PROTEINS: Structure, Function, and Genetics 11:271–280 (1991)

Protein-Protein Recognition Analyzed by Docking Simulation

Jacqueline Cherfils, Stéphane Duquerroy, and Joël Janin
Laboratoire de Biologie Physicochimique, C.N.R.S. UA 1131, Université Paris-Sud 91405 Orsay, France

ABSTRACT **Antibody–lysozyme and protease–inhibitor complexes are reconstituted by docking lysozyme as a rigid body onto the combining site of the antibodies and the inhibitors onto the active site of the proteases. Simplified protein models with one sphere per residue are subjected to simulated annealing using a crude energy function where the attractive component is proportional to the interface area. The procedure finds clusters of orientations in which a steric fit between the two protein components is achieved over a large contact surface. With five out of six complexes, the native structure of the complexes determined by X-ray crystallography is among those retained. Docked complexes are then subjected to conformational energy refinement with full atomic detail. With Fab HyHEL 5 and lysozyme, a native-like complex has the lowest refined energy. It can also be retrieved when starting with the X-ray structure of free lysozyme. However, some non-native complexes cannot be rejected: they form large interfaces, have a large number of H-bonds, and few unpaired polar groups. While these are necessary features of protein–protein recognition, they are not sufficient in determining specificity.**

Key words: antigen–antibody recognition, protease-inhibitor complexes, simulated annealing, energy refinement, docking algorithm

INTRODUCTION

X-Ray structures of protease–inhibitor complexes and antibody–antigen complexes yield atomic details of the interactions that govern protein–protein recognition. These complexes have common features even though the proteins are not otherwise related. Their interfaces have the same size as judged from the area of buried protein surface, which is within 20% of 1,600 Å^2, they are close packed and they contain 8 to 13 hydrogen bonds or salt bridges.[1–4] This confirms that the thermodynamic basis of stable protein–protein association resides both in the hydrophobic effect, water being removed from the protein surfaces in contact, and in polar interactions.

The formation of these complexes requires no major conformation change, although small atomic movements occur. The exclusion of water at interfaces and the formation of favorable polar and van der Waals interactions imply that shape and chemical complementarity can be achieved with a minimum of structural readjustment. Complementarity is a necessary feature of protein–protein recognition. Is it sufficient or is there more than one solution to the problem of finding complementary surfaces on two protein molecules? To address this question, we create artificial complexes in the computer by docking protein components of real complexes in an arbitrary position and orientation. We use an efficient docking algorithm[5–7] that operates on a "simplified protein" model.[8] In this model, each amino acid residue is a single sphere, so that the number of degrees of freedom is minimum. With protease–inhibitor and antibody–lysozyme complexes, the procedure finds several plausible solutions. The native complex is among them in five out of six cases. Yet, other complexes have interfaces of similar size, and, when energy refinement is applied to their interface residues, they yield structures which have equivalent numbers of H-bonds, with few unpaired polar groups or none. Thus, the correct solution cannot be selected solely on the basis of these criteria. We suggest that some of the non-native solutions are alternative modes of association, that could be observed when mutations or chemical modifications prevent forming the native complex.

METHODS

Atomic Coordinates and Rigid Body Parameters

Atomic coordinates used in this work are from the X-ray structures listed in Table I. "Simplified protein" models are obtained by replacing each amino acid residue with a sphere centered on the center of gravity of its side chain. Cα included; the sphere radii are taken from Levitt.[8]

Following Levinthal et al.,[9] the six rigid-body pa-

Received March 20, 1991; revision accepted June 24, 1991.
Address reprint requests to Joël Janin, Professor of Biophysics, Laboratoire de Biologie Physicochimique, C.N.R.S. UA 1131, Bât. 433, Université Paris-Sud, 91405 Orsay, France.

TABLE I. X-Ray Structures of Protein–Protein Complexes*

Complex	Contact residues	H-bonds	Interface area (Å^2)	PDB file	Reference
Fab HyHEL5–lysozyme	30	11	1600	2HLF	10
Fab HyHEL10–lysozyme	34	13	1600	3HFM	11
Fab D1.3–lysozyme	27	12	1250	(*)	12
Trypsin–BPTI	30	11	1400	2PTC	13
Subtilisin–leech eglin C	35	12	1500	1CSE	14
Leukocyte elastase–ovomucoid	32	10	1300	(*)	15

*Data from Janin and Chothia.[1] Atomic coordinates are from Protein Data Bank files[16] or a gift from the authors (*). Contact residues have atoms within van der Waals radii plus 0.5 Å of atoms across the interface. Interface areas are evaluated from atomic coordinates with program ASA of Pr. A. Lesk (Cambridge) implementing the Shrake and Rupley[17] algorithm for accessible surface area calculations[18] with a probe radius of 1.4 Å. H-bonds assume O. . .A donor–acceptor distances less than 3.4 Å and C=O. . .A angles greater than 110°.

rameters that define the position and orientation of one molecule relative to the other are five angles and a distance: θ_1 and φ_1 locate the center O_2 of molecule 2 relative to molecule 1; θ_2 and φ_2 do the same for O_1, the center of molecule 1; χ is a spin angle about the center line; ρ is the center-to-center distance (Fig. 1A). All angles are set to zero in the orientation found in X-ray structures of the complexes. The angle convention is convenient in the sense that φ_1 and φ_2 are the longitudes of the regions of the two protein surfaces that form the contact, θ_1 and θ_2 their latitudes. In protease–inhibitor complexes, molecule 1 is the protease; in antibody–lysozyme complexes, it is the Fv part of the Fab comprising the two variable domains. Therefore, θ_1 and φ_1 define the position of the inhibitor binding site on the protease and of the combining site on the Fab; θ_2 and φ_2 define complementary sites on the inhibitors and lysozyme surfaces.

The Docking Algorithm

The docking algorithm has been described before.[6] Given a set of angles θ_1, φ_1, θ_2, φ_2, and χ, it brings molecule 2 in contact with molecule 1 by translation along the line of centers. Briefly, it operates as follows:

1. The two molecules are far apart; for each pair of residues i and j, we calculate the translation length ξ that brings them into contact; the N smallest values are retained (we use $N=10$).

