Molecular Biology, Vol. 37, No. 1, 2003, pp. 148–155. Translated from Molekulyarnaya Biologiya, Vol. 37, No. 1, 2003, pp. 164–173.
Original Russian Text Copyright © 2003 by Drozdov-Tikhomirov, Linde, Poroikov, Alexandrov, Skurida, Kovalev, Potapov.

STRUCTURAL–FUNCTIONAL ANALYSIS OF BIOPOLYMERS AND THEIR COMPLEXES

UDC 579.252.5

What Forces Can Determine the Formation of Highly Specific Protein–Protein Complexes?

L. N. Drozdov-Tikhomirov[1], D. M. Linde[2], V. V. Poroikov[2], A. A. Alexandrov[1], G. I. Skurida[1], P. V. Kovalev[1], and V. Yu. Potapov[2]

[1] Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182 Russia;
E-mail: drozdov@img.ras.ru

[2] Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, 110121 Russia

Received June 10, 2002

Abstract—A software package was designed and used in a detailed study of the contact regions (interfaces) of a large number of protein–protein complexes using the PDB data. It appeared that for about 75% of the complexes the amino acid composition of the subunit surface in the contact region is not essential. Thus one may suggest that, along with the amino acid residues at the interface, the residues in the interior of the globules substantially contribute to protein–protein recognition. Such interactions between quite remote residues are most probably of electrical nature, and are involved in recognition by contributing to the overall electric field created by the protein molecule; the configuration of this field is perhaps the definitive factor of recognition. The overall field of the protein molecule is additively built of the fields created by each constituent residue, and it can be calculated as a sum of the fields created by the protein multipole (aggregate of "partial" electric charges assigned to every atom of the protein molecule). Preliminary assessment of the remote electrostatic interaction has been performed for ribonuclease subunits in vacuum. The results are indicative of a real possibility that the electric field created by the protein multipole can strongly influence the mutual orientation of molecules before Brownian collision.

Key words: complex formation, molecular mechanism, interface, contacts, contact pair frequency, electrostatic interaction, partial charges, multipole

INTRODUCTION

The highly specific and rapid formation of a complex by a pair of protein subunits observed in the experiment raises a number of questions concerning the physical nature of this process; taken together, they are known as the problem of protein–protein recognition. The three basic questions are:

(1) What forces and factors do determine the structure and stability of the complex?

(2) What forces and factors do ensure the high rate of complex formation?

(3) Is the interface region structurally encoded (by the set of amino acid residues)?

Numerous and detailed studies on the structure of protein complexes [1–18] show that the complexes are mainly held together by forces involving the amino acid residues on the surfaces of the globules: hydrophobic interaction [4, 6, 8, 11], hydrogen bonding [15], salt bridging [15, 17], electrostatic interaction [2, 8, 9, 15], and interface hydration [5, 7, 14, 15]. Further, the specificity of binding, i.e., the choice of a few definite mutual orientations of subunits among the vast number of those possible in random Brownian collisions, is determined by shape complementarity of the contact regions [13], which allows simultaneous approach of interface atoms within distances required for establishment of H-bonds, salt bridges, and other short-range connections that together ensure tight association of subunits.

Much less clear is the physical nature of the mechanism ensuring rapid association of subunits. The experimentally measured rate of protein–protein association (~10^6 to 10^9 M^{-1} s^{-1}) exceeds by far the rate (~10^1 M^{-1} s^{-1}) [21] expected in the absence of long-range (non-contact) interactions that may orient the subunits prior to Brownian collision; this very fact irrefutably testifies to the existence of strong enough remote interactions between subunits during their approach. The overwhelming majority of researchers believe that these orientating remote forces are of electrical nature [19–22]. However, there are alternative suggestions. Thus Northrup and Erickson [23] by computer simulation of Brownian dynamics came to a conclusion that association rates about ~10^6 M^{-1} s^{-1} can be attained in a purely diffusional model, if one takes into consideration the hydration shell that can substantially increase the time of close contact between colliding subunits.

