Bioorganic & Medicinal Chemistry 16 (2008) 9378–9382

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Evaluation of ligand-binding affinity using polynomial empirical scoring functions

Walter Filgueira de Azevedo Jr. *, Raquel Dias

Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681, Bairro Partenon, Porto Alegre CEP 90619-900, Brazil

ARTICLE INFO

Article history:
Received 23 April 2008
Revised 29 July 2008
Accepted 4 August 2008
Available online 7 August 2008

Keywords:
Empirical scoring functions
Virtual screening
Protein–ligand
Docking

ABSTRACT

Assessing protein–ligand interaction is of great importance for virtual screening initiatives in order to discover new drugs. The present work describes a set of empirical scoring functions to assess the binding affinity, involving terms for intermolecular hydrogen bonds and contact surface. The results show that our methodology works better to predict protein–ligand affinity when compared with XSCORE, a popular empirical scoring function.

1. Introduction

Interactions involving protein and ligand are of pivotal importance in many fundamental biochemical processes, such as molecular recognition and enzymatic reactions.[1] One of the open problems in field of computational chemistry is the direct calculation of binding affinities, using solely information contained in the atomic coordinates of the protein–ligand complexes. Such approach is of extreme relevance for the process of scanning virtual libraries of small molecules, in order to identify new medications and biological probes. Indeed, the discovery of a new lead compound that binds tightly to a protein is the main goal of early-stage drug discovery and also of chemical genomics projects seeking inhibitors to elucidate gene function.[2]

The potential energy of a ligand or a binary complex involving protein and ligand in a given conformation can be calculated applying quantum mechanical methods. These ab initio quantum-chemistry techniques explicitly account for the electronic structure of the molecules present in the system. However, application of ab initio methods for calculation of binding affinities is time-consuming, even considering clusters of computers, grid computation, and supercomputers. Recent studies made some progress in this area, reducing the CPU time for evaluation of binding affinities. These studies demonstrated a qualitative relation between the electric characteristics and binding affinity of a protein–ligand complex; a high binding affinity correlates with a high charge transfer. This allows analysis of binding interactions of any binary complex, reducing computational time while maintaining reliability of the results. Nevertheless, these studies are not yet applied to virtual screening methods, and they are still restricted to simulations involving peptides.[3]

Computation of binding affinities can be addressed using empirical scoring methods, based on functions that use few terms such as intermolecular hydrogen bonds, van der Waals interaction, deformation effect, hydrophobic interaction, entropy, and others. This approach to the study of ligand-binding affinity started with the pioneering work of Böhm.[4] Many reviews summarize the recent progress.[5–8]

The majority of the empirical scoring functions in current use are based on the model, where binding affinity can be decomposed into terms that reflect the various contributions to the binding. This hypothesis can be used to build empirical scoring functions to estimate Gibbs free energy of binding ($\Delta G_{\text{binding}}$). This function is expressed as a sum of interactions multiplied by weighting coefficients (c_j), as follows:

$$\Delta G_{\text{binding}} = c_o + \sum_{j=1}^{N} c_j \cdot f_j(x, y, z) \quad (1)$$

where each term depends on the atomic coordinates of the ligand and protein, c_0 is a regression constant, f_j's are functions that account for van der Waals interactions, intermolecular hydrogen bonds, deformation, hydrophobic effect, and others that may be included. The summation is taken over all terms used to build the scoring function.

In the proposition of new empirical scoring functions it is necessary to use a training set of crystallographic structures of binary

* Corresponding author. Tel.: +55 51 33203500.
E-mail address: walter@azevedolab.net (W.F. de Azevedo).

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.08.014

complexes of proteins and ligands, where experimental information about the binding affinities is available. The empirical scoring function is built using a training set to obtain weights for each term present in the scoring function. Multivariate regression analysis is then applied to find the best fit between the predicted and experimental protein-binding affinities.

