A binary logistic regression model for discriminating real protein-protein interface*

LIN Wei, SUN Ping and LIU Xiangjun**

(Bioinformatics Laboratory, School of Medicine, Tsinghua University, Beijing 100084, China)

Received November 18, 2002; revised January 20, 2003

Abstract The selection and study of descriptive variables of protein-protein complex interface is a major question that many biologists come across when the research of protein-protein recognition is concerned. Several variables have been proposed to understand the structural or energetic features of complex interfaces. Here a systematic study of some of these "traditional" variables, as well as a few new ones, is introduced. With the values of these variables extracted from 42 PDB samples with real or false complex interfaces, a binary logistic regression analysis is performed, which results in an effective empirical model for the evaluation of binding probabilities of protein-protein interfaces. The model is validated with 12 samples, and satisfactory results are obtained for both the training and validation sets. Meanwhile, three potential dimeric interfaces of staphylokinase have been investigated and one with the best suitability to our model is proposed.

Keywords: protein-protein recognition, binary logistic regression, interfacial descriptive variable.

Non-covalent protein-protein recognition is essential to biological functions. Biologists are interested in how protein-protein recognition is fulfilled and what features are common among protein-protein complex interfaces. Different types of complexes, including subunit-subunit interfaces, protease-inhibitor bindings, growth hormone receptor complex, and antibody-antigen interactions, etc. have been studied[1~7]. Duquerroy, Cherfils and Janin studied known protein-protein interfaces and found that non-native alternative predicted dockings often had as many hydrogen bonds as interfacial areas[8]. Lawrence, and Colman developed methods to study shape complementarity at protein-protein interfaces and showed that antibody-antigen interfaces are generally less well-packed than other protein-protein interfaces, etc.[9]. Manocha and Wright computed interfacial surfaces that pass between two protein molecules[10]. Keskins proposed solvent inter-residue contact potential for protein-protein recognition[11]. In summary, structural and computational biologists have employed as many types of interfacial variables as they could find to describe the geometric or energetic features of the complex binding sites at different levels in these papers. Generally, it was well recognized that two types of complementarity ought to be satisfied: geometric complementarity and energetic complementarity. So far, not a single variable mentioned above can satisfactorily fulfill the determination of interface formation.

It is also important to distinguish the interface variables from surface ones. For example, Jones and Thornton tried to reveal the difference between solvent-accessible surface of protein and buried interface with 6 variables[12]. Their variables only characterize the properties of the individual surface patch of each molecule, but not those of the interaction interface. Neither did they give a quantified prediction model.

In our work, we studied many variables which were previously used or newly defined to analyze the interfacial features. We proposed an empirical model by binary logistic regression analysis. This work is based on an assumption that there is a dependent variable which has separate distributions between real and false interfaces. Such a dependent variable was derived from many interfacial variables.

1 Materials and methods

Twenty-five protein complex structures with interfaces of biological function and 17 monomer proteins with "pseudo-interfaces" of crystallographic packing were extracted from Brookhaven protein databank (PDB). The 25 complex structures belong to different classifications defined by Lo Conte et

* Supported by Trans-Century Training Program Foundation for the Talents by the Ministry of Education of China

** To whom correspondence should be addressed. E-mail: frankliu@tsinghua.edu.cn al.[13], and have different biological functions and structural details. Meanwhile, 17 monomer proteins without real complex interfaces but with crystallographic interfaces presented in their PDB structures were also employed. The PDB IDs of real complex samples and monomer samples are listed in Table 1.

