ELSEVIER

Molecular recognition via face center representation of a molecular surface

Shuo Liang Lin* and Ruth Nussinov†‡

**Laboratory of Mathematical Biology, NCI-FCRF, Frederick, Maryland, United States*
†Laboratory of Mathematical Biology, SAIC, Frederick, Maryland, United States
‡Sackler Institute of Molecular Medicine, Tel Aviv University, Tel Aviv, Israel

While docking methodologies are now frequently being developed, a careful examination of the molecular surface representation, which necessarily is employed by them, is largely overlooked. There are two important aspects here that need to be addressed: how the surface representation quantifies surface complementarity, and whether a minimal representation is employed. Although complementarity is an accepted concept regarding molecular recognition, its quantification for computation is not trivial, and requires verification. A minimal representation is important because docking searches a conformational space whose extent and/or dimensionality grows quickly with the size of surface representation, making it especially costly with big molecules, imperfect interfaces, and changes of conformation that occur in binding. It is essential for a docking methodology to establish that it employs an accurate, concise molecular surface representation.

Here we employ the face center representation of molecular surface, developed by Lin et al.,[1] to investigate the complementarity of molecular interface. We study a wide variety of complexes: protein/small ligand, oligomeric chain–chain interfaces, proteinase/protein inhibitors, antibody/antigen, NMR structures, and complexes built from unbound, separately solved structures. The complementarity is examined at different levels of reduction, and hence roughness, of the surface representation, from one that describes subatomic details to a very sparse one that captures only the prominent features on the surface. Our simulation of molecular recognition indicates that in all cases, quality interface complementarity is obtained. We show that the representation is powerful in monitoring the complementarity either in its entirety, or in selected subsets that maintain a fraction of the face centers, and is capable of supporting molecular docking at high fidelity and efficiency. Furthermore, we also demonstrate that the presence of explicit hydrogens in molecular structures may not benefit docking, and that the different classes of protein complexes may hold slightly different degrees of interface complementarity.

Keywords: molecular recognition, molecular shape, surface representation, complementarity, docking

Color Plates for this article are on pages 95–97.

Address reprint requests to: S.L. Lin, NCI-FCRF, Bldg. 469, Rm. 151, Frederick, Maryland 21702.

Received 6 February 1996; accepted 26 March 1996.

INTRODUCTION

Biomolecule association, such as the binding of antibodies to antigens and of enzymes to substrates, as well as oligomer formation, involves molecular recognition. As early as half a century ago, Linus Pauling envisioned biomolecular association to be like that of die and coin, stabilized by the complementarity of surface shapes.[2] Since then abundant evidence supporting the notion has emerged from solved molecular complex structures. The concept of shape complementarity has inspired the exercise of docking, namely simulating molecular recognition by computers. To express the concept of complementarity into computable data structures for docking, it is necessary to quantify the shape of the molecular surface. Most of the docking methods that have been developed treat the molecules as rigid bodies, and abstract certain geometric features of them for the quantification, on which various docking schemes are applied. For example, Kuntz and colleagues used spheres that filled invaginations on a surface[3,4]; Connolly prescribed various forms of shape functions at local surface areas[5,6]; Janin and colleagues used spherical approximation of surface residues[7,8]; Jiang and Kim[9] used cube blocks while Bacon and Moult[10] used B-splines to approximate surfaces; and Walls and Sternberg[11] and more recently Helmer-Citterich and Tramontano[12] both used a stack of slices of the molecule, whose rims were to be matched. All these docking methods have demonstrated success in modeling selected molecular complexes, in spite of their wide diversity in accuracy and efficiency (see, e.g., Refs. 1, 4, and 12–14a). Obviously, the quality of the docking simulation is determined both by the choice of quantification and the choice of the docking scheme. How well the surface

Journal of Molecular Graphics 14:78–90, 1996

655 Avenue of the Americas, New York, NY 10010

0263-7855/96/$15.00
PII S0263-7855(96)00030-6

complementarity is materialized by a quantification imposes a limit on how well the results of docking can possibly be. From this point of view the task of a docking scheme is to realize the potential given by the quantification. It is informative and constructive to assess the quality of the quantification chosen by a docking method, in addition to assessing the method's overall outcome. While the importance of surface complementarity has long been recognized, quantitative knowledge about molecular surface is rare. Given so many different surface quantifications that have been employed, much can possibly be learned if they are properly analyzed. However, there has been no adequate examination in this regard, despite the abundance of docking method developments.

We have devised a new surface quantification, the face center representation of molecular surfaces.[1] An example of this surface representation is illustrated in Color Plate 1,[14b] depicting a human immunodeficiency virus type 1 (HIV-1) protease inhibitor. In this representation discrete points are sparsely doped at key locations of molecular surfaces for a concise depiction of the molecular shape. In the complementary areas of a bimolecular interface, the representative points between the coupling surfaces are close in space, and the surface normals at these points are fairly well aligned. We have shown that, using this representation, accurate and efficient docking is achieved with a wide variety of protein complexes.[1,13] With the surfaces represented by the sparse points, successful molecular recognition occurs when an adequate number of the points ("*a*toms") are superimposed, their associated vectors ("*b*onds") such as the surface normals and the inter-point edges are aligned, and the *in situ* surface properties such as convexity ("*c*olor") are matched. We may coin the essence of such a recognition mechanism as the "*abc* law." Obviously, by applying a docking scheme that exercises the *abc* law, accurate docking is possible when the interface is in good complementarity. Efficient docking is feasible, because the *abc* law dictates that the conformational space be sampled only where points are superimposed and vectors are aligned. The search is confined to the vicinity where complementarity can possibly exist, avoiding the common drawback of indiscriminate search. In those docking simulations, efficiency has been exceptionally high, due to the use of a sophisticated scheme based on a pattern recognition technique. The latter belongs to the geometry hashing paradigm of computer vision, which observes the *abc* law.[1,13]

In this work we examine the interface complementarity and molecular recognition supported by the face center surface representation alone, detached from any docking scheme. In light of the *abc* law, we assess the quality of surface complementarity by how well the points are matched, and by how well the surface normals are aligned in the interface of a molecular complex, particularly in the convex–concave areas. We simulate molecular recognition by computing a complex conformation from the matched points, and assess it by how well this conformation reproduces the original one. A variety of molecular complexes with different characteristics, whose structures have been solved either by X-ray crystallography or by solution nuclear magnetic resonance (NMR), are included in the examination. We also include complexes built from unbound molecules, whose structures are solved separately. While the fidelity of molecular recognition can be conveniently evaluated by the familiar root-mean-squares deviation (RMSD) of atomic positions, we notice that there is not an accepted calibration for the quality of surface shape complementarity. To fill this void we create an artificial complex with an idealized interface geometry as a reference structure, and one that bridges the gap between the ideal and the natural complexes. The complementarity of the natural complexes is gauged against that of the ideal ones, and against each other. Because the set of face centers can be tailored to represent the surfaces in variable details, and consequently may support the shape complementarity and molecular recognition differently, we conduct the examination with four differently chosen point sets to cover a wide range of surface roughness. We demonstrate that within this range, good complementarity is realized for all of the complexes by the surface representation from subatomic details to the greater features of residual sizes, and molecular recognition can be achieved with high fidelity. Besides showing the potency of the representation for accurate docking simulation, we also find that the inclusion of explicit hydrogens may not be beneficial to docking, and that surface complementarity apparently shows a slight distinction between different complex classes.