2. Each of these N translations is checked for overlaps between pairs that yield smaller values of ξ; we retain the largest translation with no overlap.

The smallest translation, obtained for $N=1$, does not always yield the correct solution. Step 2, required in cases such as illustrated in Figure 1B, involves little overhead, since only $N(N-1)$ pairs are tested. Overlaps between residues are defined by the condition:

$$x = d/(r + r_j) < s$$

where r_I and r_j are Levitt's radii, d is the distance of the residue centroids, s is a parameter fixed to a value $s = 0.75$, making the spheres soft. Interior residues having less than 2% accessible surface area are excluded from the calculation.

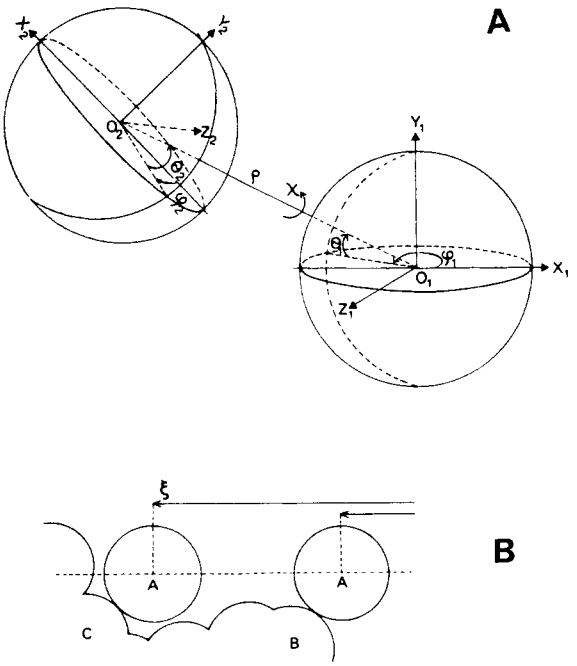

Fig. 1. Angular coordinates and the docking procedure. **(A)** ρ is the distance of the centers of mass O_1 and O_2; φ_1 and θ_1 are the longitude and latitude of O_2 in a reference frame attached to molecule 1: φ_2 and θ_2, the corresponding angles for O_1. These angles locate the region of contact on each molecule. The spin angle χ defines the relative orientation about the line of centers. All angles are set to 0 in the native complexes. **(B)** A sketch of the docking procedure: molecule 2, drawn as single sphere, moves along the line of centers (dashed); as it approaches molecule 1 from the right, it first touches atom B, but must move further to the left to reach its correct docking position corresponding to a translation ξ.

Evaluation of Docked Complexes

Two quantities are calculated on docked complexes: an approximate interface area B^* and an approximate repulsive energy E^*. The interface area is

the surface area buried in a complex, equal to the sum of the solvent accessible surface areas[18] of the two components less that of the complex. B^* is evaluated with a fast analytical approximation,[19] which underestimates interface areas by 10 to 15%. When necessary, exact accessible surface areas and interface areas are also calculated on detailed atomic models.

E^* is an estimate of molecular overlaps calculated from the repulsive components of Levitt's "soft" 6–8 potential. For two spheres with radii r_j and r_j and centers distant of d, we take

$$E_{ij} = \epsilon_{ij}\,(1+3/x^8 - 4/x^6) \quad \text{for } x = d/(r_i + r_j) < 1$$
$$E_{ij} = 0 \quad \text{for } x \geq 1.$$

The parameters ϵ_{ij} are taken from reference 8. E^* is the sum of all interaction energies, E_{ij} between pairs of residues from the two molecules. It would be zero after docking if parameter s was 1. With s = 0.75, E^* is positive but small, in the range 5–15 kcal mol^{-1}. The calculation of the translation ξ and of the two quantities B^* and E^* takes about 0.2 sec for lysozyme on a VAX station 3000.

It is convenient to bring E^* and B^* to the same scale by calculating a pseudo-energy:

$$E_B = E^* - \gamma B^*.$$

The attractive component is $-\gamma B^*$. Constant γ converts a surface area into an energy and has the dimension of a surface tension. Estimates of its value for amino acids[20,21] range from 25 to 47 cal/mol·Å^2. Here, we use γ = 50 cal/mol·Å^2 as preliminary tests suggested that smaller values give insufficient weight to B^*.

Simulated Annealing

The docking parameters are the θ_1, φ_1, θ_2, φ_2, χ angles, the calculation of the translation ξ determining the sixth parameters ρ. We use the Metropolis method[22] to explore the five parameter space as follows. A random set of angles is chosen, docking is performed, and E_B is evaluated; either one of two of the five angles are changed and the calculation repeated. If E_B has decreased, this orientation is accepted and becomes the new state. If it is higher, it can still be accepted, but with a probability $p(T)$ that depends on the pseudo-energy difference ΔE_B:

$$p(T) = \exp(-\Delta E_B/kT).$$

In this formula, inspired from Boltzmann's distribution in statistical mechanics, the "temperature" T is adjusted so that $p(T)$ is about 0.5 in early steps. T is progressively lowered, making it less and less likely that a step uphill is accepted, until the system freezes in a local minimum of E_B. The procedure then restarts from another orientation. After a sufficient number of annealing tests, all deep local minima, including the global minimum, should be sampled. This is known as a Monte Carlo simulated annealing.[23] Although the theory implies that new orientations should be near the current one, thus defining a Markov chain,[24] we find it more efficient to allow large jumps by drawing some angles at random while keeping the others unchanged. The temperature is decreased in steps of 10%, with either 100 successful jumps or 3,000 orientations tested at each temperature. Restarts are done when the success rate falls to less than 1%. Each simulation includes 20 restarts.

Conformational Energy Refinement

Energy minimization of artificial complexes is performed with program X-PLOR[25] on detailed atomic models derived from X-ray coordinates after applying the rigid-body transformations determined above. Energy parameters are taken from Brooks et al.[26] All heavy atoms and polar hydrogens of interface residues, defined as having at least one atom less than 7 Å from an atom of the other molecule, are refined. Side chain atoms of these residues are free to move in the force field. Their main chain atoms are subjected to a soft harmonic constraint keeping them near their position in the X-ray structure. A tighter constraint is applied to all atoms 7 to 9 Å from the interface, and atoms further than 9 Å are kept fixed. The interface layer includes about 800 atoms; 100 steps of conjugate gradient are performed on each complex, most side chain rearrangements being completed in about 50 steps.