Table 1. Training samples for statistical study

Protein-protein complexes with real interfaces	Monomers with "pseudo" interfaces
Protease-inhibitor (5 items):	1a8v 1a9m 1amu 1aoe 1ayf 1ba2 1bb3 1bja 1bry 1cgf
1avw 1cho 1cse 1mkw 2ptc	1crc 1dek 1dy5 1dzk 1e2w 1e4y 1eox 1fmt 1fvk 1g8i
Large protease complexes (4 items):	1h97 1iaz 1ilr 1lys 1mdv 1qin 1sei 3tmy
1bth 1tbq 1toc 4htc	
Antibody-antigen (3 items):	
1jhl 1nca 1osp	
Enzyme complexes (5 items):	
1brs 1dhk 1fss 1gla 1ydr	
G protein and signal transduction (4 items):	
1agr 1fin 1tx4 2trc	
Miscellaneous (4 items):	
1a71 1fc2 1igc 1ycs	

The extracted structural and energetic descriptors include:

(i) The area of buried interface "A_b". The accessible surface area of the complex and the components are calculated by GETAREA 1.1, a web-based program provided by Sealy Center for Structural Biology[14]. The area of buried interface, A_b, is approximately calculated by the equation as follows:

$$A_b = (A_{\text{compo1}} + A_{\text{compo2}} - A_{\text{compl}})/2, \quad (1)$$

where A_{compo1} represents the solvent accessible area of component 1 before binding; A_{compo2}, the area of component 2; A_{compl} the solvent accessible area of complex after binding; A_b is often used in the description of a complex interface.

(ii) The number of atoms buried in the interfaces "N_b". The interfacial atoms are extracted if one atom is in contact with another atom of different component with their distance less than the sum of their Van Der Vaals radii and a solvent diameter. The total number of atoms buried in the interfaces, N_b, is then summed.

(iii) The number of hydrogen bonds between complex components "N_h". N_h is obtained from the hydrogen bonds involved in the interfacial binding. The hydrogen bonds were picked when the hydrogen donor and the acceptor contact each other with the distance less than the hydrogen bond length. Structural biologists concern about hydrogen bond when they observe or propose a complex interface.

(iv) The score of charged residue pairs buried in the interface "S_{ch}". S_{ch} is calculated from the number of charged residue pairs meeting on the interfaces. The equation is:

$$S_{ch} = N_{\text{oppo}} - N_{\text{iden}}, \quad (2)$$

where N_{oppo} represents the number of opposite charged residue pairs buried in the interface; N_{iden} the number of identical charged residue pairs. The higher the S_{ch}, the stronger the two components bind together. We found that S_{ch} is a rough but effective variable.

(v) The area ratio of buried interfaces to the total area of a complex surface "R_a".

(vi) The ratio of the buried atom number to the total complex atoms number "R_n".

(vii) The ratio of hydrogen bonds between complex components "$R_{n,h}$".

(viii) The score ratio of charged residue pairs between complex components "$R_{s,ch}$".

(ix) A scoring function by interfacial residue-pair contact potential "S_{cp}". S_{cp} is an energetic scoring function, derived from pairwise contact potential which extracted from interface samples at residue level[11,15] 1):

$$S_{cp} = \frac{\sum_{n} CP_{ij}}{N_{cp}}, \quad (3)$$

where CP_{ij} represents the contact potential of residue i with residue j defined by Keskin. N_{cp} represents the number of contact residue pairs.

1) Lin, W. et al. A statistical analysis of protein-protein interaction with knowledge-based potential at residue level. Tsinghua Science and Technology, in Press.

(x) Patch interface propensities "*PIP*". *PIP* is a scoring function proposed by Jones and Thornton[12,16]:

$$PIP = \frac{\sum_{i}^{N_p} (\ln IP_{AA})}{N_p}. \tag{4}$$

This variable is derived from the relative frequency of different amino acid residues in the interface of complexes and used for prediction. Please refer to the Jones' papers[12,16] for the value of IP_{AA}.

(xi) The solvation potential of interface patch, "ΔSP", is also used to evaluate the preference of residues exposed to solvent environment. Please refer to the Jones' papers[12,16] for the calculation of ΔSP. When two components bind together, there is a change between the ΔSP of components and the ΔSP of the complex. We define S_{sp} to measure this change by the following equation:

$$S_{sp} = \frac{\sum_{i}^{N_{cp}} (\Delta SP_{compo} - \Delta SP_{compl})}{N_{cp}}, \tag{5}$$

where ΔSP_{compl} is the ΔSP of the complex and ΔSP_{compo} is the ΔSP of the components. The larger the S_{sp}, the more possible the patch is buried[12].