0026-8933/03/3701-0148$25.00 © 2003 MAIK "Nauka/Interperiodica"

The adherents of the electrostatic hypothesis base their arguments on the structure of complex interfaces [6, 10, 12, 16, 19–22] and on the data obtained upon introduction of amino acid changes [24]. These works demonstrate that the interfaces of many complexes include polar and charged residues, and their replacement with nonpolar alanine residues usually destabilizes the complex and substantially lowers the association rate. The shortcoming of this standpoint is that it relies on statistics, i.e., such conclusions pertain not to all complexes but just a large part thereof, and fails to answer the question about the forces that drive the association process when the interface lacks charged or polar residues.

As to the idea that the association rate can be enhanced by trapping of randomly colliding subunits in a hydration shell, which prevents their immediate separation, its weak point is the quite arbitrary choice of the numerical geometric criterion of collision efficacy, or admissible docking inaccuracy. If instead of the ±2 Å assumed by the authors [23] one takes a more rigorous one, say ±0.2 Å, the complex formation rate in these calculations becomes 4–5 orders of magnitude lower than the one observed experimentally. As unfortunately there is no way to strictly define the requirements for subunit docking accuracy, these results cannot convincingly argue in favor of a universal hydrophobic mechanism of protein–protein recognition.

Data that at first glance contradict the electrostatic hypothesis were obtained [25–27] in theoretical calculations of the contribution of electrostatic interaction of surface polar and charged residues into the free energy of subunit binding. In most cases, polar and charged residues at the interface were found to destabilize the complex. However, opposite data were also obtained in calculations [28], in particular, for hyperthermophilic proteins [29]. It should be noted that a destabilizing effect of electrostatic interactions between surface residues in the interface of a formed complex does not, generally speaking, rule out their substantial participation in pre-orientation of subunits during their approach, and therefore cannot be regarded as an argument against the hypothesis explaining rapid complex formation by electric interaction between interface residues.

We believed that the vast material of PDB could still yield much more information helpful in understanding the basic principles of the physical mechanism of protein–protein recognition; the present work is an attempt to move along this line.

DATA AND METHODS

Data set. Atomic structures of protein–protein complexes in PDB (http://www.rcsb.org/pdb/) were screened to choose 812 dimers solved by X-ray at resolution no worse than 3.0 Å (whereby the error in atomic coordinates is about 0.3 Å). We did not aim at comparable representation of all known classes of complexes. Accepted was any structure formed by noncovalent binding of two polypeptide chains. The list of the sample is given in Table 1[1].

Table 2. List of contacts for homodimer 1an9 (D-amino acid oxidase)

Contacts made by subunit A residues	Contacts made by subunit B residues
Sub A Sub B	Sub B Sub A
ALA 85 ➢ TRP 209	ARG 120 ➢ LEU 112, MET 110, VAL 111, d ASP 109
ARG 120 ➢ e ASP 109, MET 110	**ASN 86 ➢ LYS 271**
ASN 83 ➢ LYS 271	ASP 109 ➢ e ARG 120
ASN 86 ➢ ASP 272	**ASP 272 ➢ ASN 86**
ASP 109 ➢ d ARG 120	GLY 232 ➢ LYS 211
GLY 232 ➢ LYS 211	**LEU 233 ➢ PRO 208**
LEU 112 ➢ ARG 120	LYS 211 ➢ GLY 232, LYS 211
LYS 211 ➢ GLY 232, LYS 211	LYS 271 ➢ ASN 83, PRO 82
LYS 271 ➢ ASN 86	MET 110 ➢ ARG 120
MET 110 ➢ ARG 120	PHE 133 ➢ PHE 133
PHE 133 ➢ PHE 133, SER 93	**PRO 82 ➢ PRO 268**
PRO 82 ➢ LYS 271, PRO 268	PRO 268 ➢ PRO 82
PRO 208 ➢ LEU 233	SER 93 ➢ PHE 133
PRO 268 ➢ PRO 82	THR 90 ➢ TRP 209
THR 90 ➢ TRP 209	TRP 209 ➢ THR 90, ALA 85
TRP 209 ➢ THR 90	
VAL 111 ➢ ARG 120	

Boldface marks contact with a single residue of the opposite subunit; (e) immediate electrostatic contact ($R \leq 0.3$ Å); (d) distance electrostatic contact (0.3 Å $\leq R \leq 3.0$ Å).