In spite of many problems in the understanding of the structural features important for binding affinity, most of the experimentally available data indicate that additive functions for protein-ligand interactions might be a good approach for the development of empirical scoring functions. These functions may be expressed in Gibbs free energy of binding, as in the Eq. 1, or in pK_d as used for X-SCORE[9] or in pK_i used in the LigScore.[10] With atomic coordinates (x,y,z) available for protein–ligand complexes, the analysis of the binding can be estimated as a sum of interactions multiplied by weighting coefficients (c_j), as indicated by the following equation:

$$pK_d = c_o + \sum_{j=1}^{N} c_j f_j(x,y,z) \quad (2)$$

where pK_d is the $-\log K_d$, where K_d could also be either K_M or K_i, c_0 is a regression constant, and f_j's are functions that account for intermolecular interactions.

Recently, we proposed the use of a set of polynomial functions to estimate the binding affinity for protein–ligand complexes.[11] Here, we describe in detail this algorithm and its application to estimate protein–ligand binding affinity. The results strongly indicate that the present algorithm shows better performance to estimate binding affinity when compared to other empirical scoring functions. Furthermore, the use of a set of 25 empirical scoring functions opens the possibility to test different schemes to predict binding affinities using an adaptive approach to assess binding affinity.

2. Results and discussion

We described here 25 polynomial empirical scoring functions which use only two terms to evaluate the binding affinity, a term for intermolecular hydrogen bonds (HB) and a second for contact surface (A). Table 1 shows these polynomial functions. These functions were implemented in a program called POLSCORE. Since identification of intermolecular hydrogen bond is of pivotal importance in our model to evaluate binding affinity, we compared the capability of POLSCORE to identify intermolecular hydrogen bonds with LIGPLOT[12], a well-established program used to identify hydrogen bonds. The correlation coefficient between the present algorithm and LIGPLOT obtained from the training set was 0.81, which indicates correlation between both methods. Since our results are similar to the results of an already accepted program, it assures the validation of the new algorithm for identification of hydrogen bonds.

Analysis of polynomial functions against the training set composed of 123 structures yielded squared correlation coefficient (r^2) ranging from 0.13 to 0.64, and standard deviation (s) ranging from 1.6 to 1.9 in pK_d units. The correlation between observed binding affinities and the fitted values given by the best polynomial function is shown in Figure 1. Nearly half of the set of polynomial functions present r^2 higher than 0.4, and three functions present values higher than 0.6 (function numbers 17, 18, and 21). All functions were kept in the program POLSCORE, even the ones with low values for r^2. Keeping all polynomial functions in the program allows for a wide range of empirical functions to be tested and then applied to a specific set of proteins or structural related compounds.

POLSCORE uses the three polynomial empirical scoring functions 17, 18, and 21. It calculates the average value for these three functions. The true value of a novel empirical scoring function lies in its predictive capability. We used the program POLSCORE to predict the binding affinity for 20 protein–ligand complexes in the test set and compared with XSCORE. Polynomial functions 6, 7, 19, and 21 showed the highest correlation coefficient against the experimental binding affinities (0.65), higher than the one obtained using the XSCORE, correlation coefficient of 0.23. The results are shown in Table 2.

The program POLSCORE has also the capability to test the 25 polynomial functions against a set of protein–ligand complexes. This feature is relevant when one desires to propose an empirical scoring function specific for a protein family or specific for a family

Table 1
Polynomial functions together with weights obtained using regression analysis