METHODS

Molecular complexes

Table 1 lists the 26 complexes being examined. The complexes are divided into seven classes: (1) model complexes artificially constructed, (2) protein/small molecules, (3) protein oligomers, (4) proteinase/protein inhibitors, (5) antibody/antigens, (6) NMR structures, and (7) model complexes built from unbound structures. All the natural complexes are obtained from the Protein Data Bank (PDB)[15]; they are cocrystal structures except for the NMR and unbound structures.

Unbound complexes were built with separately solved molecular structures, by fitting individual molecules into a known complex containing the same molecules. The fitting was a least-square optimization on the C_α positions of the interface residues, which are those having atoms within 5 Å of any atoms of the countermolecule in the known complex. The antibody moiety of hflA is an exception, which the complex copies directly from the cocrystal 2hfl.

The idealized artificial complex **(I)** was created by separating a 17-layered pyramid into two moieties, *top* and *bottom*. The pyramid consisted of balls of 2.0-Å radius, closely packed in hexagonal stacking. The separation left knob, hole, plane, and terrace areas in the interface of **I** (Color Plate 2a). Both the top and bottom moieties maintained a threefold symmetry. A molecular dynamics simulation, followed by energy minimization, distorted **I** to yield complex **II** (Color Plate 2b). The molecular dynamics simulation was performed at temperatures from 10 000 K to 0 K, cooling evenly every step in 25 steps, while the balls at the exterior of the pyramid were constrained by a harmonic potential with force constant 10 kcal/Å^2. The energy minimization was a 200-step conjugated gradient optimization. The program CHARMM[16] was used to obtain **II,** wherein the balls were given unity mass and the van der Waals energy of the CH3E atom. The RMSD between **I** and **II** is 2.31 Å.

Table 1. Summary of 26 complexes examined[a]

Name	Receptor	Ligand	Resolution (Å)
Artificial complexes			
I	Bottom half of a pyramid	Top half of the pyramid	Artificial
II	Bottom half of a disturbed pyramid	Top half of the disturbed pyramid	Artificial
Protein/small ligands			
4phv	HIV-1 protease	Inhibitor L-700,417	1.7
4mbn	Metmyoglobin	Heme	2.0
3dfr	Dihydrofolate reductase	Methotrexate	2.3
Oligomers			
2mhb	Hemoglobin β subunit	Hemoglobin α subunit	2.0
4hvp	HIV-1 protease chain B	HIV-1 protease chain A	2.1
Proteinase/protein inhibitors			
1cho	α-Chymotrypsin	Ovomucoid third domain	1.8
2ptc	β-Trypsin	Pancreatic trypsin inhibitor	1.9
1tgs	Trypsinogen	Pancreatic secretory trypsin inhibitor	1.8
2sec	Subtilisin Carlsberg	*N*-Acetyl eglin-C	1.8
4sgb	Serine proteinase B	Potato inhibitor PCI-1	2.1
4cpa	Carboxypeptidase Aα	Carboxypeptidase A inhibitor	2.5
1tec	Thermitase	Eglin-C	2.2
1cpk	Protein kinase	Protein kinase inhibitor	2.7
2kai	Kallikrein A	Pancreatic trypsin inhibitor	2.5
Antibody/antigen			
1jhl	D11.15 Fv fragment	Lysozyme	2.4
1fdl[b]	D1.3 Fab fragment	Lysozyme	2.5
2hfl[b]	HyHEL-5 Fab fragment	Lysozyme	2.5
3hfm[b]	HyHEL-10 Fab fragment	Lysozyme	3.0
1nca	NC41 Fab fragment	N9 neuraminidase	2.5
NMR structures			
2bbm	Calmodulin	Myosin light chain kinase binding domain	Avg. and opt.
2bbmH	Identical to 2bbm, except for bearing all hydrogens		
1lcc	*lac* repressor head piece	Left half of the *lac* operator	Best
Complexes built from unbound structures			
ptcA	Trypsin (2ptn)	Pancreatic trypsin inhibitor (4pti)	1.6/1.5
hf1A*	HyHEL-5 Fab fragment (2hfl)	Lysozyme (1lyz)	2.5/2.0

[a]The code names of the complexes are listed in the left column. The descriptions of the receptors and ligands are in the two middle columns. The resolutions of the structures are in the right column; NMR and artificial complexes, where resolution is not applicable, are also specified in this column. The code names are after the PDB code for the native complexes. Complexes 2bbm and 1lcc are NMR structures; the former is an averaged and energy-optimized model, while the latter is the best model. Other PDB entries are crystallographic structures. Model choA is obtained by fitting the unbound 5cha and 2ovo structures into the cocrystal 1cho. Model ptcA is obtained by fitting the unbound 2ptc and 4pti into the cocrystal 2ptc. Model hflA is obtained by fitting 1lyz into the ligand structure of the cocrystal 2hfl, while the antibody remains untouched. Complexes **I** and **II** are artificially constructed, as described in Methods.
[b]The constant domain has been cut off.

For the convenience of notation, we refer to the two moieties of a complex as "receptor" and "ligand," individually, as specified in Table 1. Usually, but not always, the receptor refers to the larger molecule.

Face center surface representation

The description of the surface representation has been elaborated previously.[1] Briefly, the molecular surface is a

mosaic of three types of faces (convex, concave, and toroidal)[17,18]; each face has a geometry center at the average coordinates of the face; projecting the geometry center onto the surface yields the *face center;* the set of the face centers is our initial surface representation. A face center acquires the label "cap," "pit," or "belt," corresponding to the type of face (convex, concave, or toroidal, respectively). The geometric features associated with a face center relevant to this study are its position, the direction of its surface normal, and the area of the face it covers.

The major steps in creating the face centers include computing the geometry centers and the projection. The geometry centers were computed by numerical integration over dense surface dots, through a procedure based on Connolly's MS-DOT program.[19] We used a density of 10 dots/Å^2 in this procedure, and a probe ball of 1.8-Å radius, which is to mimic a nonhydrogen atom of organic molecules, rather than a water. We adopted the CHARMM atomic radii for the molecules, either from the extended-atom module when there was no aliphatic hydrogen, or from the all-atom module otherwise. The computation for the projection has been described analytically, and has been incorporated into the above-described procedure. The procedure introduces uncertainties in the position of the face centers and the direction of their associated surface normals. In the worst case, the uncertainty was about 0.15 Å for position, 5° for the normals on nonhydrogen atoms, and 15° for normals on hydrogen atoms, respectively.