Among the ratio variables, R_a was calculated from the buried area divided by the surface area of the whole complex. The other three ratio variables were calculated from the corresponding variables divided by the atom numbers of both component molecules. A total of 11 descriptors for each proposed interface were taken as our input variables.

With these variables of the samples, we performed binary logistic regression analysis with SPSS 10.0 package. The principle of this work is as follows: A variable "y" represents the property of a proposed interface. If an interface is a real one from a complex, then $y=1$, while $y=0$ for a false one. By binary logistic regression analysis of these variables from training samples, an empirical probability equation would be obtained as follows:

$$P_{(y=1|x_i)} = \frac{e^{b_0+\sum b_i x_i}}{1+e^{b_0+\sum b_i x_i}}, \tag{6}$$

where x_i represents the variable values extracted from the proposed interfaces. $P_{(y=1|x_i)}$ represents the probability of "$y=1$", i.e. real interface, on condition that the variables have values of x_i.

With the values of the variables of a proposed interface obtained, we are then able to evaluate the probability of the complex interface being real by this equation.

To test the validity of such a model, 6 complex structures and 6 monomer structures outside the training set were used as validation samples. With their x_i and by Eq. (6), the probabilities were calculated.

Three potential dimeric interfaces of staphylokinase were also investigated with the model to verify their suitability to our model.

2 Results and discussions

The authenticity of the biological interfaces of the 25 complex samples has been proved by experiments. The 17 pseudo-interfaces of the monomeric proteins presented only in crystal structural experiments, without functional experiments demonstrating that they are biological interfaces, thus we treat these interfaces as pseudo or false ones. The values of the interfacial variables of the forty-two training samples are summarized in Table 2.

Table 2. Values of the variables extracted from the interfaces of training samples

PDB ID	Components' chain ID	A_b(Å^2)	N_b	N_h	S_{ch}	R_a(%)	R_n(%)	R_{hb}(%)	R_{ch}(%)	S_{cp}	*PIP*	S_{sp}
1avw	A & B	871	172	4	1	4.96	4.56	0.11	1.39	0.0048	0.218	0.018
1cho	E & I	752	130	4	2	11.63	5.28	0.16	2.99	0.0048	0.159	0.006
1cse	E & I	749	144	8	0	5.39	5.12	0.28	0	0.0054	0.164	0.038
1mkw	HL & K	293	107	0	3	1.12	1.96	0	10.7	0.0035	0.107	0.004
2ptc	E & I	716	136	8	1	5.50	5.55	0.33	1.49	0.0048	0.125	0.056
1bth	HL & P	1198	236	13	5	6.99	6.27	0.35	5.15	0.0043	0.134	0.053
1tbq	HL & R	1766	315	11	4	8.23	7.10	0.25	2.85	0.0046	0.065	0.037
1toc	AB & R	1776	316	18	2	8.32	6.88	0.39	2.03	0.0047	0.085	0.032
4htc	HL & I	1204.51	274	7	6	6.84	8.81	0.23	5.26	0.0049	0.121	0.047

