# ESCHER: A New Docking Procedure Applied to the Reconstruction of Protein Tertiary Structure

**G. Ausiello, G. Cesareni, and M. Helmer-Citterich***
*Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica, 00133 Roma, Italy*

*ABSTRACT* **Evaluation of Surface Complementarity, Hydrogen bonding, and Electrostatic interaction in molecular Recognition (ESCHER) is a new docking procedure consisting of three modules that work in series. The first module evaluates the geometric complementarity and produces a set of rough solutions for the docking problem. The second module identifies molecular collisions within those solutions, and the third evaluates their electrostatic complementarity. We describe the algorithm and its application to the docking of cocrystallized protein domains and unbound components of protein-protein complexes. Furthermore, ESCHER has been applied to the reassociation of secondary and supersecondary structure elements. The possibility of applying a docking method to the problem of protein structure prediction is discussed. Proteins 28: 556–567, 1997.** © **1997 Wiley-Liss, Inc.**

**Key words: molecular recognition; automated docking; protein domains; secondary structure elements**

## INTRODUCTION

Numerous automated methods for the prediction of protein complex formation have been developed.[1,2] They are generally based on the rigid body approximation[3–5]: some rely on a geometric criterion and on a simplified representation of the protein surface, "soft docking" methods,[6–8,10] whereas others combine a shape complementarity search with subsequent energy refinement.[11,12] Many algorithms can successfully reconstruct a protein complex starting from the experimentally determined conformation of the components in the bound state. In some cases, some success, albeit with lower accuracy, has been obtained by docking elements whose structures have been determined separately.[8,10,13]

In the rigid body approach, the position and orientation of a protein is completely described by three translational and three rotational degrees of freedom. This approach is, therefore, widely used in automated methods, when molecular flexibility is not expected to affect much the geometry of the interacting structures.

On the other hand, when the docking problem involves the recognition between a rigid protein and a flexible ligand, the rigid body approximation is not applicable, and more conformational degrees of freedom have to be taken into account.[14,15] The basic principles underlying rigid body and flexible docking are common, but they generally require different computational techniques.

We explore here the possibility of applying a new soft docking rigid body method to components of intermediate size when compared with the "classical" targets of the docking methodologies discussed above. We chose to dock protein components, such as protein domains and secondary and supersecondary structures, and to explore the limits of the rigid body approximation when applied to a series of test cases displaying different sizes, structural complexities, and root-mean-square (rms) deviations from the structures of the protein in their bound conformations. The ultimate goal of this approach is the assembly of a whole protein tertiary structure starting from its secondary structure elements.

The reconstruction of a protein structure starting from its secondary structure elements has already been tried on a very small number of proteins, generally relying on knowledge-based potentials,[16–18] as well as on other methods.[19–21] To tackle this problem, a novel docking program was used as a tool. To this end, we have substantially modified the PUZZLE[8] procedure to achieve a more accurate description of the surface of small proteins and more effective matching and scoring criteria. Furthermore, the new program, ESCHER, takes advantage of two filter modules that reject solutions resulting in steric (BUMPS) or electrostatic (CHARGES) clashes. The addition of these modules allows less stringent criteria for geometric complementarity to be considered and, consequently, contributes to a further "softening" of the docking procedure.

*Abbreviations:* PDB, Protein Data Bank; rmsd, root-mean-square deviation

*Correspondence to: Manuela Helmer-Citterich, Department of Biology, University of Rome "Tor Vergata," via della Ricerca Scientifica e Tecnologica 00133 Rome, Italy.
E-mail: manuela@iris.bio.utovrm.it

Received 22 October 1996; Accepted 3 April 1997

ESCHER was applied to a series of two-domain proteins, by separating the domains and by attempting to reconstruct the native protein from the two-domain crystallographic structures, and to protein complexes starting from their unbound components. Furthermore, we defined a limited set of test cases in which the docking partners were represented by secondary and supersecondary structures separated from their protein. The docking experiments were performed with crystallographic structures and structural models. This was done to assess whether and how minor deviations from the idealized secondary structure backbone conformation or lack of knowledge of the exact side-chain conformation could affect the docking results.