Contact between two residues was registered if the distance between some atoms of these residues differed from the sum of their van der Waals radii [30] by no more than 0.3 Å (the accuracy of atomic coordinates). This contact criterion was applied in an exhaustive procedure to determine all residues of each subunit having contact(s) with residues of the other subunit; this information was stored as paired tables (as exemplified for complex 1an9(ab) in Table 2), and constituted the main material for analyzing the structure of contact regions.

The pair of values corresponding to the number of residues of each subunit in contact with the opposite subunit provides a measure close in its meaning to the

[1] Because of its volume Table 1 is not included here but is available in the Internet (http://obi.img.ras.ru/Humbio/proteins/table1.html) or from the authors upon request.

Table 3. Expectation of the contact pair frequency in random contacts. Gray shading marks values exceeding 500 (only above the diagonal, as the matrix is symmetrical)

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
ALA	130	289	201	231	24	195	318	146	62	79	134	383	40	66	166	273	237	43	104	114
ARG	289	640	446	511	52	432	704	325	138	174	296	848	88	145	368	604	525	95	231	253
ASN	201	446	311	356	36	301	491	226	96	121	206	591	61	101	256	421	366	66	161	176
ASP	231	511	356	408	42	345	562	259	110	139	236	677	70	116	294	482	419	76	185	202
CYS	24	52	36	42	4	35	58	27	11	14	24	69	7	12	30	49	43	8	19	21
GLN	195	432	301	345	35	292	476	219	93	118	200	572	60	98	248	408	355	64	156	171
GLU	318	704	491	562	58	476	775	357	152	192	326	933	97	160	405	665	578	105	255	278
GLY	146	325	226	259	27	219	357	165	70	88	150	430	45	74	186	306	266	48	117	128
HIS	62	138	96	110	11	93	152	70	30	38	64	183	19	31	79	131	114	21	50	55
ILE	79	174	121	139	14	118	192	88	38	47	81	231	24	40	100	164	143	26	63	69
LEU	134	296	206	236	24	200	326	150	64	81	137	392	41	67	170	279	243	44	107	117
LYS	383	848	591	677	69	572	933	430	183	231	392	1123	117	193	487	800	696	126	307	335
MET	40	88	61	70	7	60	97	45	19	24	41	117	12	20	51	83	72	13	32	35
PHE	66	145	101	116	12	98	160	74	31	40	67	193	20	33	84	137	119	22	53	58
PRO	166	368	256	294	30	248	405	186	79	100	170	487	51	84	211	347	302	55	133	145
SER	273	604	421	482	49	408	665	306	131	164	279	800	83	137	347	570	496	90	218	239
THR	237	525	366	419	43	355	578	266	114	143	243	696	72	119	302	496	431	78	190	208
TRP	43	95	66	76	8	64	105	48	21	26	44	126	13	22	55	90	78	14	34	38
TYR	104	231	161	185	19	156	255	117	50	63	107	307	32	53	133	218	190	34	84	91
VAL	114	253	176	202	21	171	278	128	55	69	117	335	35	58	145	239	208	38	91	100

water-inaccessible surface area [16]. These characteristics reflect the size of the interface, and are listed in two columns of Table 1.

Experimental frequency of every kind of contact pair (totaling 210 combinations of 20 elements taken two at a time) was determined using an *ad hoc* program for the entire interface data set.

Expected frequency of contact pairs was calculated with a model of two randomly colliding identical spheres each partitioned at random into 20 segments with areas relating as those occupied by corresponding residues on the surface of an "overall" protein [15]. In this model, the probability that segment (residue) of type **i** (**i** = 1, 2, …, 20) of one sphere comes in contact with segment of type **j** on the other sphere is obviously

$$P_{\mathbf{ij}} = \gamma_{\mathbf{i}}\gamma_{\mathbf{j}},$$

where γ is the respective share of surface. The expectation of the occurrence of **i**–**j** contact in N random collisions is

$$M_{ij}(N) = NP_{ij}.$$

Table 3 presents the full matrix thus generated.

