to find a contact between arbitrary protein and ligand atoms at the distance r; $Pa = \sum_{b,r} N_{ab}(r) / \left(\sum_{abr} N^{ab}(r) \right)$ is the probability to find the ligand atom type A in a contact with an arbitrary protein atom.

2.3.3. Reference state 3: the volume probability depends on protein atom types

The correlation function $G_{ab}(r)$ is defined by

$$P_{ab}(r) = P_a P_b(r) G_{ab}(r), \quad (13)$$

where $P_b(r) = \sum_a N_{ab}(r) / \left(\sum_s N^{l(s)} N^{p(s)} \right)$ is the probability to find the protein atom type B at the distance r from an arbitrary ligand atom. The probabilities $P_b(r)$ are defined under an assumption of equal volumes of ligand atoms.

2.3.4. Reference state 4: random reference state

The correlation function $G_{ab}(r)$ is defined by

$$P_{ab}(r) = P^*_{ab}(r) G_{ab}(r), \quad (14)$$

where $P^*_{ab}(r) = N^*_{ab}(r) / \left(\sum_s N^{l(s)} N^{p(s)} \right); N^*_{ab}(r)$ is a number of contacts between atoms A and B, located at the distance r in the training set with random-mixed atoms in molecules. The random-mixed set was obtained from the training set of 1807 entries by the random permutation of heavy atoms in molecules.

3. Results

Deriving potentials on 49 atom types, we found that some of potentials are strongly unsteady functions and thus not reliable. The most evident unsteady character of potential functions was found for atom types with the concentration lower than 1% in the training set. Descending the atom type tree and uniting atom types with similar properties, we got a set of 18 atom types with the content of every atom type not less than 1%: C3—nonpolar sp3 carbon; C3P—polar sp3 carbon bound with at least one atom other than C, H; CR—aromatic nonpolar carbon; CRP—aromatic polar carbon; C2—nonpolar sp2 carbon; C2P—polar sp2 carbon (e.g., bound to carbonyl oxygen); O2C—oxygen in negatively charged group (e.g., carboxylate), acceptor and donor; O2—oxygen as a hydrogen bond acceptor, not charged (e.g., keto, amide oxygen); O3—oxygen bound with two heavy atoms (in an ether bond), acceptor; O3D—oxygen bound to hydrogen, acceptor and donor; NN—nitrogen as neither donor nor acceptor; NA—nitrogen as the hydrogen bond acceptor; NB—nitrogen as the donor or acceptor (e.g., histidine); NDC—charged donor (e.g., NH_3+ or guanidino groups); ND—nitrogen as a hydrogen bond donor only (e.g., amide nitrogen); P—phosphorus; S—sulfur; HAL—halogens F, Cl, Br and I.

To make a choice between the reference states we inspected docking accuracy of the corresponding potentials on the test set of 130 PDB complexes. Three-dimensional grids (18×18×18 Å) of 0.12 Å resolution were used to compute atom-to-atom interaction energies. Protein–ligand energies (1) were minimized with the direction set (Powell's) method in 6 dimensions of relative translational and rotational motions.

Table 1 shows results of rigid docking with four potentials for the 18 atom types and the optimal cutoff radii in terms of the percentage of the top-ranked (lowest-energy) pose within a defined RMSD from the experimentally determined pose. The numeration of the reference states corresponds to their ordering in Methods and materials section. The increasing or decreasing of the cutoff radii (Fig. 2) from the values, stated in Table 1, makes worse docking accuracy. Here we supposed that the cutoff radii do not depend on atom types. Also we used the potential of the rigid sphere in the repulsion region of $N_{ab}(r)$=0. It may be that more accurate docking results can be derived as a result of variation of cutoff radii for all atom type pairs separately and the substitution of potential walls by the variational functions.

The results in Table 1 indicate that potentials derived with the random reference state have the higher percentage ratios. To account effect of multiple binding modes of a ligand in a protein binding site we analyzed the percentage of best solutions (closest to the X-ray structure) selected from the top ten of all ranked solutions (Table 2). These results also demonstrate that potentials based on the random reference state perform better than other ones.