Table 2 enumerates for each molecule the number of caps, pits, and belts on its surface, and the number of its total and surface atoms. The latter does not include hydrogen atoms.

Reduction of critical points

Critical points refer to both the initial face centers and their descendent points derived from the purging operations described below. Purging reduces the number of critical points, resulting in a rougher surface representation. We will investigate four levels of roughness: one the initial, and three obtained from various purging sequences.

Three kinds of purging operations are used in this study: (1) selection by face type, (2) fusion of close points, and (3) selection by the extent of area coverage. Type selection keeps a subset of critical points, either the caps, the pits, or the belts, or any combination of them. The fusing operation merges the points that are closer than a given distance to each other, averaging their positions and normal directions. Area selection retains only the points that cover a surface area within a given range.

Three sequences of purging have been formulated to yield three sets of critical points. Initially, the density of the critical points is about six points per surface atom[1] (Table 2), close to the theoretical limit of one cap, two pits, and three belts for atoms close-packed on an infinite flat surface. The first roughness level results from a type selection, which keeps only the caps for the ligands and the pits for the receptors. At this level about one cap and two pits per surface atom remain in the corresponding surfaces. The second sequence performs the same type selection, followed by an area selection of the caps, and fusing the pits before their area selection. The area selection of the caps is to dismiss those whose areas are less than 5% of the probe sphere, expected to be located at dented surface areas not much more than two atoms wide. Such small areas are either shallow dents not crucial to molecular recognition, or are at the bottom of a shaft along which bigger caps have been preserved. Pit fusing combines those close to each other at subatomic distances, using 1.5 Å as the threshold. Subsequently, small pits covering areas less than 1% of the probe sphere are dismissed. Also dismissed are those with areas greater than 10% of the probe sphere, for these big pits are likely resulting from fusion across atoms, which is too widespread to preserve the shape of the area. The roughness that the second sequence leads to can be characterized by the resulting point density, about 0.5 points per surface atom. In the third sequence, only the caps and pits whose areas are less than 15 and 8% of the probe sphere are retained for the ligand and receptor, respectively. The thresholds are chosen to achieve the deepest possible purging across all of the complexes being studied, to leave as few as three critical point pairs coupled in an interface. The 15:8 ratio results in close point densities, regardless of caps or pits. For the molecules in this study, the third sequence maintains about one-quarter to one-third of the initial caps, and about one-tenth to one-sixth of the initial pits. It keeps caps and pits primarily on prominent surface features, namely protrusions and invaginations of multiatom size, respectively, while the other sequences result in a relatively even distribution of the points.

We refer to the initial roughness as level 0, and the ones resulting from the three purging sequences as levels 1 to 3.

Assessing surface complementarity and molecular recognition

We conduct the assessment using the critical points that are coupled in the complex interfaces. We pick the pairs of points that are closer than 2 Å across the interface as the *quality pairs* for the assessment, on the basis of the following finding. Examining the intermolecular nearest-neighbor distances between the face centers, we have found a common distance distribution among the complexes. The distribution rises to a sharp peak at about 1 Å, and drops to background between 2 and 4 Å. Beyond that, the distribution either tails off if the interface is small, or shows humps that can be traced to the rim areas of the interface (Figure 1). The bell-shaped peak between 1 and 2 Å arises from the interfacing face centers between the molecules. The distribution suggests that most of the complementarity information an average point can receive is condensed in the range of 2 Å. Therefore we examine all of the pairs that are coupled within 2 Å, to obtain the best signal-to-noise ratio regarding interface complementarity.

We employ three attributes of the quality pairs as indicators of interface complementarity. They include the following: the count of quality pairing, the median (value below which 50% incidents occur) distance between pairing points, and the median angle between normals. These attributes are sensitive to various aspects of the interface complementarity: the pairing count indicates the extent of the complementary areas, the distance median is best for

Table 2. Caps, pits, and belts, and number of total and surface atoms, on each molecule studied[a]

	Receptor					Ligand				
	Atoms					Atoms				
Name	Total	Surface	Caps	Pits	Belts	Total	Surface	Caps	Pits	Belts
Artificial complexes										
I	518	298	298	606	865	451	340	340	711	984
II	518	287	293	596	820	451	319	325	664	914
Protein/small ligands										
4phv	1520	639	673	1435	1877	46	44	51	104	129
4mbn	1217	479	512	1128	1464	44	44	59	120	147
3dfr	1294	576	620	1330	1741	33	32	43	72	99
Oligomers										
2mhb	1178	469	491	1065	1413	1113	419	434	969	1262
4hvp	758	393	425	904	1210	758	386	416	912	1216
Proteinase/protein inhibitors										
1cho	1749	620	647	1406	1815	400	226	236	509	664
2ptc	1629	579	605	1347	1728	454	232	248	528	695
1tgs	1646	594	622	1370	1786	416	231	249	567	736
2sec	1920	616	640	1388	1815	530	264	278	605	783
4sgb	1306	477	490	1062	1400	380	214	229	464	624
4cpa	2437	761	787	1741	2270	385	166	173	383	498
1tec	2004	654	673	1442	1876	522	250	262	566	746
1cpk	2770	939	979	2165	2816	157	115	119	255	338
2kai	1797	656	698	1547	2016	437	226	231	497	669
Antibody/antigen										
1jhl	1749	640	676	1468	1913	996	396	409	884	1173
1fdl	1729	646	671	1437	1896	1001	421	439	958	1237
2hfl	1685	613	649	1470	1920	999	408	436	946	1222
3hfm	1712	621	645	1409	1820	1001	410	439	936	1223
1nca	3321	1183	1262	2752	3574	3075	928	970	2148	2774
NMR structures										
2bbm	1164	611	698	1546	2006	210	144	156	312	428
2bbmH	2259	829	935	2114	2681	441	198	225	455	612
1lcc	399	210	220	468	621	445	301	318	678	886
Complexes built from unbound structures										
ptcA	1629	577	602	1323	1713	454	229	242	507	686
hflA	3227	1211	1287	2913	3733	1001	435	465	960	1264

[a]The face centers labeled "caps," "pits," and "belts" were generated on the molecules with hydrogens stripped, if there were any. Listed are their numbers, as well as the total number of the atoms and the number of atoms exposed on the surface, for each of the receptors and ligands. A rule of thumb indicates that there are six face centers per surface atom, in the ratio of 1:2:3 between the caps, pits, and belts, respectively.

monitoring the interface intimacy, while the angle median reveals the relative surface orientation around the pairing points. These attributes are compared among the complexes at all roughness levels, including the artificial complex **I**, which is composed with a nearly ideal interface.