RESULTS

Reconstitution of Native Complexes

As a test of the docking procedure, each of the six complexes is dissociated and reassembled by docking after applying small step rotations to molecule 2 while keeping molecule 1 fixed. Figure 2 illustrates the result obtained with lysozyme and Fab HyHEL5: the interface area B^* after docking has a maximum at angles less than 2° from the native complex. In this orientation, the docked position of the molecular center of lysozyme is within 0.1 Å of that of the X-ray structure, and the interface area B^* is the same. The maximum of B^* is also a minimum of the pseudo-energy E_B, since values of E^* after docking are rather insensitive to orientation parameters. In contrast, the interface area is quite sensitive to orientation: B^* drops to values less than 500 Å^2 upon rotations of a few degrees. Thus (1) the native Fab HyHEL5-lysozyme complex can be reconstituted by docking; (2) it is very near a local minimum of E_B; and (3) there is no other minimum in the angular range explored in Figure 2.

In this case, the native orientation is actually the global minimum of E_B for all θ_2, φ_2, and χ rotations. In other words, no other orientation of lysozyme and no other orientation of lysozyme and no other region of its surface forms a larger interface with the com-

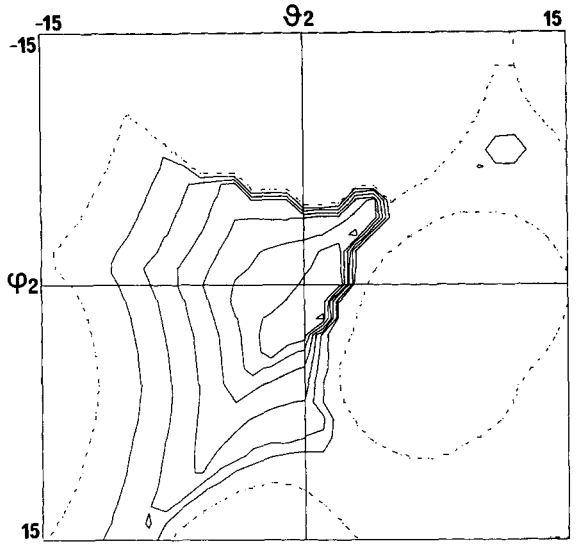

Fig. 2. Reconstitution of the Fab HyHEL5–lysozyme complex. Lysozyme is docked onto the combining site of Fab HyHEL5 after undergoing θ_2 and φ_2 rotations of up to 15° relative to the X-ray structure of the complex.[10] Values of the interface area are B^* estimated with the analytical approximation.[19] Contour lines range from 350 Å^2 (dashed line) to 1,300 Å^2 in steps of 150 Å^2. The maximum of B^* is 1,420 Å^2 at angles $\theta_2 = 1°$, $\varphi_2 = 2°$; the native complex is at the origin.

bining site of Fab HyHEL5. This is apparent in Figure 3, which is a histogram summarizing more than 3×10^4 dockings performed after applying grid step rotations to lysozyme while keeping the Fab fixed. The complex reconstituted in the native orientation is at the top of the histogram. In most orientations, E_B is above -25 kcal mol^{-1} and B^* less than 500 Å^2. Only 14 orientations yield values of B^* larger than 1,000 Å^2.

The X-ray structure of the three protease–inhibitor complexes is also near a minimum of the pseudo-energy E_B. Histograms similar to Figure 3 are obtained when all orientations of the inhibitors are tested, the location of the target surface on the enzyme remaining fixed. Less than 0.3% of 3×10^4 dockings yield B^* larger than 1,000 Å^2. The docking performed in the native orientation, or an orientation one grid step away, is among the top five. The lysozyme complex reconstituted with Fab HyHEL5 also behaves in this way, while that reconstituted with Fab D1.3 ranks only 61th.

Artificial Complexes From Simulated Annealing

A complete grid search on five angles would require excessive computer time, and most of it would be wasted in testing uninteresting orientations. Simulated annealing is a more efficient way of exploring the parameter space. We applied it to the six complexes, keeping θ_1 and φ_1 in a 30° window around the native orientation, while using the whole range of φ_2, and φ_2, and χ. This amounts to testing all parts of the lysozyme or inhibitor surface for association with a region that extends about 4 Å beyond the edge of the combining site of the Fab or the active site of the proteases.

Figure 4A represents those orientations of lysozyme that simulated annealing selects as having low values of pseudo-energy E_B after docking onto Fab HyHEL5. They cluster in about 30 groups. The cluster at the origin represents native-like complexes, others point to regions of the lysozyme surface that are sterically complementary to the combining site of the Fab. They are centered near the points where lines drawn on Figure 4B cut the surface of the lysozyme molecule. They spread over the whole surface, with no obvious relationship to curvature or other simple geometric features.

Results obtained with the other two Fabs and with the trypsin-inhibitor complexes are qualitatively similar. The regions of the lysozyme surface that are found to be complementary to Fab HyHEL10 and Fab D1.3 are however less extensive than with Fab HyHEL5. When docking the ovomucoid inhibitor onto human elastase, simulated annealing detects potential binding sites covering most of the inhibitor surface. This is not true of BPTI or of eglin. The native binding site of these two inhibitors is a strongly curved surface at the end of an elongated molecule. It is detected by the simulation, but most other potential binding sites are 90° away on flatter parts of the inhibitors' surface. Simulated annealing confirms that the native trypsin–BPTI and subtilisin–eglin complexes are near local minima of E_B. Still, many deeper minima are found, some of which correspond to artificial complexes having interfaces 20 to 60% larger than the native structures (Table II).

Energy Refinement of Artificial Complexes

Up to now, we have used "simplified protein" models with a crude energy function that ignores details of atomic packing and has no electrostatic component. We can easily recover the detailed atomic structure by applying to the original X-ray coordinates, rigid-body transformations determined with the simplified model. However, side chains conformations do not necessarily fit in their new environment, and chemical complementarity is not achieved. Therefore, the reconstituted interfaces contain bad contacts and electrostatic interactions are poor.