Table 1
Percentage of the top-ranked solutions within a defined RMSD from the experimentally determined pose

RMSD, Å	Ref. state 1	Ref. state 2	Ref. state 3	Ref. state 4
≤0.5	16	29	17	30
≤1	26	48	31	50
≤1.5	31	52	40	55
≤2.0	37	53	44	57
≤2.5	39	59	53	62
≤3.0	41	62	60	66
Cutoff radius, Å	5.0	4.0	4.5	4.0
Averaged RMSD, Å	4.51	3.04	3.33	2.84

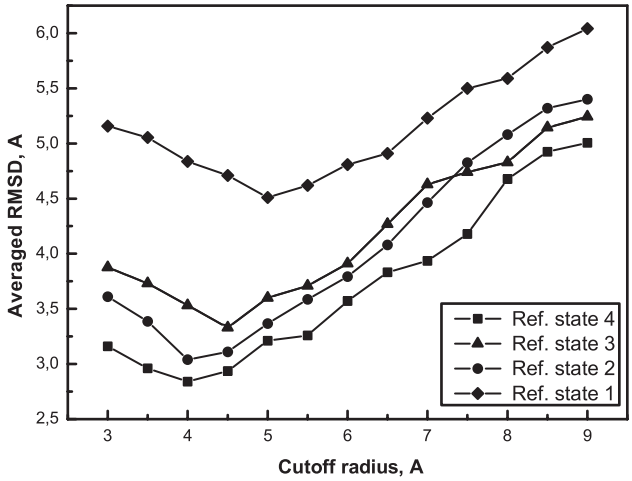

Fig. 2. Averaged RMSD (heavy atoms) of the top-ranked solution pose from the experimentally determined pose depending on the cutoff radius.

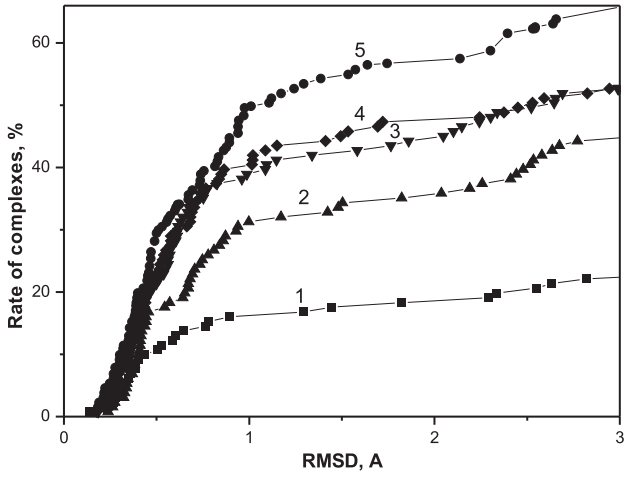

Fig. 3. Percentage of the top-ranked solutions within a defined RMSD from the experimentally determined pose for different number of atom types. Curves 1, 2, 3, 4, 5 correspond to 6 (C, O, N, S, P, HAL), 9, 12, 14, 18 atom types.

The docking results show that the main effect comes from protein and ligand atoms located at the distances not greater than 4–5 Å. We derived potentials for offered reference states with the account of volumes of protein surface atoms, obtained in Ref. [40], and got practically identical potentials and the same docking results as in the case of equal atom volumes.

Further we test the performance of the potentials evaluated with the random reference state (N 4 in Tables 1 and 2) in relation to subsequent atom type unification. Fig. 3 shows the percentage of the top-ranked solutions within a defined RMSD from the experimentally determined poses for different number of atom types, starting from 6 atom types (C, N, O, P, S, HAL) and up to 18 atom types.

The worst docking results are obtained for atom types uniquely corresponding to the chemical elements C, N, O, P, S and one halogen type HAL (Fig. 3, curve 1). Docking accuracy essentially improves for splitting of the single carbon type (Fig. 3, curve 1) up to 4 types (Fig. 3, curve 2) according to the atom hybridization and inclusion in aromatic systems. Further splitting of carbon types accounting bonds with polar atoms (not C) also increases success rates (Fig. 3, curve 3). The oxygen splitting (Fig. 3, curve 4) in accordance with the atom hybridization and the donor–acceptor classification rises the curve on 3–5%. The next step (Fig. 3, curve 5) is the nitrogen splitting into 5 types, considering donor–acceptor properties. It is interesting to note that simultaneous splitting of oxygen and nitrogen raises the percentage a lot more effectively in comparison with the case of 4 oxygen types and the 1 unsplit nitrogen (Fig. 3, curve 4). It means that only the simultaneous splitting of O and N adequately describes potentials of hydrogen bonds. Further we tested alternative choices of atom types, obtained for another atom tree pathes, and variants of new atom types introduction in the system of the 18 atom types, but did not find improvement of docking accuracy. Thus, on the basis of docking results we obtained 18 optimal atom types.