## METHODS

### Overview

ESCHER is a rigid docking program, written in C language, assembled from three modules that work in series: SHAPES, BUMPS, and CHARGES. The first module takes as input the solvent-accessible surface of the target and probe peptides and provides a list of 500 solutions, ranked on the basis of geometric complementarity. Each solution is identified by the translations and rotations to be applied to the probe to match the target. The second module, BUMPS, analyzes the molecular collisions within the solution complexes, and CHARGES evaluates their electrostatic complementarity. All the solutions generated by SHAPES are subjected to the BUMPS filter and grouped together whenever their deviation from the member of the cluster with the highest geometric complementarity is lower than 2.5 Å. Finally, each group is represented by the solution with the best CHARGES value.

### SHAPES Module

SHAPES analyzes two peptide structures, evaluates their complementarity in all possible orientations, and proposes a number of docking solutions, each identified by the $T_i$ translations and the $R_i$ rotations to be applied to the second element (the probe) with respect to the first element (the target). Each structural element is cut in parallel slices, and each slice is described as a polygon with sides of identical length. The polygons representing the target are compared with the probe polygons, and regions of complementarity are evaluated. A complete search in the rotation space is done by rotating the probe in all possible orientations with respect to the target.

The flowchart of the geometric module is reported in Figure 1. Each box in the flowchart (b-1 to b-7) corresponds to one paragraph in the following detailed SHAPES description.

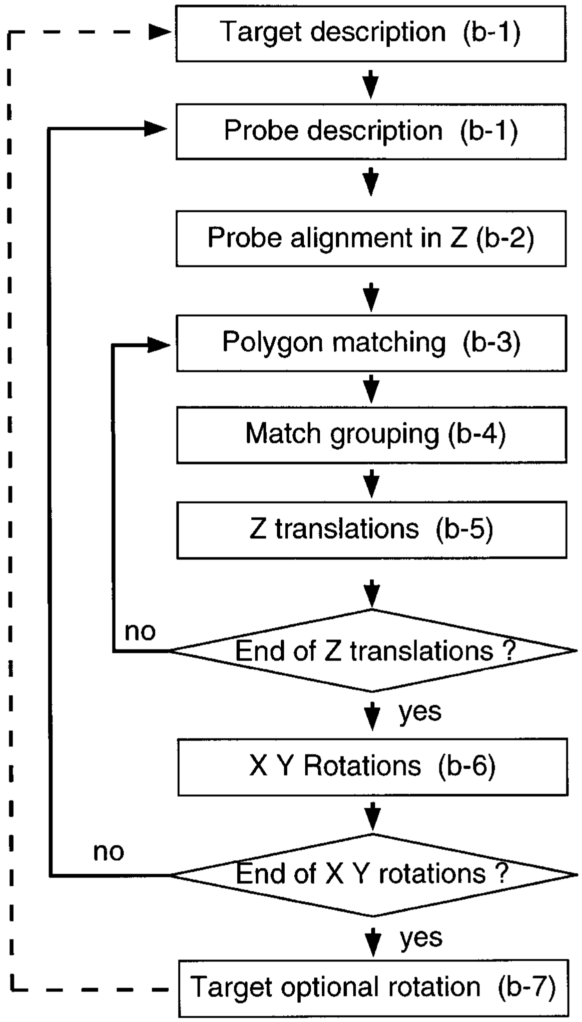

Fig. 1. Flowchart of ESCHER geometric module. Each box in the flowchart corresponds to a paragraph in Methods.

#### *(b-1) Structural elements description*

The solvent-accessible surface[22,23] of each protein structural element is generated from its atomic coordinates by using a probe radius of 1.8 Å (a density of 10 dots/Å$^2$). Each surface is cut in parallel slices 1.5 Å thick, orthogonal to the Z axis (Fig. 2a). Each slice is transformed into a polygon (a few hundred of vertices) by an "intelligent contouring" algorithm (Fig. 2b,c,d).

#### *(b-2) Starting position in Z translations*

The probe polygons are translated along the Z axis until the probe top polygon is at the level of the target bottom polygon (Fig. 3a).