These defects are corrected by applying energy refinement to atoms in the interfaces. For each complex, we refine 30 structures representative of clusters detected by simulated annealing. All structures reach low conformational energies in 100 steps of conjugate gradient. Examination of their interfaces indicates that bad contacts have been removed and H-bonds with correct geometries have appeared. Ta-

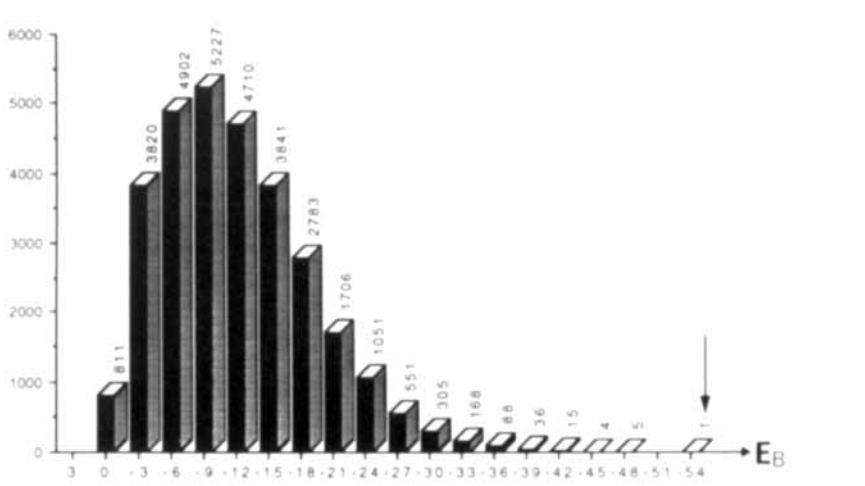

Fig. 3. Histogram of Fab HyHEL5–lysozyme dockings. 30,784 orientations of lysozyme are generated by sampling θ_2 in steps of 5° over 360°, φ_2 in steps of 10° over 180° and χ in steps of 10° over 360°. In each orientation, lysozyme is docked onto the combining site of Fab HyHEL5 and the pseudo-energy E_B is evaluated. The mean value of E_B is -12.4 kcal mol^{-1}, that of B^* is 300 Å^2. The arrow points to the complex reconstituted in the native orientation ($\theta_2 = \varphi_2 = \chi = 0$), for which $E_B = -56.0$ kcal mol^{-1}, $B^* = 1{,}200$ Å^2.

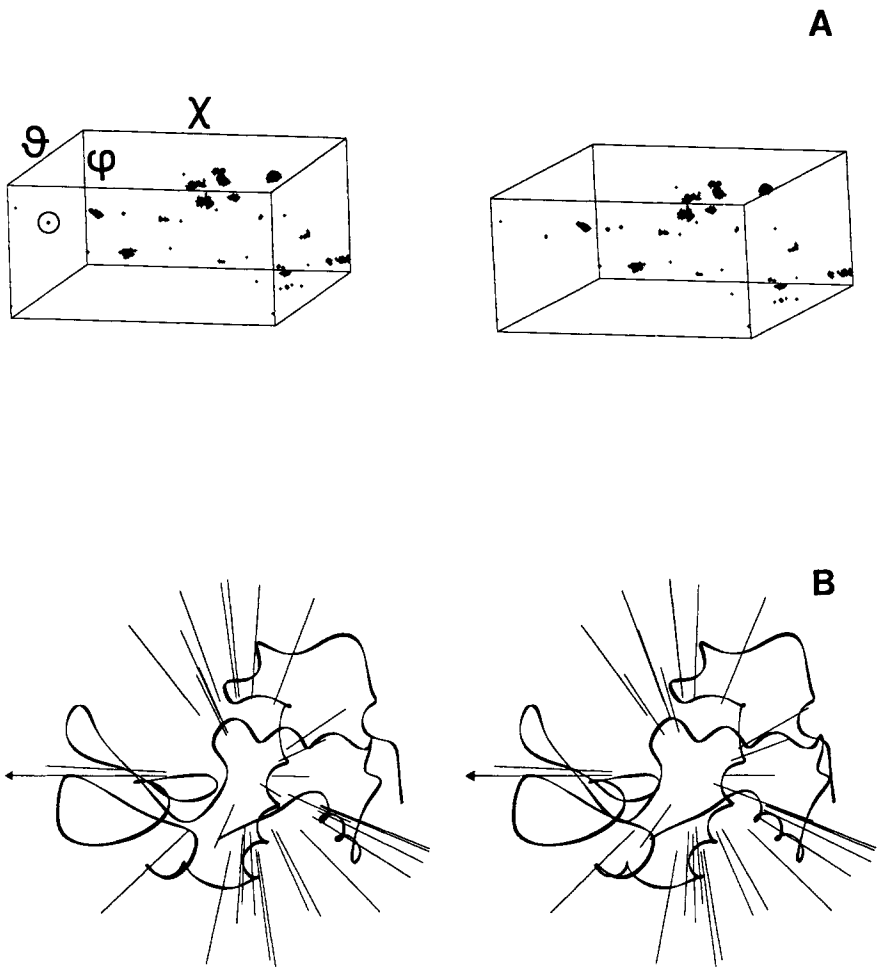

Fig. 4. Lysozyme orientations in artificial Fab HyHEL5 complexes derived from simulated annealing. **(A)** 3,300 orientations of lysozyme with E_B less than -60 kcal mol^{-1} were found by simulated annealing; they are plotted as dots in a stereo view of the θ_2, φ_2, χ space, θ_1 and φ_1 (not shown) are in the range (-15,15°). The circle points to the native orientation. **(B)** Stereo view of the lysozyme backbone; orientations representative of clusters in **(A)** are drawn here as lines joining the molecular centers. The orientation corresponding to the X-ray structure is shown by an arrow.