Orienting effect of remote electrostatic interaction was evaluated by calculating *in vacuo* the energy of a system of two subunits with centers of gravity fixed at ~100 Å from each other at different angles of mutual rotation, and determining the depth of potential wells.

The energy of the overall electrostatic interaction was calculated as proposed by Kong (ftp://dashes.wustl.edu/pub/papers/kong-thesis.pz.gz) to reduce the computing time. The partial charge distribution over the residue atoms was calculated using the SYBYL package tools (Gasteiger, Gasteiger–Huckel, Huckel).

RESULTS

We analyzed the amino acid composition and structure of the interfaces of 812 protein–protein complexes, and determined the total number of residue–residue contacts, the number and kind of electrostatic contacts (charged residue with oppositely charged residue), and the number of hydrophobic contacts (between nonpolar residues) for every complex (Table 1, http://obi.img.ras.ru/Humbio/proteins/table1.html).

For 289 complexes, the data on their interfaces were presented as matrices (e.g., Table 4) containing the number of contacts of every possible kind. The matrix approach was used to test Mekler's [31] (Table 5) and related hypotheses [32] that the contact region is encoded through specific complementary interaction

Table 4. Matrix of the number of contact pairs at the interface of the dimer of subunits A and B of *Bac. stearothermophilus* phosphofructokinase (6PFK)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
N	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 5. Matrix representing Mekler's AA hypothesis: unities for contact (complementary) pairs that should be present in interfaces of complexes

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
R	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
N	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
C	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
G	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
I	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
L	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
P	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
S	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
T	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
V	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0

Table 6. Testing the validity of (I) Mekler's AA hypothesis [31] and (II) the first and (III) the second hypotheses of Blalock and colleagues [32]

	I	II	III
Total number of complexes examined	289	289	289
Number of complexes whose interface contains at least one contact of those predicted by the hypothesis	94	116	150

Table 7. Frequencies of contact pairs in the sample of 812 complexes. Gray shading marks values exceeding 500 (only above the diagonal, as the matrix is symmetrical)

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
ALA	316	296	256	170	56	172	158	170	108	184	236	102	204	206	96	184	170	102	272	188
ARG	296	548	382	974	72	406	876	330	200	192	446	230	156	298	328	290	342	164	524	280
ASN	256	382	328	194	24	178	222	266	102	192	268	292	88	172	160	206	240	96	288	126
ASP	170	974	194	244	66	198	132	170	284	84	156	538	88	136	124	276	220	144	262	88
CYS	56	72	24	66	156	38	46	76	40	36	32	62	38	46	58	52	30	28	68	36
GLN	172	406	178	198	38	400	196	198	130	190	240	188	156	170	200	206	202	88	258	194
GLU	158	876	222	132	46	196	216	198	252	200	196	634	122	198	208	270	252	80	386	222
GLY	170	330	266	170	76	198	198	360	108	138	202	166	108	124	214	210	202	100	314	190
HIS	108	200	102	284	40	130	252	108	256	128	216	106	138	76	96	138	184	96	240	180
ILE	184	192	192	84	36	190	200	138	128	268	338	150	142	304	78	138	186	130	198	318
LEU	236	446	268	156	32	240	196	202	216	338	796	216	208	338	210	274	272	162	378	398
LYS	102	230	292	538	62	188	634	166	106	150	216	196	102	154	120	216	194	76	218	136
MET	204	156	88	88	38	156	122	108	138	142	208	102	292	174	140	108	104	98	214	192
PHE	206	298	172	136	46	170	198	124	76	304	338	154	174	376	210	162	236	208	312	232
PRO	96	328	160	124	58	200	208	214	96	78	210	120	140	210	252	176	162	138	324	158
SER	184	290	206	276	52	206	270	210	138	138	274	216	108	162	176	320	258	96	218	158
THR	170	342	240	220	30	202	252	202	184	186	272	194	104	236	162	258	368	98	246	232
TRP	102	164	96	144	28	88	80	100	96	130	162	76	98	208	138	96	98	180	136	76
TYR	272	524	288	262	68	258	386	314	240	198	378	218	214	312	324	218	246	136	324	246
VAL	188	280	126	88	36	194	222	190	180	318	398	136	192	232	158	158	232	76	246	328

in pairs of amino acids determined by the complementarity of their genetic codes. The results of testing are given in Table 6, which shows how many complexes among those studied have in their contact region at least one amino acid pair of those implied by a hypothesis.

