ISSN 0006-2979, Biochemistry (Moscow), 2008, Vol. 73, No. 2, pp. 231-233. © Pleiades Publishing, Ltd., 2008.
Published in Russian in Biokhimiya, 2008, Vol. 73, No. 2, pp. 283-286.

Interfacial Atom Pair Analysis

Yong-Chao Li and Zong-Hao Zeng*

Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China; E-mail: zzh@ibp.ac.cn

Received June 25, 2007
Revision received July 10, 2007

Abstract—The relations of the binding free energies in a dataset of 69 protein complexes with the numbers of interfacial atom pairs, as well as with the atomic distances of the pairs, are analyzed. It is found that the interfacial main-chain atom pairs contribute more to the correlation than the interfacial side chain atom pairs do, and the polar atom pairs contribute more than the non-polar atom pairs do. Interfacial atom pairs with atomic distance in the range of 6-12 Å are the most important to explain the differences in binding free energies in the datasets.

DOI: 10.1134/S0006297908020156

Key words: interfacial atom pairs, statistical potentials, complementarity

Works on "statistical potentials", or "potentials of mean force", are impressive [1-12]. The potentials are statistical results from databases of monomer structures [1] or databases of protein complexes [3]. Uniform potentials were also published, which can be used for both folding and binding [1, 6, 7]. The potentials have been tested not only by theoretical works, but also by fitting with experimental binding free energies [1-4, 6-10]. Different groups used different definitions for the "pair potentials", $G_{ij}(r)$, of two atoms of type i and j, which are functions of the atomic distance r. The atoms may be classified into several, or dozens of, types [1-3, 12]. If we take the differently defined pair potentials as empirical, for all these methods the common approximation to the free energies of folding, or binding, is to establish a linear relation between the free energies ΔG and the pair potentials by linear fitting, $\Delta G = \Sigma G_{ij}(r)$, where the summation runs over all types of atom pairs and all distances less than a cut-off distance r_{cut}.

There may exist quantities beneath these "statistical potentials", or "potentials of mean force", which are important to correlate with the free energies of folding or binding. We selected the dataset used in reference [2], which has the largest number of protein–protein and protein–peptide complexes, and analyzed the relation between the binding free energies with the numbers of interfacial atom pairs. In this letter, we report the results.

* To whom correspondence should be addressed.

METHODS OF INVESTIGATION

The 69 protein complexes were kept as they are in reference [2], except the experimental binding free energy of the complex with PDB ID 1efn was changed to −8.8 kcal/mol in accordance with the experimental k_d value [13]. Atoms are divided into four types: main-chain polar, main-chain non-polar, side chain polar, and side chain non-polar. In an interfacial atom pair, the two atoms are each in one monomer of a complex. The atomic distances of interfacial atom pairs, labeled by the types of the two atoms are calculated and recorded if their values are in the range 2.0-16.0 Å. Not full occupations of atomic coordinates as determined by the X-ray structures are considered as to produce partial pairs. Non-standard amino acids are excluded from the statistics.

RESULTS

The correlation coefficients of binding free energies with various types of atom pair numbers summed over the distance interval of 2.0-9.5 Å are listed in the table. The results show that main chain atom pairs, which give correlation coefficients about 0.72-0.73 (rows 1-3 in the table), are more important than that of the side chains, which give correlation coefficients less than 0.62 (rows 4-6 in the table). Whereas the side chain atom pairs alone are unlikely to be correlated with the binding free ener-

Correlation coefficients between binding free energies and various types of atom pair numbers summed up in the distance interval of 2.0-9.5 Å

Atom pair type		Correlation coefficient
partner I	partner II	
Main chain polar	main chain polar	0.72
Main chain non-polar	main chain non-polar	0.73
Main chain all atoms	main chain all atoms	0.72
Side chain polar	side chain polar	0.45
Side chain non-polar	side chain non-polar	0.59
Side chain all atoms	side chain all atoms	0.62
All polar atoms	all polar atoms	0.78
All non-polar atoms	all non-polar atoms	0.72
All atom types	all atom types	0.77

gies, their inclusions do improve the correlation (rows 7-9 in the table).

Higher correlations are achieved if polar atoms are involved (rows 7 and 9 in the table). If all atoms, or only all the polar atoms, are used, correlation coefficients about 0.77-0.78 can be achieved. It is especially interesting that the constant term, –4.7 kcal/mol, in linear fitting is almost the same as that in reference [2].