We also obtain an RMSD value to indicate the quality of simulated molecular recognition, as described below. By superimposing the ligand critical points onto their receptor partners in quality pairing, we simulate the optimal docking achievable with the complementarity supported by the surface representation. We employ the unit quaternion least-square fitting algorithm[20] for the superposition, which reaches the analytic optimum in a single step. The superposition moves the ligand away from its original position, while the receptor is fixed. As the quality indicator for the simulated recognition, we calculate the positional RMSD between the moved and the original interface ligand atoms, which are those within 5 Å of any receptor atom. We include also the RMSD calculated from all ligand atoms in Tables 3–6, just to provide a reference. As pointed out by Shoichet and Kuntz,[4] the RMSD loses its statistical origin when the atoms move dependently. The all-atom RMSD is particularly marred because it reflects less the quality of interface reconstruction than the degree of "tail-wagging" in the part of the ligand far from the interface. In general, the RMSD is affected both by the number of quality pairs and their spatial arrangement. Because the latter varies rather unpredictably from complex to complex and between

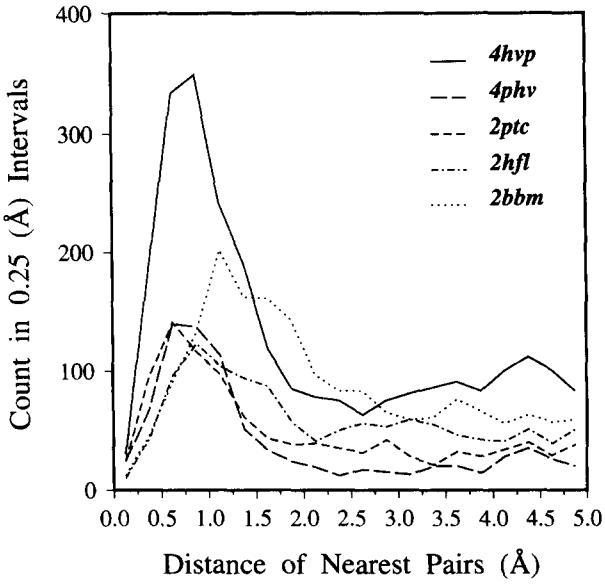

Figure 1. The distribution of nearest face center pairs between coupling molecules is shown for the complexes 4hvp, 4phv, 2ptc, 2hfl, and 2bbm, which each belong to one of the natural complex classes.

different levels of roughness, comparison between RMSD values is meaningful only if they are substantially distinct, or if the source of differences can be traced.

RESULTS AND DISCUSSION

In the following we analyze the surface representation at the four roughness levels. Tables 3–6 enumerate for the individual levels the values of the quality indicators, namely, count of quality pairing, the distance and angle medians, as well as the RMSD. Color Plate 3 summarizes them by complex classes. Color Plate 3 conveys these messages: between the roughness levels the pairing count drops by 2 to 3 decades, while the other indicators fluctuate more and more but do not decay substantially on average; at the beginning, the distance and angle medians show a slight ranking distinction corresponding to the chemical nature of the classes, which diminishes with the fluctuations in the middle, but reemerges in the end; at the last roughness level, RMSD values as poor as 3 Å appear where there are only a few quality pairs; however, most of the RMSD values are still remarkably low. In the following sections we investigate the individual roughness levels. In Color Plate 4, we demonstrate the effect of minimalizing the representation with an antibody–lysozyme complex. An illustration is presented in the section Antibody–Antigen Complex (below).

Roughness level 0: Subatomic details with all face centers

At roughness level 0 a molecular surface is represented by all the initial face centers. More specifically, on the surface a cap occupies the tip of an exposed atom, around which are a few pits and belts doping the dents and seams the atom shares with its neighbors (Color Plate 1). With about six centers per surface atom, the intimacy of an interface can be monitored by certain subatomic details. Before proceeding to the natural complexes, we examine the artificial ones first, in order to understand the interface complementarity under ideal or near-ideal conditions. The examination is aided by the trace of the distances and angles of quality pairs of complex **I**, as shown in Figure 2. The quantities cited in the following should be found in Table 3.

The idealized complex **I** has a count of quality pairing among the fewest of all complexes, but not out of their range. With a more undulating interface, complex **II** has a count about the average. In terms of interface size they are both reasonably constructed. As expected, complex **I** possesses the shortest distance median and the highest normal angle median. Nonetheless they are not perfect values, being 1.23 Å and 175°, respectively. They reflect chiefly the characteristics of the pairing with belts (Figure 2c, e, and f), because they dominate in terms of population over the caps and pits (3:1:2). On the contrary, cap–pit pairing finds a near-perfect cluster at ~0.0 Å and ~180° (Figure 2b), with nonzero distance only owing to the gap left between convex and concave faces of unmatched radii. The contrast indicates that the surface representation is able to recognize an intrinsic distinction in complementarity between different types of faces of molecular surface by definition. That is, convex faces can complement concave faces, and vice versa, while toroidal faces cannot find any tightly complementary counterpart. Therefore, molecular surfaces cannot be perfectly complementary everywhere, no matter how ideal the interface is. The surface representation observes this maxim. The representation detects perfect surface orientation by the clusters of ~180° angles, which occur in the flat areas of the interface. In other areas, it monitors the deterioration of complementarity, for example, by a trail of cap–pit pairing departing from 0 Å and 180° (Figure 2b). Interestingly, the trail declines linearly toward about 2 Å and 120°, suggesting a strong correlation in complementarity between position and orientation. The RMSD between the docked and original ligand of complex **I** is not the lowest among all the complexes. In particular, it is higher than that of the randomized complex **II.** This apparent paradox is due to the fact that the belt-involving pairs in complex **I** are mostly aligned with the pyramid axis; closing their large gaps by superposition results in a systematic shift of the ligand. This incident demonstrates that the RMSD should be interpreted with care.

For all the complexes we find extensive quality pairing in their interfaces. Most of the pairing counts fall in a relatively narrow range, consistent with the notion that the buried interfaces of natural complexes are fairly uniform in size. Exceptions are perhaps the complexes 4hvp and 2bbmH, which have greater counts, owing to a large interface between the two protease monomers of the former and the hydrogens present in the latter, which is an NMR structure.