	Polynomials	c_0	c_1	c_2	c_3	c_4	c_5
1	$c_0 + c_1.x + c_2.y$	4.245	0.05105	5.387×10^{-3}	0	0	0
2	$c_0 + c_1.x + c_2.y^2$	5.114	0.07455	3.597×10^{-6}	0	0	0
3	$c_0 + c_1.x + c_2.x^2 + c_3.y$	3.537	0.2794	−0.01520	5.846×10^{-3}	0	0
4	$c_0 + c_1.x + c_2.x^2 + c_3.y^2$	4.569	0.2668	−0.01309	4.148×10^{-6}	0	0
5	$c_0 + c_1.x + c_2.x^2 + c_3.y + c_4.y^2$	2.270	0.2244	-6.920×10^{-3}	0.01571	-12.03×10^{-6}	0
6	$c_0 + c_1.x + c_2.y + c_3.y^2$	2.518	0.1235	0.01596	-12.57×10^{-6}	0	0
7	$c_0 + c_1.x^2 + c_2.y + c_3.y^2$	2.959	5.687×10^{-3}	0.01593	-12.48×10^{-6}	0	0
8	$c_0 + c_1.x^2 + c_2.y$	4.462	260.6×10^{-6}	5.648×10^{-3}	0	0	0
9	$c_0 + c_1.x^2 + c_2.y^2$	5.427	1.841×10^{-3}	3.872×10^{-6}	0	0	0
10	$c_0 + c_1.x + c_2.x.y$	5.386	0.03340	219.8×10^{-6}	0	0	0
11	$c_0 + c_1.x.y + c_2.y$	4.219	8.225×10^{-3}	-233.5×10^{-6}	0	0	0
12	$c_0 + c_1.x + c_2.x^2 + c_3.x.y$	4.676	0.2906	−0.02157	366.5×10^{-6}	0	0
13	$c_0 + c_1.x + c_2.x.y + c_3.y$	2.282	0.3243	0.01177	-728.5×10^{-6}	0	0
14	$c_0 + c_1.x + c_2.x^2 + c_3.x.y + c_4.y$	2.410	0.1399	0.01951	0.01373	-1.018×10^{-3}	0
15	$c_0 + c_1.x^2 + c_2.x.y + c_3.y$	2.781	0.02880	-1.077×10^{-3}	0.01409	0	0
16	$c_0 + c_1.x^2 + c_2.x.y$	5.598	-4.005×10^{-3}	318.8×10^{-6}	0	0	0
17	$c_0 + c_1.x + c_2.x^2 + c_3.x.y + c_4.y + c_5.y^2$	2.005	0.1560	0.01198	-639.5×10^{-6}	0.01721	-7.825×10^{-6}
18	$c_0 + c_1.x + c_2.x.y + c_3.y + c_4.y^2$	1.885	0.2627	-429.9×10^{-6}	0.01652	-8.751×10^{-6}	0
19	$c_0 + c_1.x.y + c_2.y + c_3.y^2$	3.225	32.18×10^{-6}	0.01526	-11.36×10^{-6}	0	0
20	$c_0 + c_1.x^2 + c_2.x.y + c_3.y^2$	5.333	7.316×10^{-3}	-238.0×10^{-6}	6.035×10^{-6}	0	0
21	$c_0 + c_1.x^2 + c_2.x.y + c_3.y + c_4.y^2$	2.428	0.02251	-714.3×10^{-6}	0.01753	-7.630×10^{-6}	0
22	$c_0 + c_1.x.y + c_2.y^2$	5.483	5.935×10^{-6}	4.076×10^{-6}	0	0	0
23	$c_0 + c_1.x.y$	5.528	249.4×10^{-6}	0	0	0	0
24	$c_0 + c_1.x + c_2.x.y + c_3.y^2$	4.748	0.1671	-289.4×10^{-6}	6.547×10^{-6}	0	0
25	$c_0 + c_1.x + c_2.x^2 + c_3.x.y + c_4.y^2$	4.560	0.25365	-9.358×10^{-3}	-130.3×10^{-6}	5.319×10^{-6}	0

x = number of intermolecular hydrogen bonds (HB). $y = (A_1)^2/(A_2 - A_1)$, where A_1 is contact area between protein and ligand, A_2 is the total area of the ligand, determined using the program AREAIMOL.[20]