To be continued

Continued

PDB ID	Components' chain ID	$A_h(\text{Å}^2)$	N_b	N_h	S_{ch}	$R_a(\%)$	$R_n(\%)$	$R_{hb}(\%)$	$R_{ch}(\%)$	S_{cp}	PIP	S_{sp}
1jhl	HL & A	633	112	5	2	3.80	3.80	0.17	4	0.0045	0.138	0.062
1nca	HL & N	990	167	9	1	2.91	2.34	0.13	1.11	0.0047	0.069	0.039
1osp	HL & O	756	158	5	−1	2.24	2.19	0.07	−1.89	0.0047	0.003	0.083
1brs	A & D	786	151	10	4	7.35	6.89	0.46	6.90	0.0049	0.074	0.055
1dhk	A & B	1538	285	12	5	5.75	3.57	0.15	4.07	0.0049	0.072	0.022
1fss	A & B	995	177	7	6	4.16	3.38	0.13	8.33	0.0052	0.133	0.036
1gla	F & G	657	86	1	6	2.50	1.59	0.02	12.00	0.0045	0.080	0.030
1ydr	E & I	1012	157	9	7	5.71	4.21	0.24	10.29	0.0046	0.089	0.030
1agr	A & E	838	143	7	5	3.37	2.82	0.14	7.81	0.0059	−0.024	0.089
1fin	A & B	1722	274	12	5	6.56	5.05	0.22	4.1	0.0047	0.090	0.041
1tx4	A & B	1154	176	9	6	6.09	4.72	0.24	6.67	0.0051	0.054	0.039
2trc	BG & P	2246	361	27	14	7.05	6.21	0.46	8.05	0.0041	0.101	0.045
1a71	A & B	1715	304	11	4	5.67	4.54	0.16	2.67	0.0044	0.124	−0.00113
1fc2	C & D	656	107	3	−1	4.26	3.96	0.11	−2.27	0.0036	0.258	0.000481
1igc	HL & A	683	117	8	0	2.33	2.69	0.18	0	0.0048	−0.097	0.104564
1ycs	A & B	745	140	6	5	3.87	3.80	0.16	11.1	0.0046	0.153	0.033946
1a8v	A & B	471	93	4	4	3.30	3.71	0.16	9.09	0.0058	0.085	0.080767
1a9m	A & B	1887	335	20	−1	14.02	15.1	0.9	−0.70	0.0044	0.157	−0.02235
1ba2	A & B	275	27	0	0	1.12	0.58	0	0	0.0066	−0.144	0.075485
1bb3	A & B	512	72	3	−1	3.81	2.67	0.11	−2.78	0.0049	0.076	0.042423
1bja	A & B	926	128	4	−1	8.19	6.15	0.19	−1.37	0.0044	−0.032	0.063337
1bry	Y & Z	85	3	0	−1	0.39	0.07	0	−14.29	0.0072	−0.260	0.092002
1cgf	A & B	197	28	2	1	1.20	0.80	0.06	9.09	0.0055	−0.022	0.068514
1e4y	A & B	554	73	0	6	2.65	1.66	0	15.79	0.0063	−0.160	0.132043
1fmt	A & B	600	74	2	0	2.11	1.32	0.04	0	0.0050	0.066	0.028203
1fvk	A & B	759	113	5	−1	4.05	3.08	0.14	−2.13	0.0052	0.075	0.054835
1h97	A & B	319	30	1	3	2.07	0.33	0.01	15.79	0.0058	0.042	0.058503
1iaz	A & B	411	63	3	0	2.60	1.72	0.08	0	0.0050	−0.058	0.040858
1lys	A & B	305	45	1	−1	2.33	1.99	0.04	−4.55	0.0064	−0.108	0.0706
1mdv	A & B	750	64	2	0	4.55	1.83	0.05	0	0.0059	−0.053	0.151799
1qin	A & B	3758	650	19	0	15.90	17.2	0.50	0	0.0044	0.128	0.013068
1sei	A & B	316	47	1	−1	2.17	1.64	0.03	−5	0.0061	−0.005	0.085762
3tmy	A & B	135	29	2	−2	1.11	1.17	0.08	−25	0.0064	0.079	0.080441

Different combinations of the eleven variables were analyzed with the regression program of the SPSS package. The parameter (b_i) of each variable as in Eq. (6), the statistical significance of each variable and the correct percentage of the model were computed by the program. Two indexes were used to monitor the selection of the discriminative variable combinations. One is the correct percentage of the model. The other is the significance index of each variable. The higher the correct percentages and the lower the significance indexes, the better. Furthermore, it is necessary to judge whether the model is consistent with well-accepted principles on protein-protein recognition. For instance, one of the combinations contains the interfacial area ratio (R_a) as a variable (x_i), with a negative value of parameter (b_i) in Eq. (6). As is well known, the bigger the interfacial area, the higher possibility the complex components bind together. Therefore a negative value of parameter for the interfacial area ratio is irrational, probably induced by noise, and should be eliminated.