The distance medians of the natural complexes range from 1.45 to 1.60 Å, and the angle medians range from 139° to 157°. Their subset pairing shows greater dispersion mainly parallel to the 0 Å, 180°–2 Å, 120° diagonal, indicating that the surface representation recognizes the higher undulation of the surfaces of these complexes in comparison with complex **I** (data not shown). Simulating molecular recognition results in RMSD values from 0.10 to 0.53 Å, all

Table 3. Details of complexes at roughness level 0[a]

Name	No. of pairs	Median		RMSD	
		Distance	Angle	Interface	All-atom
Artificial complexes					
I	1430	1.23	175	0.28	0.28
II	1890	1.48	156	0.18	0.18
Protein/small ligands					
4phv	1946	1.49	157	0.14	0.14
4mbn	2740	1.45	152	0.10	0.10
3dfr	1358	1.49	149	0.13	0.13
Oligomers					
2mhb	1585	1.49	149	0.39	0.51
4hvp	4556	1.49	151	0.18	0.23
Proteinase/protein inhibitors					
1cho	1673	1.50	151	0.28	0.58
2ptc	1791	1.49	154	0.15	0.50
1tgs	2221	1.50	152	0.13	0.17
2sec	1675	1.49	148	0.22	0.45
4sgb	1685	1.50	151	0.19	0.29
4cpa	1237	1.51	151	0.40	0.41
1tec	1935	1.51	151	0.21	0.42
1cpk	1925	1.52	149	0.20	0.21
2kai	1828	1.48	150	0.12	0.23
Antibody/antigen					
1jhl	1539	1.55	152	0.53	0.60
1fdl	1457	1.53	143	0.44	0.54
2hfl	1406	1.54	147	0.34	0.40
3hfm	1906	1.54	150	0.30	0.30
1nca	2160	1.52	151	0.40	0.55
NMR structures					
2bbm	1742	1.60	148	0.14	0.15
2bbmH	3676	1.60	146	0.22	0.23
1lcc	1639	1.52	151	0.22	0.24
Complexes built from unbound structures					
ptcA	1486	1.53	145	0.12	0.39
hflA	1322	1.58	139	0.33	0.49

[a]The count of quality pairing, the distance medians, the normal angle medians, and the ligand RMSD of the complexes are listed for roughness level 0. At this level all the face centers are used.

well below 1 Å. Obviously, with all of the initial face centers, the complex conformations can be reproduced with high fidelity.

The complexes modeled from unbound structures, ptcA and hflA, are inferior in three of the four indicators (namely, count of quality pairing, angle median, and RMSD) than their native counterparts, 2ptc and 2hfl, respectively, suggesting that the surface representation is able to recognize the less complementary interfaces of the unbound. We find it interesting that although quality pairing is twice as dense in the complex 2bbmH than in its twin complex 2bbm, the former does not demonstrate better complementarity. This can be understood on the basis of the only structural difference between the two, i.e., that the former has all the hydrogens, while the latter has none. The hydrogens, with radii of only 0.6–0.8 Å, cannot fit the concave faces generated by a 1.8-Å probe ball as tightly as the heavy atoms. Their presence on a surface effectively splits the faces of the one generated without hydrogens, thus dispersing the critical points and their coupling.

There appears to be a weak but discernible distinction in the range of distance medians between the complex classes. They are as follows: 1.45–1.49 Å for the protein/small molecule complexes, 1.49 Å for both of the protein oligomers, 1.48–1.52 Å for the proteinase/protein inhibitor complexes, 1.52–1.55 Å for the antibody/antigen complexes, 1.52–1.60 Å for the NMR complexes, and 1.53–1.58 Å for the complexes built from unbound molecules. The angle medians show a similar trend: 149–157° for the complexes with small ligands, 149–151° for the oligomers, 148–154° for the proteinase/inhibitors, 143–152° for the antibody/antigens, 146–151° for the NMR complexes, and 139–145° for the models from unbound structures. The trend agrees with the interface intimacy one can perceive qualitatively from molecular graphics; the fact that the values are poorer for the NMR and unbound complexes appears to be consistent with

Table 4. Details of complexes at roughness level 1[a]

Name	No. of pairs	Median		RMSD	
		Distance	Angle	Interface	All-atom
Artificial complexes					
I	127	1.08	150	0.08	0.08
II	165	1.35	152	0.12	0.12
Protein/small ligands					
4phv	125	1.46	156	0.28	0.28
4mbn	150	1.42	153	0.18	0.18
3dfr	92	1.53	147	0.20	0.20
Oligomers					
2mhb	109	1.45	151	0.42	0.52
4hvp	289	1.51	154	0.22	0.31
Proteinase/protein inhibitors					
1cho	115	1.56	146	0.24	0.38
2ptc	114	1.45	150	0.21	0.44
1tgs	149	1.56	152	0.10	0.15
2sec	105	1.45	151	0.24	0.45
4sgb	121	1.56	148	0.33	0.52
4cpa	79	1.59	147	0.50	0.50
1tec	106	1.49	154	0.37	0.75
1cpk	128	1.52	151	0.18	0.18
2kai	131	1.44	149	0.10	0.28
Antibody/antigen					
1jhl	101	1.54	151	0.44	0.63
1fdl	90	1.46	149	0.46	0.79
2hfl	90	1.42	148	0.32	0.84
3hfm	129	1.55	149	0.23	0.33
1nca	145	1.52	152	0.33	0.44
NMR structures					
2bbm	134	1.55	149	0.22	0.24
2bbmH	205	1.62	146	0.15	0.18
1lcc	100	1.42	146	0.37	0.56
Complexes built from unbound structures					
ptcA	98	1.53	141	0.32	1.02
hflA	81	1.45	143	0.47	1.29

[a]The count of quality pairing, the distance medians, the normal angle medians, and the ligand RMSD of the complexes are listed for roughness level 1. At this level the caps of the ligands and the pits of the receptors are used.

one's chemical intuition. They again demonstrate that the surface representation is capable of monitoring the shape complementarity.

Roughness level 1: Convex–concave complementarity with the cap–pit subset

As discussed above, convex and concave faces are naturally complementary, while toroidal faces have no complementary images. It follows that computer docking without the latter may be plausible. An immediate advantage of such a scheme is to have fewer objects to match, hence a smaller conformational space to search. For instance, the use of caps for the ligand and pits for the receptor at roughness level 1 reduces the critical point density to ~1 and ~2 points per surface atom, respectively. However, the loss of a large amount of subatomic details raises a question of whether a reduced representation still maintains the capacity to monitor interface complementarity and to support molecular recognition. The examination here, the results of which are shown in Table 4 and Color Plate 3, answers the question.

The response of the idealized complex **I** is easy to interpret. It essentially adopts the medians of the distance and angle seen in Figure 2b. The RMSD diminishes to 0.08 Å, as the result of removing the aforementioned systematic shift inflicted by the belts. Complex **II** demonstrates similar behavior with smaller changes, indicating a shape dependence of the impact of surface reduction.