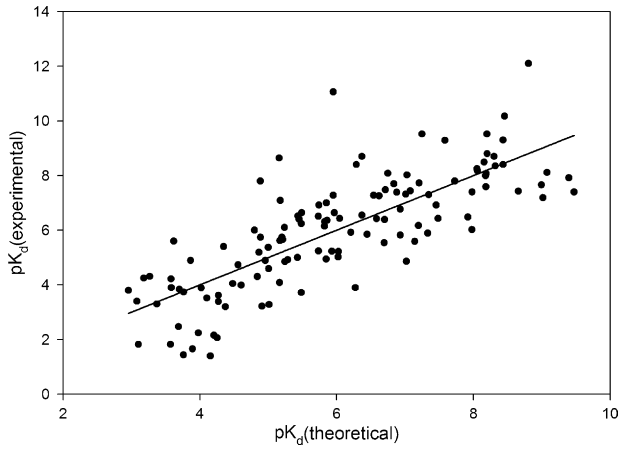

Figure 1. Dispersion plot for experimental and theoretical pK_d functions (function 21).

Table 2
Binding affinities

PDB access code	Experimental pK_d	pK_d (POLSCORE)	pK_d (X-SCORE)
10GS	6.40	7.29	6.47
1AI5	3.72	4.72	5.31
1B9J	5.96	7.13	7.05
1BMA	4.59	6.48	7.07
1HI4	4.49	5.71	5.36
1E5A	7.64	5.55	5.21
1JQE	6.44	10.1	7.12
1KV1	5.94	8.05	7.01
1OLS	5.82	8.19	6.86
1PPM	5.80	8.28	7.60
1Q8T	4.76	6.45	5.81
1SV3	4.74	6.26	4.81
1TSY	4.96	6.73	5.48
1U33	4.60	7.26	6.22
1PWY	4.04	5.10	6.21
1V2H	4.54	4.73	5.23
1RFG	4.85	4.97	6.82
1RCT	4.39	5.04	5.77
1YRY	3.45	4.72	5.95
1V45	4.42	3.98	6.61

of compounds for which structural information and experimentally determined binding affinities are available. It has been demonstrated that dividing up a set of known inhibitors into clusters of chemically similar molecules, and then deriving a specific scoring function for each molecule sets increase the accuracy of the scoring function.[1,11,13] We illustrated this application on the study of a new empirical function for complexes involving human PNP.[11] In this study, we have used 15 binary complexes involving human PNP and different ligands as a test set. We also used the program XSCORE to evaluate the binding affinity of the ligands for human PNP. A correlation coefficient of 0.65 and a standard deviation of 0.79 in pK_d units were obtained using POLSCORE, these results are better than the ones obtained using XSCORE empirical scoring function, 0.32 and 0.92, respectively. Furthermore, our novel empirical scoring functions were able to identify the ligand with highest affinity for human PNP, the inhibitor 5-deoxy-5-iodo-9-deazainosine, described in a previous study of our laboratory.[11] This capability is of pivotal importance when scanning larger libraries with thousands or even millions of compounds. Such initiatives intend to identify the ligands with high affinity present in the library.

Empirical scoring functions that decompose the binding free energy into a sum of terms, present an intrinsic problem in physical sense, since this decomposition is not allowed. Free energy of binding is a state function but its terms are not.[14] Furthermore, additive methods are unable to describe subtle cooperative effects.[15] Despite these problems, simulation of biomolecular systems in silico has great potential to make predictions and give insights that can guide molecular design. The present set of empirical scoring functions partially overcomes the problems cited above, introducing cross-terms involving intermolecular hydrogen bonds (HB) and contact surface (A).

3. Conclusions

One of the weaknesses of empirical scoring methods is the low reliability for unknown systems. The present set of functions was tested on systems not used in the training set. The use of simple scoring functions to estimate the pK_d for binary complexes showed good agreement with experimentally determined affinities for two different test sets. The scoring capability of our novel scoring functions identified 5-deoxy-5-iodo-9-deazainosine as the ligand with the highest affinity, as previously reported in the study of human PNP.[11] The same scoring functions were used to estimate the binding affinity of the ligands for 20 complexed structures and showed predictive capability higher than XSCORE. Strictly speaking, neither traditional scoring function nor the cross-term function (described here) present direct physical basis. Nevertheless, using the cross-term empirical scoring function was able to predict the binding affinity for complexed structures, with results better than previously described empirical scoring function.