With these 3 aspects satisfied, we obtained the following model as in Eq. (7) (Model A). The data of regression analysis are shown in Table 3.

$$P_{(y=1|x_i)} = \frac{e^{21.442+1.062\cdot S_{ch}-5635.242\cdot S_{cp}+79.449\cdot S_{sp}+27.462\cdot PIP}}{1+e^{21.442+1.062\cdot S_{ch}-5635.242\cdot S_{cp}+79.449\cdot S_{sp}+27.462\cdot PIP}}. \tag{7}$$

Table 3. Parameters for 4 variables of Model A and 2 variables of Model B

	Variables	S_{ch}	S_{cp}	S_{sp}	PIP	Constant
Significance	Model A	0.011	0.011	0.053	0.063	0.015
	Model B	0.011	0.008			0.008

It is found that the 4 variables in this equation are all energetic or semi energetic descriptive variables. The probabilities of the training samples were calculated. Thirty-eight of the forty-two samples are correctly classified, as shown in Table 4 and Fig. 1 (a). Model A presents the highest correct percentage with every significance index of the input variables and of the constant part (b_0) lower than 0.07, a satisfactorily acceptable value.

Table 4. Self classification of samples with regression Model A and Model B

		Pseudo	Real	Percentage correct
Model A	Pseudo	15	2	88.2
	Real	2	23	92.0
		Overall percentage		90.5
Model B	Pseudo	14	3	82.4
	Real	3	22	88.0
		Overall percentage		85.7

The cut value is 0.5.

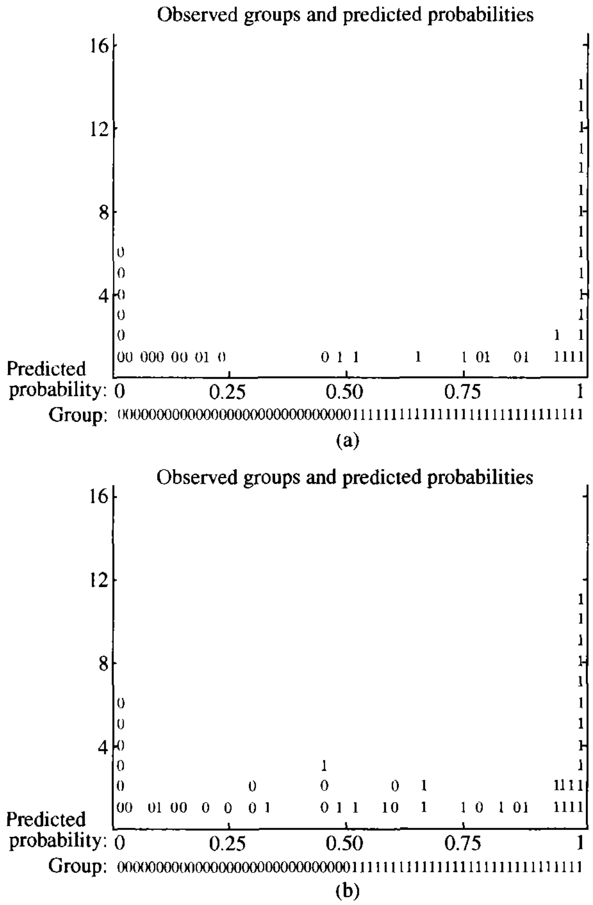

Fig. 1. Observed group and predicted probability with regression Model A (a) and Model B (b). Pseudo interfaces (0) and real interfaces (1) are finely clustered.