For the native complexes, the counts of quality pairing decrease by an order of magnitude, in comparison with the full face center representation. The decrease is fairly uniform across the complexes. In particular, the counts of the unbound complexes, ptcA and hflA, are still less than their native counterparts, 2ptc and 2hfl, respectively, indicating that the representation still maintains the ability to distinguish their different complementarity. No definite trend of

Table 5. Details of complexes at roughness level 2[a]

Name	No. of pairs	Median		RMSD	
		Distance	Angle	Interface	All-atom
Artificial complexes					
I	94	1.04	153	0.12	0.17
II	43	1.36	154	0.16	0.16
Protein/small ligands					
4phv	15	1.62	157	0.54	0.54
4mbn	7	1.16	152	0.53	0.53
3dfr	5	1.08	142	0.65	0.65
Oligomers					
2mhb	23	1.40	150	0.42	0.51
4hvp	21	1.39	147	0.40	0.64
Proteinase/protein inhibitors					
1cho	13	1.49	150	0.56	1.21
2ptc	22	1.56	139	0.32	0.66
1tgs	16	1.27	157	0.57	0.99
2sec	12	1.20	159	0.45	0.36
4sgb	14	1.58	144	0.55	0.68
4cpa	15	1.76	133	0.88	0.98
1tec	14	1.64	155	0.63	0.68
1cpk	21	1.34	157	0.31	0.34
2kai	26	1.54	143	0.31	0.36
Antibody/antigen					
1jhl	14	1.05	154	0.48	0.83
1fdl	11	1.72	146	0.58	0.65
2hfl	15	1.20	143	0.52	1.45
3hfm	19	1.59	148	0.66	1.21
1nca	19	1.57	150	0.46	0.71
NMR structures					
2bbm	24	1.28	150	0.56	0.56
2bbmH	21	1.63	146	0.58	0.67
1lcc	13	1.60	139	0.62	0.84
Complexes built from unbound structures					
ptcA	19	1.45	131	0.41	0.46
hflA	16	1.43	149	0.77	1.64

[a]The count of quality pairing, the distance medians, the normal angle medians, and the ligand RMSD of the complexes are listed for roughness level 2. At this level the caps of the ligands and the pits of the receptors have been purged to a density of about 0.5 points per surface atom.

improvement or decay is found in terms of other quality indicators. The fluctuation of changes within individual complex classes causes the slight class distinction observed before to diminish, although not completely disappear. The overall ranges of the quality indicators, 1.42 to 1.62 Å for distance median, 141 to 156° for angle median, and 0.10 to 0.50 for RMSD, remain comparable to those at level 0. Obviously, the surface representation maintains its high capacity in monitoring interface complementarity and in supporting molecular recognition, even with the substantial critical point reduction performed at this level.

Roughness level 2: Preserving only the atomic details

At level 1 the densities of caps and pits are unbalanced, ~1 and ~2 per surface atom, respectively. At the present level they are both purged to a density close to 0.5. Accordingly, the counts of quality pairing decrease roughly by another order of magnitude for the natural complexes, as shown in Table 5. The fluctuating impact of roughing the surfaces is evident on their distance and angle medians. They now fall in wider ranges: 1.05 to 1.76 Å for distances, and 131 to 159° for angles. Class distinction is breached at this level, as a result of the fluctuations. All RMSD values but one decay from the previous level. However, remarkably, none of them are yet to exceed 1 Å. Note that this purging sequence has been used in real docking simulations, resulting in all-atom RMSD values mostly below 1 Å, and none exceeding 2 Å.[1,13] (Here the all-atom RMSD values range from 0.36 to 1.64 Å for the natural complexes.)

The count of quality pairing for complex **I** reduces from 127 only to 94, because few of its rather uniform faces fall outside the range of selection. Although the mild count reduction seen by complex **II** confirms its intermediate status, it becomes dubious how to transfer the observations on

Table 6. Details of complexes at roughness level 3[a]

Name	No. of pairs	Median		RMSD	
		Distance	Angle	Interface	All-atom
Artificial complexes					
I	0	–	–	–	–
II	37	1.00	152	0.47	0.47
Protein/small ligands					
4phv	8	1.02	155	0.59	0.59
4mbn	11	1.27	155	0.48	0.48
3dfr	6	1.44	145	0.48	0.48
Oligomers					
2mhb	9	1.11	162	1.13	2.69
4hvp	14	1.45	157	0.57	0.80
Proteinase/protein inhibitors					
1cho	4	1.55	156	3.35	7.39
2ptc	8	1.52	157	1.13	2.17
1tgs	14	1.08	151	0.28	0.36
2sec	11	1.22	159	0.94	1.52
4sgb	8	1.56	154	0.92	1.77
4cpa	6	1.79	150	1.35	2.25
1tec	7	1.12	156	0.53	1.02
1cpk	12	1.57	154	0.51	0.52
2kai	11	1.68	149	0.34	0.39
Antibody/antigen					
1jhl	3	1.72	156	3.13	5.97
1fdl	6	1.82	156	1.32	2.44
2hfl	7	1.41	158	0.73	1.57
3hfm	4	1.62	167	1.74	3.66
1nca	6	1.76	151	0.66	0.80
NMR structures					
2bbm	14	1.64	158	0.57	0.67
2bbmH	6	1.81	152	1.10	1.06
1lcc	12	1.57	151	0.82	1.16
Complexes built from unbound structures					
ptcA	9	1.51	148	1.05	2.07
hflA	5	1.23	148	0.83	2.19

[a]The count of quality pairing, the distance medians, the normal angle medians, and the ligand RMSD of the complexes are listed for roughness level 3. At this level the caps of the ligands and the pits of the receptors have been purged, so that the remaining caps are chiefly on the surface protrusions while the remaining pits are chiefly on the invaginations.

these artificial complexes to the natural complexes, and we do not refer to them further here.

Roughness level 3: Prominent multiatomic surface features

On the surfaces of molecules there exist protrusions and invaginations of multiatomic size, in about the volumes of amino acid residues. Large convex faces are often located in the protrusions, or "knobs," while the large concave faces are often located in the invaginations, or "holes." The reason for the former is that the surface atoms are exposed more while extruding. The same argument applies to the latter, if we view the concave faces as being "exposed inward" toward the interior of the molecule. By using more selective area criteria, such as those adopted at this roughness level, we can represent molecular surfaces with critical points doping the bigger surface features. Depending on the frequency of such features, the number of retained critical points vary widely between molecules, and so does the count of quality pairing. In general, the points are not evenly positioned on the surface. Table 6 shows the results obtained from such a rough surface representation.

It is immediately apparent that RMSD values as poor as 3 Å occur. However, all the RMSD values greater than 1.5 Å are associated with three and four counts of quality pairing, which impose the lowest or next to lowest constraints for computing a docking conformation. More impressively, most of the RMSD values remain near 1 Å or lower, even with low counts of quality pairing.