Table 2
Percentage of the best solutions (closest to the X-ray pose) selected from the top ten of all solutions within a defined RMSD from the experimentally determined pose

RMSD, Å	Ref. state 1	Ref. state 2	Ref. state 3	Ref. state 4
≤0.5	22	37	21	39
≤1	47	67	52	66
≤1.5	59	73	65	75
≤2.0	63	75	73	79
≤2.5	67	79	79	81
≤3.0	69	83	84	86
Cutoff radius, Å	5.0	4.0	4.5	4.0
Averaged RMSD, Å	2.54	1.61	1.64	1.39

4. Conclusions

We derived structure-based potentials in the frame-work of the optimization problem in the multiparametric model of atom types, interaction radii and reference states. The values of these parameters were optimized to minimize RMSD in docking tests. We showed that potentials derived with the random reference state perform better than others. We obtained the set of optimal atom types. We found a connection between the size of the training set and the number of statistically reliable atom types. The set of optimal atom types can be enlarged with the following expansion of the PDB, but most likely there is no sense to introduce atom types with the abundance lower than 1% of the training set.

Developed potentials assess pair-wise part of binding free energy and thus can be successful in the definition of ligand docked positions. The offered variational approach to evaluate potentials and to construct scoring functions with non pair-wise terms provides a basis for a promising hybrid approach of knowledge-based and regression-based methods.

Acknowledgements

We thank to A.M. Dykhne, M. Levitt, S.M. Nikitin, V.D. Ozrin, D.N. Tarasov, E.V. Tkalya, N.I. Zaitseva and V.V. Zosimov for interesting discussions.