TABLE II. Artificial Complexes From Simulated Annealing*

Complex	E_B (kcal mol^{-1})	B^* (Å^2)
Fab HyHEL5–lysozyme		
Native	−61.7	1330
Native local minimum	−73.2	1630
Global minimum	−78.4	1920
Fab HyHEL10–lysozyme		
Native	−46.7	1520
Native local minimum	−59.1	1360
Global minimum	−71.3	1650
Fab D1.3–lysozyme		
Native	−41.6	1070
Global minimum	−85.0	1970
Trypsin—BPTI		
Native	−56.1	1240
Native local minimum	−54.6	1270
Global minimum	−70.9	1660
Subtilisin–eglin		
Native	−55.3	1250
Native local minimum	−59.1	1340
Global minimum	−73.4	1600
Leukocyte elastase–ovomucoid		
Native	−53.1	1250
Native local minimum	−48.2	1120
Global minimum	−79.0	1690

*Artificial complexes are generated by docking lysozyme onto Fab's or inhibitors onto proteases in arbitrary orientations. The range of φ_1 and θ_1 is (−15°,15°); that of θ_2 (−90,90°) φ_2 and χ (−180,180°). "Native" complexes there are X-ray structures changed to Levitt's simplified protein model. "Native local minima" are the result of docking in an orientation less than 10° from the native. B^* is the interface area calculated with the analytical approximation.[19]

ble III describes features of some refined artificial complexes, selected as having the lowest energies and the largest number of interface H-bonds. At the top of the list is a native-like Fab HyHEL5-lysozyme complex, that is, a complex reconstituted in an orientation very close to the X-ray structure. A native-like complex is also detected with Fab HyHEL10, but not with Fab D1.3 or with protease–inhibitor complexes. Both native-like lysozyme–Fab complexes have interfaces that are sensibly larger than in the X-ray structures. This is due in part to energy refinement which pulls some flexible side chains from the periphery into the interface. When the same refinement procedure is applied to the X-ray structures themselves, we observe the same effect. The resulting interface areas and number of contact residues are quoted in Table III; they are about 10% larger than in Table I.

Refinement is also efficient in creating H-bonds. The native-like Fab HyHEL5–lysozyme complex has the correct number of H-bonds, that with Fab HyHEL10 misses three. The procedure is less successful in reconstituting other complexes. When the trypsin–BPTI and subtilisin–eglin complexes are reconstituted near the native orientation, the resulting interfaces contain fewer than the 8 H-bonds cutoff used in Table III; with elastase–ovomucoid complexes, the largest number of H-bonds achieved is only six.

Docking With Models of the Free Proteins

Atomic coordinates of protein molecules derived from the complexes are biased in favor of the native orientation, where side chains fit at the interface. Though the effect of preformed side chains should largely vanish when residues are reduced to spheres, we also performed docking simulations starting with the atomic coordinates of "free" lysozyme, BPTI, and trypsin, that is, starting from X-ray structures of the proteins instead of the complexes. Results are mentioned in Table III along with those obtained above, the complexes listed being also selected on the basis of their conformational energy and of the number of H-bonds after refinement. Again, a native-like Fab HyHEL5–lysozyme complex is at the top of the list, implying that the free lysozyme structure fits well in the combining site of this Fab, and that most side chain conformation changes required for correct interaction can be simulated in 100 steps of conjugate gradient. Still, 3 H-bonds are lost when compared to the X-ray structure or to the complex reconstituted from its elements. An even larger number of H-bonds is lost with trypsin and BPTI starting from the free proteins models.

DISCUSSION

Our procedure includes a selection of artificial complexes followed with refinement. The first step relies on the docking procedure of Janin and Wodak[6] to match surfaces with complementary shapes. Docking rigid bodies is a classical problem in molecular recognition. Most proposed solutions deal with a small molecule ligand binding to a macromolecule, which is of major interest to pharmacology. Nevertheless, the a priori more complex docking of two macromolecules has been attempted as early as 1972 by Blow et al.[29] They proposed a model of BPTI bound to trypsin, based on biochemical as well as on structural information, which the X-ray structure of the complex showed to be essentially correct.[30] Automatic procedures developped later rely either on electrostatics or on a geometric description of the molecular surface. The simulation of redox complexes by Salemme[31] is of the first kind: electric charges are matched on the surface of two molecules. Warwicker[32] also finds matching electrostatic potentials on the surfaces of trypsin and BPTI, and of Fab HyHEL5 and lysozyme. The match is highly sensitive to the conformation of long flexible side chains that carry most of the relevant charges.

Our algorithm is of the second kind. Like those of Greer and Bush,[33] Kuntz et al.,[34] Lee and Rose,[35] Connolly,[36] and the recent algorithm of Jiang and

TABLE III. Characteristics of Artificial Complexes*

		Rank	Contract residues	H-bonds	Interface area (Å^2)
Fab HyHEL5–lysozyme					
From complex	(a)	—	34	15	1790
	(b)	1	40	11	2060
		2	34	12	1900
		3	33	10	1580
		4	30	8	1770
		5	30	9	1560
		6	29	11	1820
		7	31	10	1700
Free protein	(b)	1	36	8	1890
		2	32	10	1690
		3	34	8	1760
		4	33	8	1970
		5	41	9	2090
		6	32	11	1990
Fab HyHEL10–lysozyme					
	(a)	—	37	11	1820
		1	39	10	2040
		2	39	12	2070
	(b)	3	33	10	1660
		4	30	9	1450
		5	30	9	2000
		6	34	8	1380
		7	29	8	1770
		8	35	11	1380
		9	38	9	1820
Fab D1.3–lysozyme					
	(a)	—	29	11	1360
		1	28	9	1480
		2	32	11	1570
		3	30	11	1470
		4	24	11	1080
Trypsin–BPTI					
From complex	(a)	—	34	14	1510
		1	29	8	1770
		2	31	10	1700
		3	29	9	1720
		4	35	9	1460
Free proteins		1	36	8	1770
		2	31	8	1640
		3	29	8	1610
		4	31	8	1710
Subtilisin–eglin					
	(a)	—	38	13	1600
		1	30	11	1500
		2	31	9	1650
		3	35	10	1730
		4	36	9	1780
		5	34	8	2010
		6	37	9	1740
		7	29	9	1660
		8	37	12	1810
Elastase–ovomucoid					
	(a)	—	32	7	1410
		1	31	6	1630
		2	27	5	1370
		3	31	5	1560

*Artificial complexes are ranked according to the energy of interaction calculated after refinement with X-PLOR. Each complex is representative of a cluster from simulated annealing. Only complexes having more than 8 H-bonds (or 5 H-bonds for elastase–ovomucoid) after refinement are listed. H-bonds and interface areas are evaluated as in Table I. "Free proteins" refers to dockings performed with models of free lysozyme (Protein Data Bank file 6LYZ; ref.27), trypsin (File 1TPO; ref. 28), and BPTI (file 4PTI; ref. 28). (a) X-ray structures after energy refinement. (b) Native-like complexes within 7° from the orientation of X-ray structures.