Interestingly, class distinction in distance median appears to have resurrected at this level. For the native complexes, the distance medians range from 1.02 to 1.44 Å for those with small ligands, 1.11 to 1.45 Å for the oligomers, 1.08 to

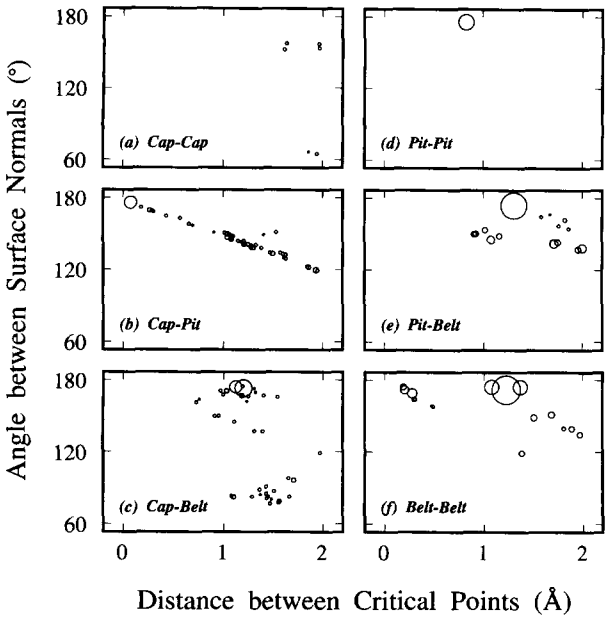

*Figure 2. Quality pairing in the interface of complex **I** is demonstrated by the distance–angle plots. (a) to (f) show the pairing subsets between the top and bottom moieties of the artificial complex. The subsets pit–cap, belt–cap, and belt–pit are not shown, which are similar to the subsets cap–pit, cap–belt, and pit–belt, respectively. The center of a circle marks the distance and normal angle of a pair; the area of the circle is proportional to the number of pairs that are degenerate with the pair.*

1.79 Å for the proteinase/inhibitors (mostly below 1.6 Å except for two), 1.41 to 1.82 Å for the antibody/antigens (mostly above 1.6 Å except for one), and 1.57 to 1.81 Å for the NMR complexes, wherein the all-hydrogen 2bbmH holds the worst value. The values are more polarized than they are at level 0, where class distinction was first observed. For the best two classes, the distance medians are even shorter. On the contrary, the last two classes demonstrate longer distance medians, with only one exception on 2hfl. In between, the enzyme/protein inhibitors show both longer and shorter distance medians. Chemical intuition would agree that it might be better to monitor shape complementarity in surface protrusions and invaginations, if key in lock-type interactions are essential to molecular recognition. This idea is consistent with the observed greater class distinction in distance median. The angle medians of these complexes are generally improved over what they are at level 0, while class distinction is not as clear. The latter may be due to the fact that surface orientation fluctuates more at the areas of higher undulation. The worst class in terms of angle median is the protein/small ligands, whose interfaces often contain larger rim regions, where the surfaces curl apart and the normals are aligned poorly.

The unbound complexes are intriguing. They improve in both the distance and angle medians over level 0. The antibody/antigen model hflA improves considerably, while the trypsin/inhibitor model ptcA improves moderately. Because their native counterparts, 2hfl and 2ptc, show improvements in a similar fashion between the two levels, the behavior of the unbound models can be understood if the sizeable features in their interfaces remain similar to those in the natural complexes. This is true for hflA and 2hfl, because we have used the same antibody in both complexes (Table 1), and the lysozyme changes little between being bound and unbound,[21] leaving quality pairs in the same epitope–paratope combination areas (see the next section) in both the model and the native complex. However, the apparent similarity between the trypsin/inhibitor model and the native complex is the result of composite factors. The major features in their interface are a binding "wedge" of the inhibitor that fits into a "groove" of the trypsin, on which a "key" (Lys-15) inserts into the specific pocket of the trypsin. The wedge and the groove in the model maintain the overall shape as in the native complex, leaving a significant number of the quality pairs intact. The stems of two arginines located at opposite ends of the wedge change conformations in the model, chiefly by rotating each around a single bond. These rotations result in loss of quality pairs. Changes also occur to Lys-15 of the inhibitor and to Glu-174 of the trypsin, which rotate in tandem, breaching the coupling in the specific pocket while finding new complementarity between themselves. In terms of net effect, the complementarity around the big features in the interface of the model is still remarkable. Interestingly, if one could reverse these changes, one might find a stepwise docking process of the free inhibitor onto the trypsin, which would involve a tentative probing with the inhibitor's wedge, a rotation of the two flanking arginines into anchoring position, and the eventual turning of the inhibitor lysine into the key-in-lock position in the specific pocket with the assistance of the trypsin glutamine.

Antibody–antigen complex: an illustration

To visualize the effect of representing a molecular surface at various roughness levels, we illustrate the quality pairs in the interface of 2hfl, an antibody/lysozyme complex, in Color Plate 4. At level 0, where the surfaces are doped with all the initial face centers, the quality pairs cover thoroughly the areas where the surfaces are in intimate complementarity, roughly in an S-shape disposition (Color Plate 4a). There are no quality pairs found on the bubble at the lower right corner of the interface, which could possibly house water molecules (the original structural determination of this complex did not place interface waters,[21] while later discussions mentioned two in the interface[22]). At level 1, only the caps of the lysozyme and the pits of the antibody are allowed on the surfaces (Color Plate 4b). The quality pairs are now more clearly coupled along the knob-and-hole and ridge-and-groove areas. The normals at the coupled points are capable of depicting the undulation of the surfaces. Most of the paired normals are fairly parallel, with a few poor ones at the lower right corner and the left edge of the interface, where the surfaces part abruptly. At level 2, pairing density is reduced by more than an order of magnitude from level 1 (Color Plate 4c). However, it is obvious that the remaining, sparse pairs still cover the entire interface. They couple quite well both in position and in normal orientation, except for one or two along the left edge of the interface. At level 3, the deep purge leaves only five and seven critical points that can pair with each other on the

lysozyme and the antibody, respectively (Color Plate 4d). The few surviving quality pairs have covered the major contact areas between the lysozyme and the antibody. They include the central ridge of the lysozyme epitope, which consists of Arg-45 and the "critical" residue Arg-68, the residues Thr-43 and Pro-70 on either ends of the epitope, and five of the six CDRs (complementarity-determining regions) of the antibody.[21] The coupling of the quality pairs preserves the intimate complementarity at the spots where they stand, as can be judged by the closeness of the points and the parallelism of the normals, except for a disparity of normal orientation on the curling left edge of the interface. They are able to support an excellent docking simulation (Table 6).

CONCLUSION

Simulating molecular recognition by geometry requires quantification of the shape of molecular surfaces. The quantification constitutes one of the basic elements of a docking method: its data structure. The other element is the docking scheme. The capacity of the data structure in supporting surface complementarity underlies the fidelity of the simulation. Analyzing this capacity is therefore important in understanding the potency of the quantification the method adopts, and in turn the potency of the method itself. This would have been especially vital for those methods that have not yet succeeded in high docking accuracy. However, this analysis has been omitted so far.