With the full sample (812) we determined the frequencies of each of the 210 possible kinds of contact pairs in the interfaces (Table 7). These values were compared with the theoretical estimates for random contact formation; the observed/expected frequency ratios (see Methods) are listed in Table 8.

Further, we assessed the influence of electric fields created by the partial charges of constituent atoms on the mutual orientation of subunits during their Brownian approach. For the *Streptomyces aureofaciens* ribonuclease homodimer (1rge), the energy of a system of two electrically interacting subunits mutually rotated at 100 Å between their centers of gravity (subunit diameter ~50 Å) was found to pass through nine almost equally deep minima. The well depth (ca. $1.2 \cdot 10^{-2}$ e/Å^2 or $2.9 \cdot 10^{-20}$ J) exceeded by an order of magnitude the kT at room temperature ($4.14 \cdot 10^{-21}$ J) (Table 9). Additionally it was found that the "sharpness of tuning" (the inverse of the angular deviation from the minimal-energy orientation that increases the system energy by 1/2 kT) in each minimum was quite high for rotations about the Y, Z and Y', Z' axes (~0.25 deg^{-1}) and much lower about X, X' (~0.03 deg^{-1}). The calculated orientation angles for spaced subunits at energy minima proved rather close to the values for the subunits fixed in complex (0, 0, 0, 0, 0) though not coincident (15°–25° deviations).

Table 8. Ratios of the real incidence of contact pairs to expectation for random contacts. Gray shading marks more than fivefold excess and frames, more than tenfold (only above the diagonal, as the matrix is symmetrical)

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
ALA	2.4	1.0	1.3	0.7	2.4	0.9	0.5	1.2	1.7	2.3	1.8	0.3	5.1	3.1	0.6	0.7	0.7	2.4	2.6	1.6
ARG	1.0	0.9	0.9	1.9	1.4	0.9	1.2	1.0	1.4	1.1	1.5	0.3	1.8	2.0	0.9	0.5	0.7	1.7	2.3	1.1
ASN	1.3	0.9	1.1	0.5	0.7	0.6	0.5	1.2	1.1	1.6	1.3	0.5	1.4	1.7	0.6	0.5	0.7	1.4	1.8	0.7
ASP	0.7	1.9	0.5	0.6	1.6	0.6	0.2	0.7	2.6	0.6	0.7	0.8	1.3	1.2	0.4	0.6	0.5	1.9	1.4	0.4
CYS	2.4	1.4	0.7	1.6	36.5	1.1	0.8	2.9	3.5	2.5	1.3	0.9	5.3	3.9	1.9	1.1	0.7	3.6	3.6	1.7
GLN	0.9	0.9	0.6	0.6	1.1	1.4	0.4	0.9	1.4	1.6	1.2	0.3	2.6	1.7	0.8	0.5	0.6	1.4	1.7	1.1
GLU	0.5	1.2	0.5	0.2	0.8	0.4	0.3	0.6	1.7	1.0	0.6	0.7	1.3	1.2	0.5	0.4	0.4	0.8	1.5	0.8
GLY	1.2	1.0	1.2	0.7	2.9	0.9	0.6	2.2	1.5	1.6	1.3	0.4	2.4	1.7	1.1	0.7	0.8	2.1	2.7	1.5
HIS	1.7	1.4	1.1	2.6	3.5	1.4	1.7	1.5	8.6	3.4	3.4	0.6	7.2	2.4	1.2	1.1	1.6	4.7	4.8	3.3
ILE	2.3	1.1	1.6	0.6	2.5	1.6	1.0	1.6	3.4	5.7	4.2	0.7	5.9	7.7	0.8	0.8	1.3	5.0	3.1	4.6
LEU	1.8	1.5	1.3	0.7	1.3	1.2	0.6	1.3	3.4	4.2	5.8	0.6	5.1	5.0	1.2	1.0	1.1	3.7	3.5	3.4
LYS	0.3	0.3	0.5	0.8	0.9	0.3	0.7	0.4	0.6	0.7	0.6	0.2	0.9	0.8	0.2	0.3	0.3	0.6	0.7	0.4
MET	5.1	1.8	1.4	1.3	5.3	2.6	1.3	2.4	7.2	5.9	5.1	0.9	24.1	8.7	2.8	1.3	1.4	7.5	6.7	5.5
PHE	3.1	2.0	1.7	1.2	3.9	1.7	1.2	1.7	2.4	7.7	5.0	0.8	8.7	11.4	2.5	1.2	2.0	9.6	5.9	4.0
PRO	0.6	0.9	0.6	0.4	1.9	0.8	0.5	1.1	1.2	0.8	1.2	0.2	2.8	2.5	1.2	0.5	0.5	2.5	2.4	1.1
SER	0.7	0.5	0.5	0.6	1.1	0.5	0.4	0.7	1.1	0.8	1.0	0.3	1.3	1.2	0.5	0.6	0.5	1.1	1.0	0.7
THR	0.7	0.7	0.7	0.5	0.7	0.6	0.4	0.8	1.6	1.3	1.1	0.3	1.4	2.0	0.5	0.5	0.9	1.3	1.3	1.1
TRP	2.4	1.7	1.4	1.9	3.6	1.4	0.8	2.1	4.7	5.0	3.7	0.6	7.5	9.6	2.5	1.1	1.3	12.7	3.9	2.0
TYR	2.6	2.3	1.8	1.4	3.6	1.7	1.5	2.7	4.8	3.1	3.5	0.7	6.7	5.9	2.4	1.0	1.3	3.9	3.9	2.7
VAL	1.6	1.1	0.7	0.4	1.7	1.1	0.8	1.5	3.3	4.6	3.4	0.4	5.5	4.0	1.1	0.7	1.1	2.0	2.7	3.3