#### *(b-3) Polygon matching*

Each polygon of the target can now be compared with the corresponding probe polygon belonging to

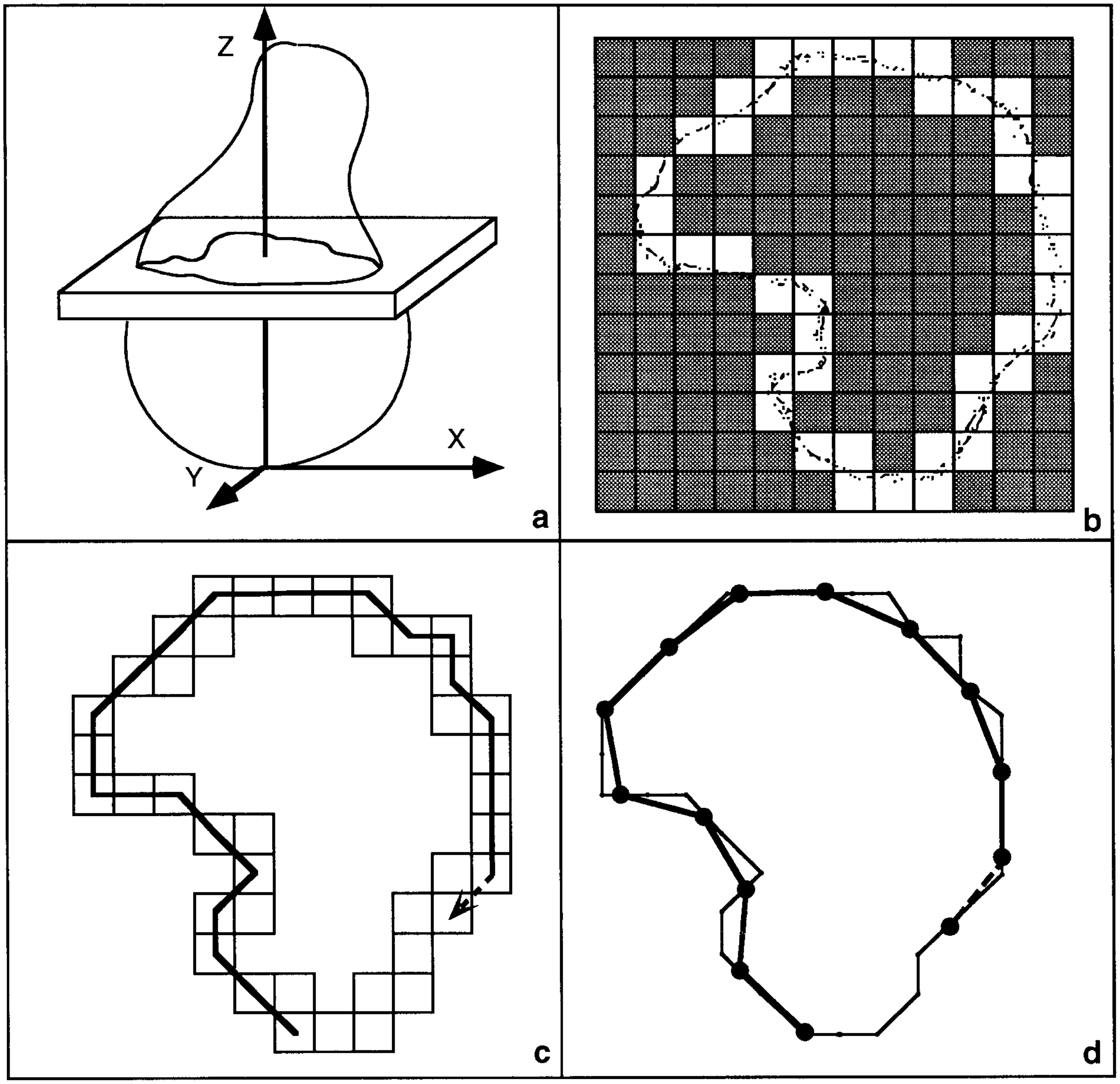

Fig. 2. Schematic representation of four steps needed to transform a three-dimensional structure in a set of polygons. **a:** A section orthogonal to the Z axis is cut on the structural element solvent accessible surface. **b:** The grid squares are defined filled (in white) or empty (in gray) whether they contain surface dots. **c:** A first polygonal line is drawn joining the center of the external filled squares. **d:** A second polygon with even sides (shown in bold) is drawn over the first polygon.

the same Z plane. Each one of the *m* probe polygon sides is superposed to each one of the *n* corresponding target polygon sides, by applying the necessary translations (Tx and Ty) and rotation (Rz). A measure of geometric complementarity ("complementarity score") is associated to each one of the *m*x*n* different superpositions and to the corresponding *Rz*, *Tx* and *Ty* transformation (Fig. 3b); the complementarity score is the number of corresponding consecutive polygon vertices whose distance is lower than 1.6 Å (Fig. 3c). Whenever a small cavity is present in the "interface" between two polygons, the complementarity of the preceding or of the following polygon sides still can contribute to the geometric score.