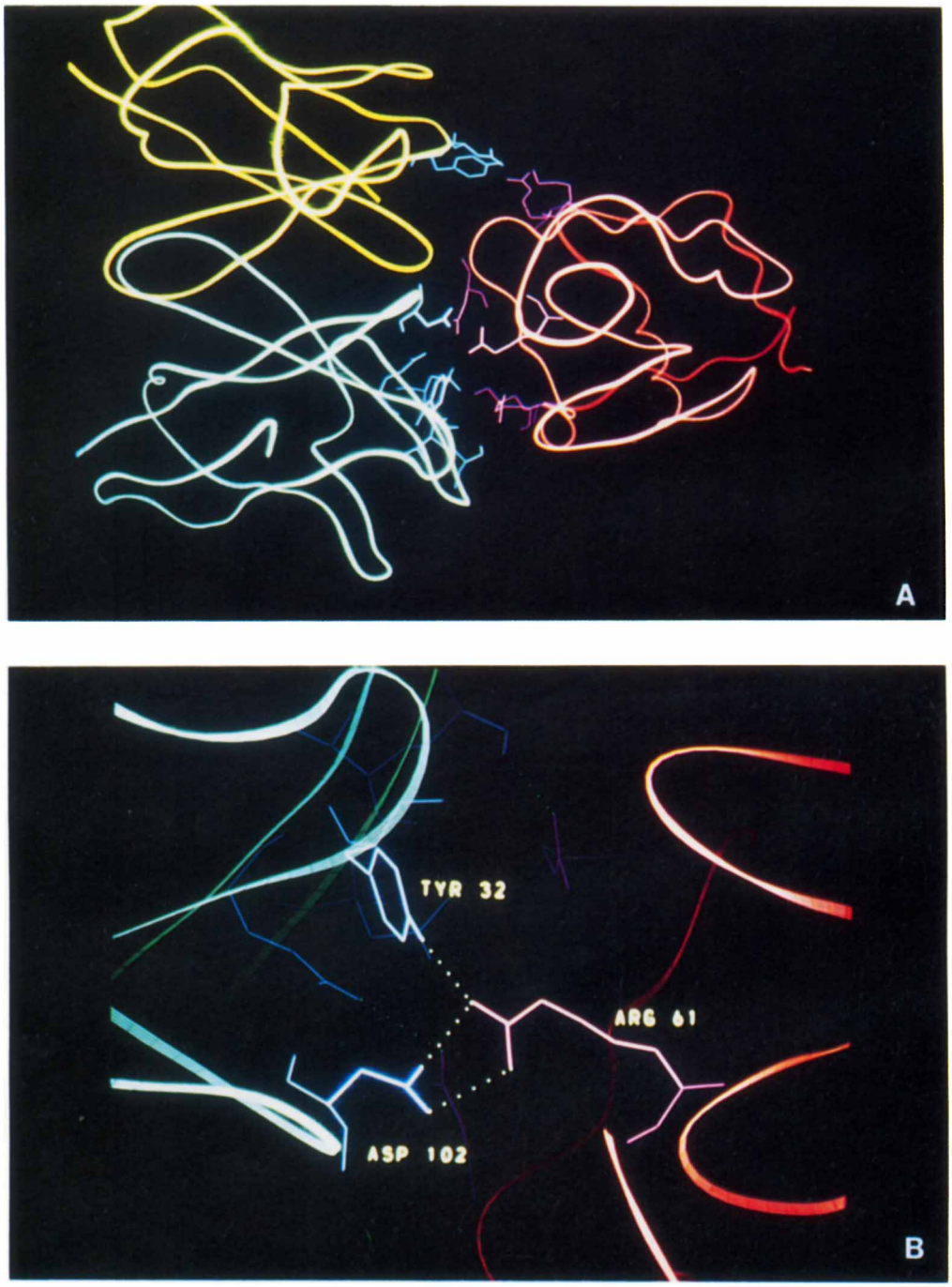

Fig. 5. An artificial Fab HyHEL5–lysozyme complex. The non-native complex illustrated here ranks second in Table III. **(A)** C_α trace of the variable domains of the light chain (green) and of the heavy chain (blue) of antibody HyHEL5; lysozyme is in red. Side chains involved in H bonds are shown. **(B)** Close-up view of the salt bridge appearing after energy refinement. Residues Asp-102 and Tyr-32 in blue belong to the heavy chain, Arg-61 in red, to lysozyme.

Kim,[37] it attempts to match geometric features, "knobs and holes", on two molecular surfaces. The algorithm of Kuntz et al.,[34] which uses a negative image of the protein surface rather the surface itself and fits ligand atoms to that imprint, has been most extensively tested for small ligand building.[38] The procedure of Jiang and Kim[37] has been designed more specifically for protein docking. Each molecule is reduced to a set of surface grid points and a set of internal volume grid points. Docking molecules in a given orientation is done by matching surface grid points. Optimal orientations are those with many

matching surface points and few overlapping volume cubes. Qualitative H-bonding and hydrophobicity scores are then applied to contact residues in order to evaluate interactions. This procedure appears to be of comparable efficiency to ours, and it has also been applied to the trypsin–BPTI and Fab HyHEL5–lysozyme complexes. The native structure is detected among the best solutions, but several solutions have higher scores, and they cannot be rejected on simple inspection or on the basis of Eisenberg–McLachlan[39] hydrophobicity parameters.