DISCUSSION

As evident from Table 6, none of the above-mentioned coding hypotheses [31, 32] stands the matrix test, as they can hardly explain the formation of even half of the existing complexes. Thus, elegant as they are, these assertions give no clue to the molecular mechanisms of protein–protein recognition.

Inspecting the matrix of contact pair occurrence in the 812 interfaces (Table 7), one can see that contacts between polar and charged residues are most frequent; this might have supported the mechanism based on electrostatic interaction between residues on the contact surfaces. However, the dozen and a half pairs with frequency significantly exceeding the chance expectation (more than fivefold, as marked in the table) are formed by nonpolar (hydrophobic) residues, except for two pairs involving positively charged histidine. And yet one cannot say that these are the contacts definitive for recognition in all cases, because such contacts are found in only 24% of the complexes examined.

The appreciable overall frequency of contacts between oppositely charged residues is in line with the important role of electric interactions on the protein surfaces; yet 56% of the interfaces studied do not have a single electrostatic contact. It also does not help to assume that recognition is based on hydrophobic and/or electrostatic contacts, as neither type is found in 48% of cases when tight complexes are nonetheless rapidly formed.

On the strength of these results, we think that the association mechanism cannot be reduced to interaction between residues on the subunit surfaces, and attention should be paid to interactions involving residues in the globule interior. There are grounds [27] for supposing that such interactions may take place through the electric fields created by the multipole of each subunit (aggregate of 'partial' electric charges assigned to every atom of the overall electroneutral macromolecule).

The observation of deep energy minima for spaced ribonuclease subunits (Table 9) suggests that multipole interaction may give rise to a strong remote ori-

Table 9. Potential wells on the electrostatic energy surface for a system of spaced (100 Å between centers of gravity) ribonuclease subunits in vacuum