4. Methods

4.1. Evaluation of binding affinity

Most of the empirical functions use the size of the contact surface at the protein–ligand interface to estimate the hydrophobic interaction. Several programs adopt this methodology to estimate the hydrophobic interaction.[16] It has been noted that a reasonable correlation between experimental affinities can be obtained with contact surfaces alone. Furthermore, it is well established that intermolecular hydrogen bonds are of pivotal importance for the protein–ligand interaction. We described here 25 scoring functions which are based on the observation that the major determinants to ligand specificity are intermolecular hydrogen bonds and hydrophobic contacts.[17–19] Table 1 shows all 25 polynomial empirical scoring functions.

Determination of the intermolecular hydrogen bonds and contact areas of 123 binary complexes available (training set) were used to propose these functions (Supplementary material). These novel functions are based exclusively on the hydrogen bonds (HB), contact area (A_1), and accessible surface area for the ligand (A_2).

The additive nature of the empirical scoring functions generally leads to large ligands, obtaining high scores. This effect is undesirable when trying to estimate the binding affinity. Trying to avoid the overestimation of the contact surface term contribution to the empirical scoring function, we devise a simple scheme to reduce this contribution. We introduced a penalty term that diminishes the dependence of the score on molecular size. We divided the squared contact area ($A_1 * A_1$) by the term $A_2 - A_1$, which presents the highest values for complexes with A_1 much smaller than A_2. This scheme reduces the contribution of the contact surface term in the empirical function for those ligands with relatively small contact area (A_1) when compared with the total area (A_2). The empirical functions are composed of a polynomial function

involving the number of intermolecular hydrogen bonds (HB) and the modified contact surface (A) as follows.

$$pK_d = c_o + \sum_{j=1}^{N} c_j f_j(x,y,z) \cdot g_j(x,y,z) \quad (3)$$

where f_j and g_j are terms for intermolecular hydrogen bonds (HB) and/or modified contact surface (A), this summation may involve quadratic terms of HB and/or A and/or cross-terms (HB.A). Our set of functions comprises polynomials up to degree 2, c_0 is the regression constant, and c_j is the weighting coefficient for the terms $f_j \cdot g_j$.

The modified contact surface is determined by the following equation,

$$A = \frac{AI^2}{(A_2 - A_1)} \quad (4)$$

The terms A_1 and A_2 are calculated using the program AREAIMOL in the CCP4.[20]

4.2. Hydrogen bonds

Specificity of ligand for a protein depends on several features such as intermolecular hydrogen bonds, van der Waals contacts, and shape and charge complementarity between ligand and protein.[21–23] Therefore, assessing intermolecular hydrogen bonds from crystallographic data is of pivotal importance for determining specificity of ligand against a protein. A hydrogen bond is an interaction involving two electronegative atoms, a donor and an acceptor.[24–26] A hydrogen atom lies aligned between them and is covalently bound to the donor atom. In this electrostatic interaction, the donor attracts the electron on the hydrogen from its orbital towards the donor itself. This results in a partial positive charge on the hydrogen, which is electrostatically attracted towards the electronegative acceptor.

Our algorithm does not use a fuzzy function as described by Wang et al.[9] Instead we use two functions that evaluate the distance ($f(d)$) and the angles ($f(\theta)$) for each pair of atoms.

The present algorithm assumes a set of ideal parameters for each pair of atoms present in the binary complexes. Intermolecular hydrogen bonds (D–H···A) are usually described by the distance between the hydrogen (H) and acceptor atom (A), but most of the time hydrogen atoms are not revealed in X-ray crystallography analysis. Crystallographic structures refined to resolution 1.2 Å or higher are suitable for identification of hydrogen atom positions, however, most of the structures solved by X-ray crystallography usually are good enough to locate non-hydrogen atoms only. Nevertheless, hydrogen atoms can be added and set with energy minimization method, but there are many possible low-energy positions.