To explore the effects of atomic distances of atom pairs, pair numbers of all atom types are summed in a distance interval with a fixed 1.2 Å width, while the central distance of the interval moves from 2.6 to 15.4 Å. Figure 1 displays the relation of correlation coefficients with the central distance of the intervals for summing up interfacial pair numbers. In Fig. 1b, atom pairs are taken into account only if the two atoms are in interface residues [2]. Interface residues are defined as those residues that have at least one atom in a distance less than 4.5 Å to the partner peptide chain in a complex. In Fig. 1a, no such a restriction was applied. From both of the figures, one can see that pairs with distance more than about 4 Å give major contributions to the binding free energies. Correlation coefficients approach or surpass 0.7 if the fixed width intervals are in the distance range of 4-15 Å.

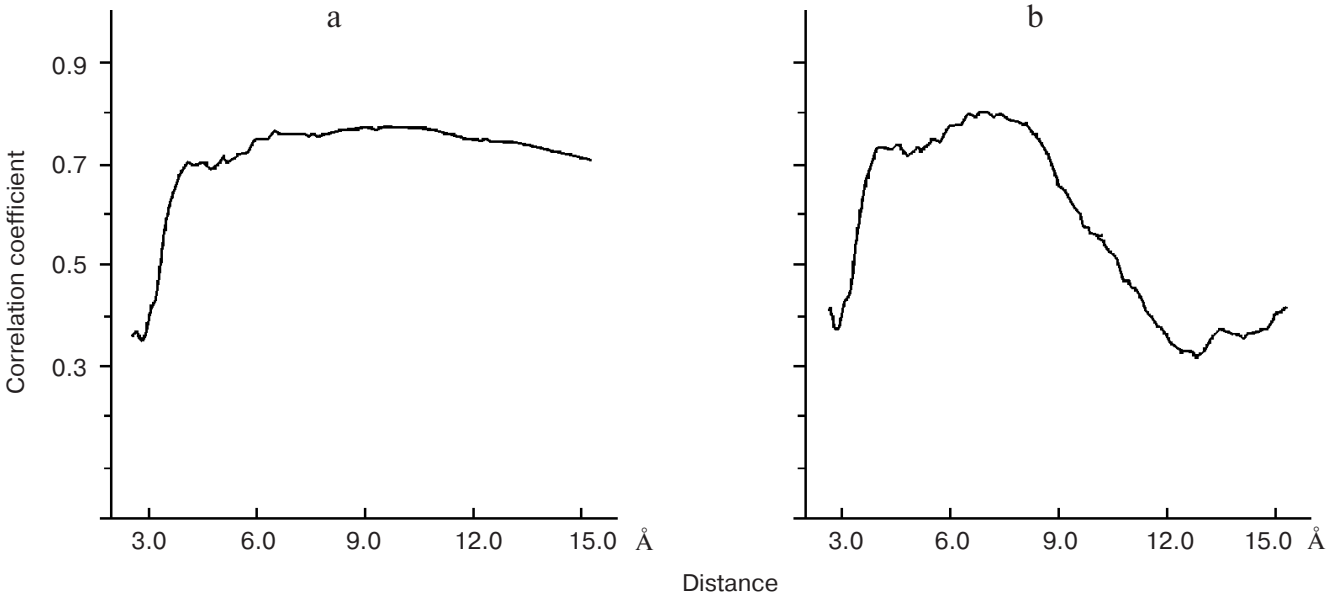

Fig. 1. Relation of correlation coefficients with the central distance of the fixed 1.2 Å width intervals for summing up atom pairs between all atoms: b) summations are over interface residues only; a) no restrictions.

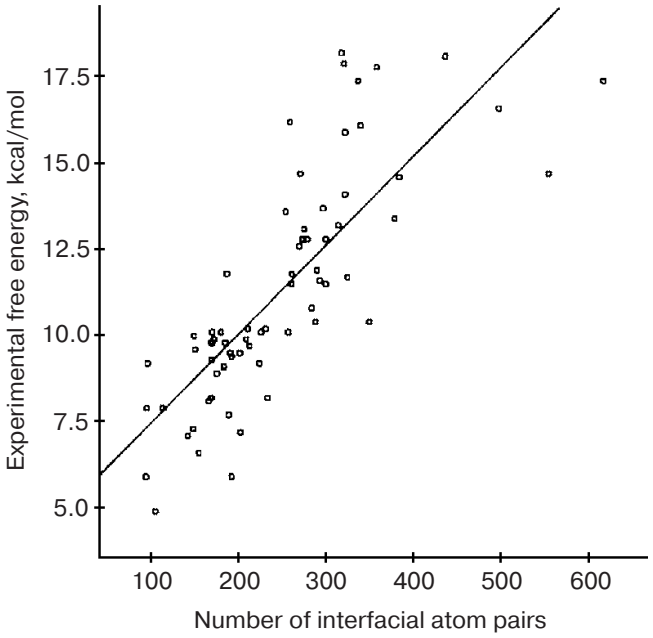

Fig. 2. Fitting of interfacial atom pairs within 7.2-8.4 Å with experimental affinities.