Gasteiger (partial charges)									
N	1	2	3	4	5	6	7	8	9
E	–1.1B2E–02	–1.165E–02	–1.157E–02	–1.162E–02	–1.167E–02	–1.163E–02	–1.167E–02	–1.161E–02	–1.161E–02
*A*cos *x*	–0.08	–0.08	0.36	–0.48	0.79	–0.82	0.92	0.95	0.97
y	0.97	0.97	0.88	0.87	0.53	0.57	0.29	0.31	0.14
z	–0.23	–0.23	–0.30	–0.13	0.29	0.02	0.25	0.11	–0.22
*B*cos *x*	0.30	–0.05	–0.29	0.45	–0.50	0.65	–0.66	0.75	0.72
y	–0.24	–0.11	0.00	–0.25	0.02	–0.28	0.08	–0.32	0.14
z	–0.92	–0.99	–0.96	–0.86	–0.87	–0.71	–0.75	–0.58	–0.68
Gasteiger–Huckel									
N	1	2	3	4	5	6	7	8	9
E	–1.252E–02	–1.251E–02	–1.249E–02	–1.254E–02	–1.255E–02	–1.250E–02	–1,253E–02	–1.247E–02	–1.242E–02
*A*cos *x*	(–0.08	(–0.08	(0.16	(–0.46	(0.79	(–0.82	(0.92	(–0.95	(0.96
y	0.97	0.97	0.95	0.67	0.53	0.57	0.29	0.31	0.26
z	–0.23)	–0.23)	–0.27)	–0.13)	–0.29)	0.02)	–0.25)	0.11)	0.08)
*B*cos *x*	(0.15	(–0.11	(–0.28	(0.45	(–0.50	(0.65	(–0.86	(0.75	(–0.79
y	–0.19	–0.11	–0.07	–0.25	0.02	–0.28	0.08	–0.32	0.31
z	–0.97)	–0.99)	–0.96)	–0.86)	–0.87)	–0.71)	–0.75)	–0.58)	–0.52)
Huckel									
N	1	2	3	4	5	6	7	8	
E	–1.185E–02	–1.186E–02	–1.183E–02	–1.188E–02	–1.182E–02	–1.187E–02	–1.181E–02	–1.183E–02	
*A*cos *x*	(–0.08	(–0.08	(–0.48	(0.79	(–0.82	(0.92	(–0.95	(0.97	
y	0.97	0.97	0.87	0.53	0.57	0.29	0.31	0.14	
z	–0.23)	–0.23)	–0.13)	–0.29)	0.02)	–0.25)	0.11)	–0.22)	
*B*cos *x*	(0.30	(–0.11	(0.45	(–0.50	(0.65	(–0.00	(0.75	(–0.72	
y	–0.24	–0.11	–0.25	0.02	0.28	0.08	–0.32	0.14	
z	–0.92)	–0.99)	–0.86)	–0.87)	–0.71)	–0.75)	–0.58)	–0.88)	

(*N*) Well number, (*E*) energy (e/Å^2), (*A*cos, *B*cos) cosines of the angles of rotation of subunits A and B about the respective axes (x, y, z) relative to their position in complex (for the latter all these values are unity).

entation effect, and seriously argues in favor of our suggestion concerning the role of multipole electric fields in protein–protein recognition.

The fact that these calculations were done *in vacuo* while the real molecules are in a dielectric milieu does not, in our opinion, depreciate the results. Indeed, the notion that water shielding should markedly attenuate the electrostatic interaction is based on formally applying the laws valid only for a homogeneous continuum. Obviously, in calculating electric fields created by internal charges of large molecules at distances commensurate with their size (and such is the case in our hands) the medium cannot be considered homogeneous. Generally speaking, there may even arise an opposite effect, namely, local enhancement of field intensity owing to unidirectional polarization of the medium near the protein surface. This question, of course, requires detailed study.

ACKNOWLEDGMENT

The authors wish to thank V.S. Skvortsov (Institute of Biomedical Chemistry) for expert help in calculating partial charges on ribonuclease atoms.

REFERENCES

1. Chotia C., Janin J. 1975. Principles of protein-protein recognition. *Nature*. **256**, 705–708.
2. Rashin A.A., Honig B. 1984. On the environment of ionizable groups in globular proteins. *J. Mol. Biol*. **173**, 515–521.

3. Thornton J.M., Singh J., Campbell S., Blundell T.L. 1988. Protein-protein recognition via side-chain interactions. *Biochem. Soc. Trans.* **16**, 927–930.

4. Young L., Jernigan R.L., Covell D.G. 1994. A role for surface hydrophobicity in protein-protein recognition. *Protein. Sci.* **3**, 717–729.