The present algorithm does not use information about the positions of hydrogen atoms, it uses the following criteria to verify whether a pair of atoms, one from the ligand and other from the protein, participates in intermolecular hydrogen bond:

(i) Criteria for hydrogen atoms. No hydrogen atoms are considered in the identification of intermolecular hydrogen bonds. Intermolecular hydrogen bonds were predicted using information about donor/acceptor pairs. Furthermore, since the angle is also an important feature in the hydrogen bond formation it has been included in the algorithm.

(ii) Donor and acceptor. The donor (D) atoms are the closest hydrogen neighbor atoms and the acceptor atoms (A) are those that receive the hydrogen bond coming from a donor atom. Donor root (DR) and acceptor root (AR) are defined as the closest donor's non-hydrogen neighbor atoms. Figure 2 illustrates a diagram showing the connection among all atoms involved in the determination of intermolecular hydrogen bonds. In the case where there is more than one donor/acceptor root, the program will take the geometric center position obtained from all the donor root atoms positions.

The algorithm calculates two functions $f(d)$ and $f(\theta)$, these functions return 1 if a pair of atoms satisfies the distance and angle conditions, respectively. Therefore, the number of intermolecular hydrogen bonds (HB) is given by the following equation:

$$HB = \sum_{j=1}^{N} f_j(d) \cdot f_j(\theta) \quad (5)$$

where the summation is taken over all pairs of atoms, one atom being from the ligand and the second atom from the protein.

To validate the capability of the present algorithm to identify intermolecular hydrogen bonds, we tested it against the previously described training set (Supplementary material). All atomic coordinate files in this dataset involve crystallographic structures of binary complexes. The files were analyzed with present algorithm and also with LIGPLOT,[12] a program used to determine intermolecular hydrogen bonds. The LIGPLOT program automatically generates schematic 2D representations of protein-ligand complexes from standard Protein Data Bank file input.

4.3. Regression analysis

The relative weight of the individual contributions for the hydrogen bond (HB) and modified contact surface (A) terms depends on the training set. We used a training set composed of the 123 crystallographic structures. All structures present non-covalent bound ligand in the binary complexes and have the experimental affinity available. Standard multivariate regression was carried out on the whole training set. The weighting coefficients obtained for each polynomial function are listed in Table 1.

4.4. Test set

To validate our set of empirical scoring functions, we used 20 binary complexes that were deliberately separated from the training set. Our 25 polynomial functions were used to estimate the pK_d and were compared with the program XSCORE. Previously published comparison studies involving 14 empirical scoring functions indicated that XSCORE[16,27] was able to obtain the best results on evaluation of binding affinities, therefore this program was chosen as comparison for this new set of empirical score functions. A complete list of the test set as well as the pK_d's estimated using all the polynomial functions can be found in the Supplementary material of this paper. Table 2 shows the pK_d's estimated using XSCORE and polynomial function number 21.

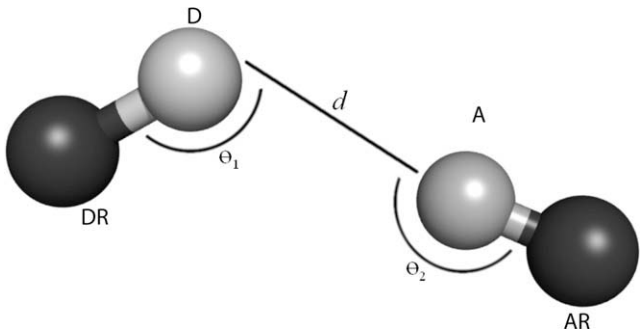

Figure 2. Diagram of intermolecular hydrogen bonds.