Twelve validation samples outside of the training set were also studied with Model A. Ten of them were correctly classified, as shown in Table 5. We also used Model A to test the three proposed dimeric structures of staphylokinase (1c77, 1c78 and 1c79) and found 1c78 presented highest probability, 0.99, indicating 1c78 is the most possible dimeric model. The results are shown in Table 6.

It was realized that two variables among the 4 in Eq. (7), patch interface propensities (*PIP*) and solvation potential of interface patch (S_{sp}), were very difficult to obtain for a proposed interface, because such an interface only provides structural details at low resolution. Therefore, we tried to eliminate these two variables from Model A and re-performed the regression analysis with the remaining two variables, shown in Table 3. It was discovered that the resulting model (Model B) was also sufficiently acceptable.

$$P_{(y=1|x_i)} = \frac{e^{17.082+0.895\cdot S_{cp}-3601.69\cdot S_{cp}}}{1+e^{17.082+0.895\cdot S_{ch}-3601.69\cdot S_{ch}}}. \quad (8)$$

The same evaluation process as Model A was performed as Model B. Thirty-six of the forty-two training samples are correctly classified, as shown in Table 4 and Fig. 1 (b). The significance indexes of the two variables and of the constant are less than 0.011. Judged with significance index values, Model B seems to be more reliable than Model A though its correct percentage is a bit lower.

For the validation of the twelve samples with Model B, as shown in Table 5, eleven were correctly classified. The three proposed dimeric structures of staphylokinase were also tested with Model B. The interface of 1c78 produced the highest probability, 0.67 (see Table 6), concurring with the analysis result with Model A in that 1c78 is the most possible dimeric model.

To further understand the discriminative capacity of the 4 individual variables used in our models, we analyzed the frequency distribution of these variables among the training samples shown as Fig. 2 (a) ~ (d). There are ambiguous separations between the distribution peaks of the real interfaces and those of the pseudo interfaces in these figures, which indicates these four individual variables could somewhat but not clearly differentiate the two types of interfaces. We also studied the frequency distribution of combination variables derived from Eqs. (7) and (8), as indicated in Eqs. (9) and (10) from Model A and Model B, and shown in Fig. 2 (e) and (f), respectively. The separations of the distributions in Fig. 2 (e) and (f) are significant, which suggests the combination variables *score*$_a$ and *score*$_b$ from Eqs. (9) and (10) are much more discriminative than the individual variables.

Table 5. Values of variables from 12 validation samples and the calculated probability of their interfaces

PDB ID	Components' chain ID	A_b(Å^2)	N_b	N_h	S_{ch}	R_a(%)	R_n(%)	R_{hb}(%)	R_{ch}(%)	S_{cp}	*PIP*	S_{sp}	$P_{md(A)}$	$P_{md(B)}$
1a2w	A & B	1998	354	24	−1	11.97	13.38	0.91	−0.66	0.0049	0.052	0.030	0.03	0.19
1a3y	A & B	786	134	2	−1	4.92	3.75	0.056	−1.85	0.0056	0.076	0.067	0.02	0.02
1afk	A & B	309	38	0	0	2.23	1.36	0	0	0.0054	−0.027	0.028	0.00	0.09
1ai9	A & B	472	75	4	1	2.29	1.83	0.1	3.84	0.0051	−0.016	0.065	0.18	0.40
1l9f	A & B	449	43	0	3	1.50	0.59	0	8.33	0.0065	−0.039	0.066	0.00	0.03
1mdt	A & B	388	49	2	−2	0.93	0.56	0.022	−8.70	0.0039	0.044	0.011	0.36	0.78
1acb	E & I	797	141	7	1	5.36	5.36	0.27	1.49	0.0045	0.193	−0.024	0.62	0.85
1bhf	A & I	275	43	1	2	3.93	3.00	0.07	13.3	0.0039	0.094	0.045	1.00	0.99
1gua	A & B	652	117	5	6	4.79	4.34	0.19	16.2	0.0046	0.047	0.080	1.00	1.00
1ppf	E & I	679	125	5	1	4.83	4.51	0.18	1.67	0.0040	0.123	−0.011	0.92	0.97
1yvn	A & G	1009	180	7	−1	4.36	3.72	0.145	−1.22	0.0044	0.122	0.006	0.35	0.58
2jel	HL & P	762	135	6	0	3.14	2.87	0.127	0	0.0047	0.080	0.034	0.46	0.54