5. Janin J. 1995. Principles of protein-protein recognition from structure to thermodynamics. *Biochimie.* **77**, 497–505.

6. Jones S., Thornton J. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA.* **93**, 13–20.

7. Janin J. 1996. Protein-protein recognition. *Prog. Biophys. Mol. Biol.* **64**, 145–166.

8. Norel R., Petrey D., Wolfson H.J., Nussinov R. 1999. Examination of shape complementarity in docking of unbound proteins. *Proteins.* **36**, 307–317.

9. Tsai J., Lin S.L., Wolfson H., Nussinov R. 1996. A dataset of protein-protein interfaces generated with a sequence-order-independent comparison technique. *J. Mol. Biol.* **260**, 604–620.

10. Ausiello G., Cesareni G., Helmer-Citterich M. 1997. ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structure. *Proteins.* **28**, 556–567.

11. Jones S., Thornton J. 1997. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**, 121–132.

12. Janin J., Miller S., Chotia C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155–164.

13. Lawrence M.C., Colman P.M. 1993. Shape complementarity at protein-protein interfaces. *J. Mol. Biol.* **234**, 946–950.

14. Janin J. 1999. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recogniti. *Struc. Fold. Des.* **7**, 277–279.

15. Conte L.L., Chothia C., Janin J. 1999. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.

16. Palma P.N., Krippahl L., Wampler J.E., Moura J.J. 2000. BiGGER: a new (soft) docking algorithm for predicting protein interaction. *Proteins.* **39**, 372–384.

17. Xu D., Tsai C.J., Nussinov R. 1997. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein. Eng.* **10**, 999–1012.

18. Larsen N.A., Olson A.J., Goodsell D.S. 1998. Morphology of protein-protein interfaces. *Structure.* **6**, 412–427.

19. Buckle A.M., Schreiber G., Fersht A.R. 1994. Protein-protein recognition: crystal statistical analysis of a barnase-barstar complex at 2.0 Å resolution. *Biochemistry.* **33**, 8878–8889.

20. Zhou H.X. 1993. Brownian dynamics study of the influence of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J.* **64**, 1711–1716.

21. Sheinerman F.B., Norel R., Honig B. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159.

22. Norel R., Sheinerman F., Petrey D., Honig B. 2001. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein. Sci.* **10**, 2147–2161.

23. Northrup S.H., Erickson H.P. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA.* **89**, 3338–3342.

24. Bogan A.A. 1998. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.

25. Froloff N., Windeman A., Honig B. 1997. On the calculation of binding free energy using continuum methods: application to MHC class I protein-peptid interaction. *Protein Sci.* **6**, 1293–1301.

26. Schapira M., Totrov M., Abagyan R. 1999. Prediction of binding energy for small molecules, peptides and proteins. *J. Mol. Recogn.* **12**, 177–190.

27. Reddy V.S., Giesing H.A., Morton R., Kumar A., Post C.B., Brooks C.L. III, Johnson J.E. 1998. Energetics of quasiequivalence: computational analysis of protein-protein interaction in icosahedral viruses. *Biophys. J.* **74**, 546–558.

28. Xu D., Lin S.L., Nussinov R. 1997. Protein binding versus protein folding: the role of hydrophilic bridges in protein association. *J. Mol. Biol.* **265**, 68–84.

29. Xiao L., Honig B. 1999. Electroststic contribution to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**, 1435–1444.

30. Li A.J., Nussinov R. 1998. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins.* **32**, 111–127.

31. Mekler L.B. 1970. On specific selective interaction between residues in polypeptide chains. *Biofizika.* **14**, 581–584.

32. Bost K.L., Smith E.M., Blalock J.E. 1985. Similarity between the corticotropin (ACTH) receptor and a peptide encoded by an RNA that is complementary to ACTH mRNA. *Proc. Natl. Acad. Sci. USA.* **82**, 1372–1375.

33. Drozdov-Tikhomirov L.N., Linde D.M., Poroikov V.V., Alexandrov A.A., Skurida G.I. 2001. Molecular mechanisms of protein-protein recognition: whether the surface placed charged residues determine the recognition process? *J. Biomol. Struct. Dyn.* **19**, 279–284.