The conclusion that geometric complementarity between protein surfaces is achieved in more than one way, had been reached earlier by Wodak and Janin[5] in a simulation of BPTI bindings to trypsin. The present study confirms that the native complexes are at a local maximum of the interface area, but not necessarily at the global maximum. The findings that chemical complementarity (e.g., H-bonding properties) is achieved in some non-native complexes are novel. While docking is performed on simplified protein models, chemical complementarity is tested on detailed atomic models with realistic energy parameters. Energy refinement carried out in a rather conservative way by moving only side chains at the interface is surprisingly efficient in creating H-bonds and other electrostatic interactions. After refinement, 7 out of 30 artificial Fab HyHEL5–lysozyme complexes have 8 or more interface H-bonds. There is no penalty against unpaired buried polar groups in energy refinement done without solvent, yet we found by inspection that with the exception of two, our artificial complexes had no more than one unpaired polar group buried at the interface. The native complex also has 8 H-bonds and an unpaired polar group when reconstituted with the free lysozyme model. Reconstitution with lysozyme from the complex yields a few more H-bonds, but then, side chain conformations are biased toward making favorable interactions in the native orientation.

Some of the refined artificial complexes come out with up to 12 H-bonds, and show features that are usually associated with real structures. This is illustrated in Figure 5 for the non-native Fab HyHEL5–lysozyme complex ranking 2d in Table III. Its interface involves residues 21–23, 47, 61–62, and 97–116 of lysozyme. It has only one residue in common with the real epitope (residues 41–53, 67–70, and 84). Still, it appears to fit nicely into the combining site. It forms a salt bridge and 12 H-bonds, very much like the native interface, but the partners are all different. We looked for buried polar groups which would not form H-bond in this artificial interface, and found none. Thus, it would be difficult to reject it in the absence of biochemical information on the epitope recognized by antibody HyHEL5.

There is a wealth of biochemical data in solution on complexes discussed here. The role of individual lysozyme residues in recognition by the D1.3, HyHEL5, and HyHEL10 antibodies, that of protease inhibitor residues in recognition by serine proteases, has been extensively studied. The most complete analysis was done on a large family of ovomucoid inhibitors.[40] It shows that single substitutions within the binding site, change the affinity by factors ranging between 1 and 10^4. In lysozyme, the substitution of Arg-68 by Lys lowers the affinity for antibody HyHEL5 by a factor of over 100 (see ref. 41); that of Gln-121 by His has similar effects on D1.3 binding.[42] Such effects are fairly well understood[43] and can be simulated to some extent in the computer,[44,45] assuming that the same interactions are retained except for the modified residue.

Effects—or the lack of effect—of more extensive changes are difficult to analyze. Low affinity complexes formed after genetic or chemical modification of one of the partners have not been proved to involve the same sites as the high-affinity complexes. Cases where the same binding site recognizes two very different partners are not uncommon. In ovomucoid inhibitor–protease complexes, the same inhibitor may have similar affinities for very different proteases, and conversely.[40] With antibodies, structural data are available from the R. Poljak laboratory on several complexes involving the same Fab D1.3. Its combining site recognizes two very different molecular surfaces when complexed with lysozyme[12] or with an antiidiotope.[46] This implies that the protein–protein recognition problem can have several distinct solutions. A procedure that finds some of them can be valuable, even though it does not rank them in the correct order at this point.

ACKNOWLEDGMENTS

We wish to dedicate this paper to the memory of Pr. Cyrus Levinthal. Our computational approach is largely derived from ideas that he put forward more than twenty years ago. J. Janin, who owes Cy his first introduction to scientific research, also acknowledges a long standing collaboration on the subject of protein–protein interactions with Dr. S. Wodak (Université Libre, Brussels, Belgium) and Dr. C. Chothia (MRC Laboratory of Molecular Biology, Cambridge, England). We are grateful to Pr. R. Poljak (Paris) and to Dr. W. Bode (Martinsried, Germany) for the gift of atomic coordinates prior to deposition. Valuable help from staff of the computer center at LURE, (Orsay, France) is also acknowledged.

REFERENCES

1. Janin J., Chothia C. The structure of protein-protein recognition sites. J. Biol. Chem. 256:16027–16030, 1990.
2. Chothia C., Janin J. Principles of protein-protein recognition. Nature (London) 256:705–708, 1975.
3. Davies D.R., Sheriff S., Padlan E.A. Antibody-antigen complexes. J. Biol. Chem. 263:10541–10544, 1988.
4. Mariuzza R.A., Phillips, S.E.V., Poljak R.J. The structural