Table 6. Values of variables from 3 proposed models of staphylokinase and the calculated probability of their interfaces

PDB ID	Components' chain ID	A_b(Å^2)	N_b	N_h	S_{ch}	R_a(%)	R_n(%)	R_{hb}(%)	R_{ch}(%)	S_{cp}	*PIP*	S_{sp}	$P_{md(A)}$	$P_{md(B)}$
1c77	A & B	542	68	2	0	3.28	2.46	0.07	0	0.0056	0.053	0.093	0.22	0.04
1c78	A & B	457	68	8	−1	2.76	2.44	0.28	−4.17	0.0043	0.051	0.096	0.99	0.67
1c79	A & B	216	17	2	1	1.31	0.61	0.07	7.14	0.0063	−0.113	0.096	0.00	0.01

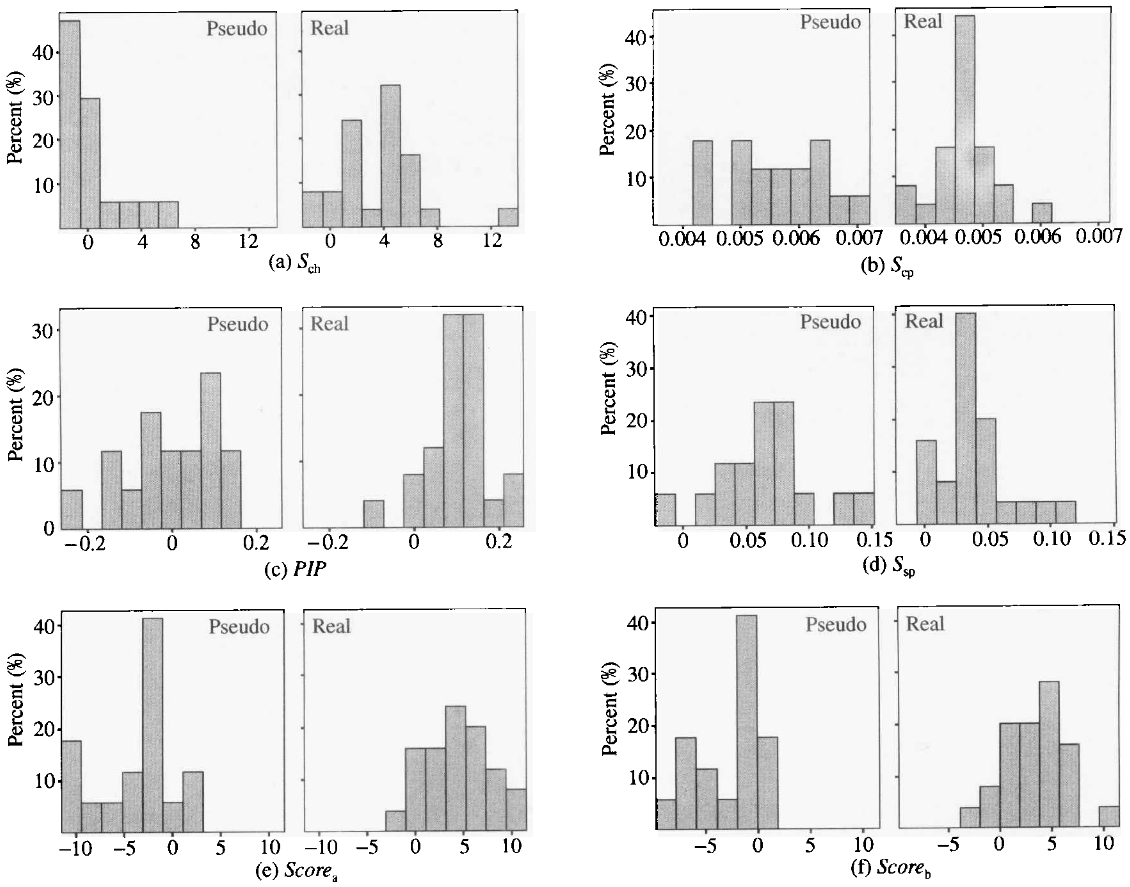

Fig. 2. Frequency distribution of four variables and two scores in real complex interfaces and pseudo interfaces. Individual variables present distribution separation ambiguously while *score*$_a$ and *score*$_b$ present significant separation.

$$score_a = 21.442 + 1.062 \cdot S_{ch} - 5635.242 \cdot S_{cp} + 79.449 \cdot S_{sp} + 27.462 \cdot PIP, \quad (9)$$

$$score_b = 17.082 + 0.895 \cdot S_{ch} - 3601.69 \cdot S_{cp}. \quad (10)$$

As is well known, the eleven variables can be divided into two classifications. One classification describes the shape features of the interfaces and the other describes the energetic features. As we can see from previous discussions about the regression analysis, the four selected variables are all energetic or semi energetic parameters. So our models strongly suggest that energetic variables contribute much more than shape variables.

Among the eleven variables we employed, score of charged residue pairs buried in the interface (S_{ch}) and interface-based residue-pair contact potential score (S_{cp}), and the two variables kept in Model B can be calculated at low resolution. Therefore, Model B provides a very practical method to evaluate the rationality of a proposed interface from two known surface patches.

In this paper, we employed the interfacial descriptors which are widely concerned by structural biologists. We previously derived S_{cp} and found that its performance is limited when used independently. However, combined with other descriptors, it contributed with very satisfactory performance. In the future, more discriminative interfacial descriptors might be introduced and included in the regression model and the prediction would be further improved.