If the intervals are within 6-12 Å, the correlation coefficients will be more than 0.75 (Fig. 1a). Atom pairs in 4-9 Å are the most important if summations are over interface residues (Fig. 1b). This phenomenon may be looked upon as a "long range" effect.

DISCUSSION

Here the analysis suggests that an approximately linear relation exists between binding free energies and total numbers of interfacial atom pairs. As shown in Fig. 2, fitting of interfacial atom pairs in 7.2-8.4 Å with experimental affinities results in a correlation coefficient of almost 0.80. The importance of the buried molecular surface in evaluating the binding has been pointed out before [14-16]. As quantities to evaluate the binding of two peptides, the total number of interfacial atom pairs is better than the buried molecular surface, since the former takes into account both the extensiveness and complementarity of an interface. The ability of polarization is also important in binding, as the correlation is improved by inclusion of polar atoms. This phenomenon has been reported in the literature [15-19].

The statistics on structures of a set of proteins can be looked upon as a statistics on the instantaneous structures of a liquid drop of an averaged size. To explain the "long range" effect in binding, we suggest that each protein is a solid body covered by a thin layer of soft materials. In this model, the "long range" effect implies that the extensiveness and complementarity of the two contacting solid bodies, not the soft materials at the interface, is the major factor to determine the binding free energy.

REFERENCES

1. Zhou, H., and Zhou, Y. Q. (2002) *Protein Sci.*, **11**, 2714-2726.
2. Liu, S., Zhang, C., Zhou, H. Y., and Zhou, Y. Q. (2004) *Proteins: Structure*, *Function*, *and Bioinformatics*, **56**, 93-101.
3. Jiang, L., Gao, Y., Mao, F., Liu, Z., and Lai, L. (2002) *Proteins*, **46**, 190-196.
4. Sippl, M. J. (1990) *J. Mol. Biol.*, **213**, 859-883.
5. Marsden, P. M., Puvanendrampillai, D., Mitchell, J. B. O., and Glen, R. C. (2004) *Org. Biomol. Chem*., **22**, 3267-3273.
6. Mitchell, J. B. O., Roman, A. L., Alexander, A., and Janet, M. T. (1999) *J. Comput. Chem.*, **20**, 1165-1176.
7. Mitchell, J. B. O., Roman, A. L., Alexander, A., Mark, J. F., and Janet, M. T. (1999) *J. Comput. Chem.*, **20**, 1177-1185.
8. Sippl, M. J. (1995) *Curr. Opin. Struct. Biol.*, **5**, 229-235.
9. Moont, G., Gabb, H. A., and Sternberg, M. J. E. (1999) *Proteins: Structure*, *Function*, *and Genetics*, **35**, 364-373.
10. Robert, C. H., and Janin, J. (1998) *J. Mol. Biol.*, **283**, 1037-1047.
11. Thomas, P. D., and Dill, K. A. (1996) *J. Mol. Biol.*, **257**, 457-469.
12. Samudrala, R., and Moult, J. (1998) *J. Mol. Biol.*, **275**, 895-916.
13. Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996) *Cell*, **85**, 931-942.
14. Chothia, C., and Janin, J. (1975) *Nature*, **256**, 705-708.
15. Xu, D., Lin, S. L., and Nussinov, R. (1997) *J. Mol. Biol.*, **265**, 68-84.
16. Ma, X. H., Wang, C. X., Li, C. H., and Chen, W. Z. (2002) *Protein Eng.*, **15**, 677-681.
17. Matthew, J. B. (1985) *Annu. Rev. Biophys. Biophys. Chem.*, **14**, 387-417.
18. Vajda, S., Weng, Z. P., Rosenfld, R., and DelLisi, C. (1994) *Biochemistry*, **33**, 13977-13988.
19. Zhang, C., Vasmatzis, G., Cornette, J. L., and DeLisi, C. (1997) *J. Mol. Biol.*, **267**, 707-726.