basis of antigen-antibody complexes. Annu. Rev. Biophys. Chem. 16:139–159, 1987.
5. Wodak S., Janin J. Computer analysis of protein-protein interaction. J. Mol. Biol. 124:323–342, 1978.
6. Janin J., Wodak S. Reaction pathway for the quaternary structure change in hemoglobin. Biopolymers 24:509–552, 1985.
7. Wodak S., de Crombrugghe M., Janin J. Computer studies of interactions between macromolecules. Prog. Biophys. Molec. Biol. 49:29–63, 1987.
8. Levitt M. A simplified representation of protein conformation for rapid simulation of protein folding. J. Mol. Biol. 104:59–107, 1976.
9. Levinthal C., Wodak S., Kahn P., Dadivanian A. Hemoglobin interaction in sickle cell fibers. I: Theoretical approaches of the molecular contacts. Proc. Natl. Acad. Sci. U.S.A. 72:1330–1334, 1975.
10. Sheriff S., Silverton E.W., Padlan E.A., Cohen G.H., Smith-Gill S.J., Finzel B.C., Davies D.R. Three-dimensional structure of an antibody-antigen complex. Proc. Natl. Acad. Sci. U.S.A. 84:8075–8079, 1987.
11. Padlan E.A., Silverton E.W., Sheriff S., Cohen G.H., Smith-Gill S.J., Davies D.R. Structure of an antibody-antigen complex: Crystal structure of the HyHEL10 FAB-lysozyme complex. Proc. Natl. Acad. Sci. U.S.A. 86:5938–5942, 1989.
12. Amit A.G., Marriuzza R.A., Phillips S.E.V., Poljak R.J. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. Science 233:747–753, 1986.
13. Huber R., Kukla D., Bode W., Schwager P., Bartel K., Deisenhofer J., Steigemann W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystallographic refinement at 1.9 Å resolution. J. Mol. Biol. 89:73–101, 1974.
14. Bode W., Papamokos E., Musil D., Seemueller U., Fritz H. Refined 1.2 Å structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. EMBO J 4:813–818, 1986.
15. Bode W., Wei A-Z, Huber R., Meyer E., Travis J., Neumann S. X-ray crystal structure of the complex of human leukocyte elastase and the third domain of the turkey ovomucoid inhibitor. EMBO J. 5:2453–2458, 1986.
16. Bernstein F.C., Koetzle T.F., Williams G.J.B., Meyer E.F., Jr., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M. The protein data bank, a computer based archival file for macromolecular structures. J. Mol. Biol. 112:535–543, 1977.
17. Shrake A., Rupley J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J. Mol. Biol. 79:351–371, 1973.
18. Lee B.K., Richards F.M. The interpretation of protein structures. Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
19. Wodak S.J., Janin J. Analytical approximation to the accessible surface area of proteins. Proc. Natl. Acad. Sci. U.S.A. 77:1736–1740, 1980.
20. Chothia C.. Hydrophobic bonding and accessible surface area in proteins. Nature (London) 248:338–339, 1974.
21. Sharp K.A., Nicholls A., Fine R.F., Honig B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effect. Science 252:106–109, 1991.
22. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. Equation of state calculation by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.
23. Kirkpatrick S., Gelatt C.D., Jr, Vecchi M.P. Optimization by simulated annealing. Science 220:671–680, 1983.
24. van Laarhoven P.J.M., Aarts E.H.L. In: "Stimulated Annealing: Theory and applications." Dordrecht, Holland: D. Reidel, 1987.
25. Brünger A. X-PLOR Version 1.5 Manual. Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, 1987.
26. Brooks B.R., Bruccoleri R.E., Olafson B.D., States J.S., Swaminathan S., Karplus M. CHARMM. A program for macromolecular energy minimization and dynamics calculations. J. Comp. Chem. 2:187–217, 1983.
27. Diamond R. Real-space refinement of the structure of hen egg-white lysozyme. J. Mol. Biol. 82:371, 1974.
28. Marquart M., Walter J., Deisenhofer J., Bode W., Huber R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta. Crystallogr. B39:480, 1983.
29. Blow D.M., Wright C.A., Kukla D., Rühlmann A., Steigemann W., Huber R. A model for association of bovine pancreatic trypsin inhibitor with chymotrypsin and trypsin. J. Mol. Biol. 69:137–144, 1972.
30. Rühlmann A., Kukla D., Schwager P., Bartels K., Huber R. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystal structure determination and stereochemistry of the contact region. J. Mol. Biol. 77:417–436, 1973.
31. Salemme F.R. An hypothetical structure for the intermolecular electron transfer complex of cytochromes c and b5. J. Mol. Biol. 102:563–568, 1976.
32. Warwicker J. Investigating protein-protein interaction surfaces using a reduced stereochemical and electrostatic model. J. Mol. Biol. 206:381–395, 1989.
33. Greer J., Bush B.L. Macromolecular shape and surface maps by solvent exclusion. Proc. Natl. Acad. Sci. U.S.A. 75:303–307, 1978.
34. Kuntz I.D., Blaney J.M., Oatley S.J. Langridge R., Ferrin T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161:269–288, 1982.
35. Lee R.H., Rose G.D. Molecular recognition: Automatic identification of topographic surface features. Biopolymers 24:1613–1627, 1985.
36. Connolly M.L. Shape complementarity at the hemoglobin $\alpha 1\beta 1$ interface. Biopolymers 25:1229–1247, 1986.
37. Jiang F., Kim S.H. "Soft docking": Matching of molecular surface cubes. J. Mol. Biol. 219:79–102, 1991.
38. DesJarlais R.L., Sheridan R.P., Seibel G.L., Dixon J.S,. Kuntz I.D., Venkataraghavan R. Using shape complementarity as an initial screen in designing ligands for a receptor site of known three-dimensional structure. J. Med. Chem. 31:722–729, 1988.
39. Eisenberg D., McLachlan A.D. Solvation energy in protein folding and binding. Nature (London) 319:199–203, 1986.
40. Laskowski M., Jr, Kato I., Ardelt W., Cook J., Denton A., Empie M.W., Kohr W.J., Park S.J., Parks K., Schatzley B.L., Scoenberger O.L., Tashiro M., Vichot G., Whatley H.E., Wieczorek A. & Wieczorek M. Ovomucoid third domain from 100 avian species. Isolation, sequence and hypervariability of enzyme-inhibitor contact residues. Biochemistry 26:202–221, 1987.
41. Smith-Gill S.J., Wilson A.C., Potter M., Prager E.M., Feldmann R.J., Mainhart C.R. Mapping the antigenic epitope for a monoclonal antibody against lysozyme. J. Immunol. 128:314–322, 1982.
42. Harper M., Lema F., Boulot G., Poljak R.J. Antigen specificity and cross-reactivity of monoclonal anti-lysozyme antibodies. Mol. Immunol. 24:97–108, 1987.
43. Fersht A.R., Shi J.P., Knill-Jones J., Lowe D.M., Wilkinson A.J., Blow D.M., Brick P., Carter P., Waye M.M.Y., Winter G. Hydrogen bonding and biological specificity analysed by protein engineering. Nature (London) 314: 235–238, 1985.
44. Bartlett P.A., Marlowe C.K. Evaluation of intrinsic binding energy from a hydrogen bonding group in an enzyme inhibitor. Science 235:569–571, 1987.
45. Novotny J., Bruccoleri R.E., Saul F.A. On the attribution of binding energy in antigen-antibody complexes McPC603, D1.3 and HyHEL5. Biochemistry 28: 4735–4749, 1989.
46. Bentley G.A., Boulot G., Riottot M.M., Poljak R.J. Three-dimensional structure of an idiotope-anti-idiotope complex. Nature (London) 348:254–255, 1990.