References

1 Janin, J. et al. Surface, subunit interfaces and interior of oligomeric proteins. Journal of Molecular Biology, 1988, 204: 155.

2 Janin, J. et al. The structure of protein-protein recognition sites. Journal of Biological Chemistry, 1990, 265: 16027.

3 Hubbard, S. J. et al. Cavities and packing at protein interfaces. Protein Science, 1994, 3: 2194.

4 Jones, S. et al. Principles of protein-protein interactions. Proc. Nat. Aca. Sci. USA, 1996, 93: 13.

5 Milligan, R. A. Protein-protein interactions in the rigor actomyosin complex. Proc. Nat. Aca. Sci. USA, 1996. 93: 21.

6 Ban, N. et al. 1996. Crystal structure of an anti-anti-idiotype shows it to be self- complementary. Journal of Molecular Biology, 255: 617.

7 Wells, J. A. Binding in the growth hormone receptor complex. Proc. Nat. Aca. Sci. USA, 1996, 93: 1.

8 Duquerroy, S. et al. Protein-protein interaction: an analysis by computer simulation. Ciba Found. Symp., 1991, 161: 237.

9 Lawrence, M. C. et al. Shape complementarity at protein/protein interfaces. Journal of Molecular Biology, 1993, 234: 946.

10 Manocha, D. et al. Conformational analysis of molecular chains using nano-kinematics. Computational Application of Bioscience, 1995, 11: 71.

11 Keskin, O. et al. Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. Protein Science, 1998, 7: 2578.

12 Jones, S. et al. Prediction of protein-protein interaction sites using patch analysis. Journal of Molecular Biology, 1997, 272: 133.

13 Lo Conte, L. et al. The atomic structure of protein-protein recognition sites. Journal of Molecular Biology, 1999, 285: 2177.

14 Fraczkiewicz, R. et al. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. Journal of Computational Chemistry, 1998, 19: 319.

15 Miyazawa, S. et al. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. Journal of Molecular Biology, 1996, 256: 623.

16 Jones, D. T. et al. A new approach to protein fold recognition. Nature, 1992. 358: 86.