4.5. Program description

The algorithm used to evaluate the intermolecular hydrogen bonds and the pK_d's based on the 25 empirical scoring functions described in Table 1 is implemented in the program POLSCORE. The inputs to the program are the PDB file for the protein, a MOL2 file for the ligand, and the areas A_1 and A_2, calculated using AREAIMOL.[20] The user may also select to test all polynomial functions available in POLSCORE_LIST, in order to select the best empirical scoring function for a given dataset. For this option, information about experimental affinities must be provided (as pK_d). The results are output into a text file in which the detailed information about the pK_d evaluated using the 25 polynomial functions is tabulated. The program is written in ANSI C++ and runs on Linux- and Windows-based computers. It is available for download at http://bioinfo.zip.net/downloads.html.

Acknowledgments

This work was supported by grants from CNPq and Instituto do Milenio (CNPq-MCT). R.D. thank CNPq for the fellowship. W.F.A. is senior researcher for CNPq (Brazil).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.08.014.

References and notes

1. Böhm, H.-J.; Schneider, G. *Protein–Ligand Interactions*; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2003.
2. Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.
3. Salazar, P. F.; Seminário, J. *Phys. Chem. B* **2008**, *112*, 1290.
4. Böhm, H.-J. J. *Comput.-Aided Mol. Des.* **1994**, *8*, 243.
5. Rajamani, R.; Good, A. C. *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 308.
6. Laurie, A. T.; Jackson, R. M. *Curr. Protein Pept. Sci.* **2006**, *7*, 395.
7. Vajda, S.; Guarnieri, F. *Curr. Opin. Drug Discov. Dev.* **2006**, *9*, 354.
8. Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. *Curr. Med. Chem.* **2004**, *11*, 3093.
9. Wang, R.; Lai, L.; Wang, S. J. *Comput.-Aided Mol. Des.* **2002**, *16*, 11.
10. Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. *J. Mol. Graph. Model.* **2005**, *23*, 395.
11. Caceres, R. A.; Timmers, L. F. M. S.; Dias, R.; Basso, L. A.; Santos, D. S.; de Azevedo, W. F. *Bioorg. Med. Chem.* **2008**, *9*, 4984.
12. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 127.
13. Rizzo, R. C.; Udier-Blagovic, M.; Wang, D. P.; Watkins, E. K.; Kroeger Smith, M. B.; Smith, R. H.; Tirado-Rives, J.; Jorgensen, W. L. J. *Med. Chem.* **2002**, *45*, 2970.
14. Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.
15. Williams, D. H.; Bardsley, B. *Perspect. Drug Discov. Design.* **1999**, *17*, 43.
16. Wang, R.; Lu, Y.; Fang, X.; Wang, S. J. *Chem. Inf. Comput. Sci.* **2004**, *44*, 2114.
17. Boehm, H.-J.; Klebe, G. *Angew. Chem. Int. Ed.* **1996**, *35*, 2589.
18. Matsumara, M.; Matthews, B. W.; Beckel, W. J. *Nature* **1988**, *334*, 406.
19. Nauchatel, V.; Villaverde, M. C.; Sussman, F. *Protein Sci.* **1995**, *4*, 1356.
20. Collaborative Computational Project, Number 4. *Acta Cryst.* **1994**, D50, 760.
21. De Azevedo, W. F.; Mueller-Dieckmann, H. J.; Schulze-Gahmen, U.; Worland, P. J.; Sausville, E.; Kim, S-H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2735.
22. De Azevedo, W. F.; Leclerc, S. J.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S. H. *Eur. J. Biochem.* **1997**, *243*, 518.
23. De Azevedo, W. F.; Canduri, F.; Fadel, V.; Teodoro, L. G.; Hial, V.; Gomes, R. A. *Biochem. Biophys. Res. Comm.* **2001**, *287*, 277.
24. Latimer, W. M.; Rodebush, W. H. *J. Am. Chem. Soc.* **1920**, *2*, 1419.
25. Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford University, 1997.
26. Huggins, M. L. *Angew. Chem. Int. Ed.* **1971**, *10*, 147.
27. Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287.