References

[1] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, Angew. Chem. Int. Ed. 41 (2002) 2644–2676.
[2] T. Lazaridis, Binding affinity and specifity from computational studies, Curr. Org. Chem. 6 (2002) 1319–1332.
[3] P. Kollman, Free energy calculations: applications to chemical and biochemical phenomena, Chem. Rev. 93 (1993) 2395–2417.
[4] D. Qui, P.S. Shenkin, F.P. Hollinger, W.C. Still, The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii, J. Phys. Chem. 101 (1997) 3005–3014.
[5] G. Archontis, T. Simonson, M. Karplus, Binding free energies and free energy components from molecular dynamics and poisson–boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase, J. Mol. Biol. 306 (2001) 307–327.
[6] R. Luo, M.K. Gilson, Synthetic adenine receptors: direct calculation of binding affinity and entropy, J. Am. Chem. Soc. 122 (2000) 2934–2937.
[7] T. Lazaridis, A. Masunov, F. Gandolfo, Contributions to the binding free energy of ligands to avidin and streptavidin, Proteins: Struct., Funct., Genet. 47 (2002) 194–208.
[8] S.M. Schwarzl, T.B. Tschopp, J.C. Smith, S. Fischer, Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? J. Comput. Chem. 23 (2002) 1143–1149.
[9] G. Verkhivker, K. Appelt, S.T. Freer, J.E. Villafranca, Empirical free energy calculations of ligand–protein crystallographic complex: I. Knowledge-based ligand–protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity, Protein Eng. 8 (1995) 677–691.
[10] A. Wallqvist, R.L. Jernigan, D.G. Covell, A preference-based free energy parametrization of enzyme-inhibitor binding-applications to HIV-1 protease inhibitor design, Protein Sci. 4 (1995) 1881–1903.
[11] I. Muegge, Y.C. Martin, A general and fast coring function for protein–ligand interactions: a simplified potential approach, J. Med. Chem. 42 (1999) 791–804.
[12] J.B.O. Mitchell, R.A. Laskowski, A. Alex, J.M. Thornton, BLEEP-potential of mean force describing protein–ligand interactions: I. Generating potential, J. Comput. Chem. 20 (1999) 1165–1176.
[13] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring function to predict protein–ligand interactions, J. Mol. Biol. 295 (2000) 337–356.
[14] A.V. Ishchenko, E.I. Shakhnovich, SMall Molecule Growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein–ligand interactions, J. Med. Chem. 45 (2002) 2770–2780.
[15] H.J. Böhm, The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure, J. Comput.-Aided Mol. Des. 8 (1994) 243–256.
[16] D.M. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, Empirical scoring functions: I. The development of a fast empirical scoring function to estimate binding affinity of ligands in receptor complexes, J. Comput.-Aided Mol. Des. 11 (1997) 425–445.
[17] R.D. Head, M.L. Smythe, T.I. Oprea, C.L. Waller, S.M. Green, G.R. Marshall, VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands, J. Am. Chem. Soc. 118 (1996) 3959–3969.
[18] Z. Weng, C. Delisi, S. Vajda, Empirical free energy calculations: comparison to calorimetric data, Protein Sci. 6 (1997) 1976–1984.
[19] L.M. Amzel, Structure based drug design, Curr. Opin. Biotechnol. 6 (1998) 336–369.
[20] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[21] S. Tanaka, H.A. Scheraga, Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins, Macromolecules 9 (1976) 945–950.
[22] S. Miyazawa, R.L. Jernigan, Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation, Macromolecules 18 (1985) 534–552.
[23] M.J. Sippl, Calculation of conformational ensembles from potential of mean force. An approach to the knowledge-based prediction of local structures in globular proteins, J. Mol. Biol. 213 (1990) 859–883.
[24] A. Godzik, A. Kolinski, J. Skolnick, Topology fingerprint approach to the inverse protein folding problem, J. Mol. Biol. 227 (1992) 227–238.
[25] F. Melo, E. Feytmans, Novel knowledge-based mean force potential at atomic level, J. Mol. Biol. 267 (1997) 207–222.
[26] B.A. Reva, A.V. Finkelstein, M.F. Sanner, A.J. Olson, Residue–residue mean-force potentials for protein structure recognition, Protein Eng. 10 (1997) 865–876.
[27] C.W. Carter, B.C. LeFebvre, S.A. Cammer, A. Tropsha, M.H. Edgell, Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations, J. Mol. Biol. 311 (2001) 625–638.
[28] H.B. Bürgi, J.D. Dunitz, Can statistical analysis of structural parameters from different crystal environments lead to quantitative energy relationships? Acta Crystallogr., B 44 (1988) 445–448.
[29] G.P. Brady, A. Szabo, K.A. Sharp, On the decomposition of free energies, J. Mol. Biol. 263 (1996) 123–125.
[30] A.V. Finkelstein, J. Janin, The price of lost freedom: entropy of bimolecular complex formation, Protein Eng. 3 (1989) 1–3.
[31] L.M. Amzel, Loss of translational entropy in binding, folding, and catalysis, Proteins: Struct. Funct. Genet. 28 (1997) 144–149.
[32] M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, The statistical-thermodynamic basis for computation of binding affinities: a critical review, Biophys. J. 72 (1997) 1047–1069.
[33] C.W. Murray, M.L. Verdonk, The consequences of rotational and translational entropy lost by small molecules on binding to proteins, J. Comput.-Aided Mol. Des. 16 (2002) 741–753.
[34] B.M. Baker, K.P. Murphy, Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry, Biophys. J. 71 (1996) 2049–2055.
[35] J.M. Bradshaw, G. Waksman, Calorimetric investigation of proton linkage by monitoring both the enthalpy and association constant of binding: application to the interaction of the Src SH2 domain with a high-affinity tyrosyl phosphopeptide, Biochemistry 37 (1998) 15400–15407.
[36] M.M. Garcia-Mira, J.M. Sanchez-Ruiz, pH corrections and protein ionization in water/guanidinium-chloride, Biophys. J. 81 (2001) 3489–3502.

[37] J.A.D'Aquino, E. Freire, L.M. Amzel, Binding of small organic molecules to macromolecular targets: evaluation of conformational entropy changes, Proteins: Struct. Funct. Genet. 41 (2000) 93–107.
[38] C.A. Croxton, Liquid state physics. A Statistical Mechanical Introduction, Cambridge University Press, 1974, p. 42.
[39] W.A. Koppensteiner, M.J. Sippl, Knowledge-based potentials—back to the roots, Biochemistry (Russia) 63 (1998) 247–252.
[40] M. Gerstein, J. Tsai, M. Levitt, The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra, J. Mol. Biol. 249 (1995) 955–966.