Here we have demonstrated that such an analysis is feasible. To cope with the lack of an objective gauge for quantitative measurement of complementarity, we have devised artificial complexes whose ideal complementarity can be directly analyzed, and complexes that can bridge the understanding of the idealized complexes and the understanding of the realistic ones. The quantification that we developed, the face center surface representation, is particularly suitable for docking schemes that observe the *abc* law, a strategy for accurate and efficient docking characterized by selective sampling of the conformational space where local complementarity exists between points and vectors. We choose the superposition of the points and the alignment of their associated surface normals for the analysis, so that the results are intimately relevant to actual docking. We also simulate the docking without the use of any actual docking scheme, to assess detachedly the potency of the surface representation toward high-fidelity molecular recognition. For a variety of natural and artificial protein complexes, we have shown that the surface representation is highly capable of monitoring the shape complementarity of their interfaces and supporting accurate docking simulation, within a wide range of surface roughness. This assessment has been reinforced by actual protein docking, which has achieved high accuracy and efficiency.[1,13]

With the analysis on the critical points and their normals we learn that in the interface between associated molecules, complementarity in the coin-and-die, or key-in-lock fashion begins with interlaced atoms, and extends to the tips and bottoms of greater knobs and holes of amino acid sizes. The complementarity can be expressed as the face center overlapping between the extruding atoms and their accommodating valleys, where the surface orientations are approximately antiparallel, as observed by the normals at the centers. The complementarity in the toroidal areas is questionable. Also, this expression of the complementarity is weakened if all explicit hydrogens, which fracture the faces, are included; docking using our data structure would not benefit from such a maneuver. Instead, it is sufficient to use extended atom models to include hydrogens implicitly. We also find that distinction in complementarity seems to exist between protein classes. The distinction is more apparent when the surfaces are doped more densely, or when the complementarity is monitored only at the sizeable knobs and holes. This observation appears to be consistent with the chemical properties of the classes, and agrees with the finding of Lawrence and Colman[23] that antibody/antigen complexes were poorer in complementarity than proteinase/inhibitor complexes and protein oligomers. However, we both have detected only slight differences. A more sensitive probe for the class distinction may be the distance distribution function of the nearest face center pairs covering longer ranges, as shown in Figure 1; we have cut the coverage at 2 Å in this analysis in order to take advantage of the high-quality part of interfacing, as mentioned above. Nonetheless, our results suggest that protein classes are distinct in residue intercalation in the interface.

We believe that the molecular surface deserves more interest as a scientific entity, beyond that of a graphical convenience. While the importance of surface shape in molecular recognition is undeniable, so far perhaps the only surface quantity that has received decent attention is the surface area. A hurdle appears to be the lack of other analyzable quantities. However, a reservoir of surface quantifications actually exists in many docking methods. They are extracted from a rich variety of surface features. A great opportunity exists to expand our knowledge of molecular surface substantially, if they are properly analyzed. In practical terms, an analysis of the quantification of a docking method, detached from the docking scheme, should shed light on the potency of the method, and on the possibilities and locations to make improvements. As automatic docking methods are becoming a popular tool in the investigation of molecular interactions and in rational drug design, such analyses may be timely.

ACKNOWLEDGMENTS

We thank Drs. Jacob Maizel and Haim Wolfson for helpful discussions, encouragement, and interest. We thank the personnel at the Frederick Cancer Research and Development Center for their assistance. The research of R. Nussinov has been sponsored by the National Cancer Institute, DHHS, under Contract No. 1-CO-74102 with Scientific Application International Co., and in part by Grant No. 91-00219 from the U.S.–Israel Binational Science Foundation (BSF) (Jerusalem, Israel) and by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences. The contents of this publication do not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

REFERENCES

1 Lin, S.L., Nussinov, R., Fischer, D., and Wolfson, H.L. Molecular surface representations by sparse critical points. *Proteins* 1994, **18,** 94–101

2 Pauling, L. The nature of the intermolecular forces operative in biological processes. *Science* 1940, **92,** 77

3 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 1982, **161,** 269–288

4 Shoichet, B.K. and Kuntz, I.D. Protein docking and complementarity. *J. Mol. Biol.* 1991, **221,** 79–102

5 Connolly, M.L. Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface. *Biopolymers* 1986, **25,** 1229–1247

6 Connolly, M.L. Shape distributions of protein topography. *Biopolymers* 1992, **32,** 1215–1236

7 Cherfils, J., Duquerroy, S., and Janin, J. Protein–protein recognition analyzed by docking simulations. *Proteins* 1991, **11,** 271–280

8 Cherfils, J., Bizebard, T., Knossow, M., and Janin, J. Rigid-body docking with mutant constraints of influenza hemagglutinin with antibody HC19. *Proteins* 1994, **18,** 8–18

9 Jiang, F. and Kim, S.-H. "Soft docking": Matching of molecular surface cubes. *J. Mol. Biol.* 1991, **219,** 79–102

10 Bacon, D.J. and Moult, J. Docking by least-squares fitting of molecular surface patterns. *J. Mol. Biol.* 1992, **225,** 849–858

11 Walls, P.H. and Sternberg, J.E. New algorithm to model protein–protein recognition based on surface complementarity; applications to antibody–antigen docking. *J. Mol. Biol.* 1992, **228,** 227–297

12 Helmer-Citterich, M. and Tramontano, A. PUZZLE: A new method for automated protein docking based on surface shape complementarity. *J. Mol. Biol.* 1994, **235,** 1021–1031

13 Fischer, D., Lin, S.L., Wolfson, H.J., and Nussinov, R. A geometry-based suite of molecular docking processes. *J. Mol. Biol.* 1995, **248,** 459–477

14a Norel, R., Lin, S.L., Wolfson, H.J., and Nussinov, R. Molecular surface complementarity at protein–protein interfaces: The critical role played by surface normals at well-placed, sparse points in docking. *J. Mol. Biol.* 1995, **252,** 263–273

14b Bone, R., Vacca, J.P., Anderson, P.S., and Holloway, M.K. X-Ray crystal structure of the HIV protease complex with L-700,417, an inhibitor with pseudo C_2 symmetry. *J. Am. Chem. Soc.* 1991, **113,** 9382–9384

15 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542

16 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 1983, **4,** 187–217

17 Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176

18 Connolly, M.L. Molecular surface-dot (MS QCPE 429). *QCPE Bull.* 1981, **1,** 75

19 Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 1983, **16,** 548–558

20 Horn, B.K.P. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am.* 1987, **A4,** 629–642

21 Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C., and Davies, D.R. Three-dimensional structure of an antibody/antigen complex. *Proc. Natl. Acad. Sci. U.S.A.* 1987, **84,** 8075–8079

22 Davies, D.R. and Padlan, E.A. Antibody–antigen complexes. *Annu. Rev. Biochem.* 1990, **59,** 439–473

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