### *(b-4) Match grouping*

SHAPES solutions are generated by grouping the polygon matches coming from different Z planes but representing the same complementary region.

To group polygon matches and define a complementary region extending through different Z planes, a

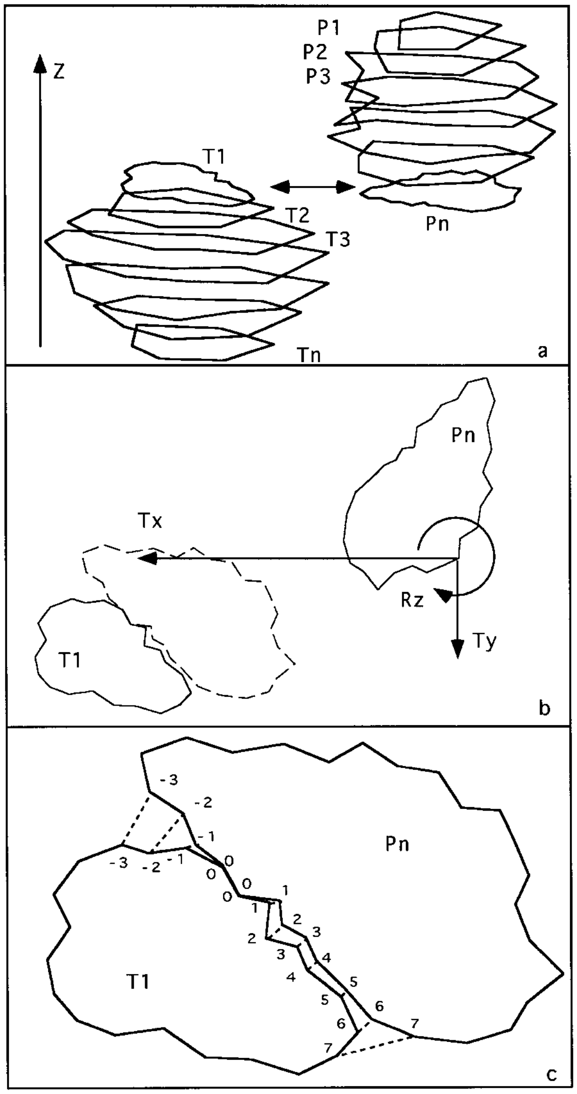

Fig. 3. **a:** The target and probe structural elements are described as $T_n$ and $P_n$ polygon sets, respectively, and are translated in the starting position along the Z axis. The probe bottom polygon ($P_n$) is aligned with the top target polygon ($T_1$). **b:** The three values that define one polygon match between $P_n$ and $T_1$ are represented: the two X and Y translations (Tx and Ty) and the rotation around the Z axis (Rz). **c:** The vertices of the superposed polygon sides are numbered with 0, and distances between the corresponding vertices on the two polygons are depicted with dotted lines. The complementarity value is the number of consecutive sides whose corresponding vertices are situated at a distance lower than 1.6 Å.

three-dimensional space is defined in which a polygon match is represented by a dot with the corresponding Tx, Ty, and Rz coordinates. In this space, we define a grouping grid with a cell size of 3 Å along the Tx and Ty axes and 12° along the Rz axis. Whenever a new polygon match is found, its complementarity value is added into the corresponding element of the grouping grid.

At the end of the match grouping step, the indexes of the grid element that scored the highest complementarity value represent the Tx, Ty, and Rz transformations to apply to the probe to correctly match the highest number of polygon pairs. These Tx, Ty, and Rz values, together with the Tz, Rx, and Ry explored stepwise as explained below, define a solution complex, associated with the complementarity value stored into its grouping matrix element.

### *(b-5) Z Translations*

All the polygons of the probe are translated by $-1.5$ Å along the Z axis, thus shifting the pairs of polygons that are matched in the subsequent "polygon matching" phase. This corresponds to a stepwise exploration of the Tz degree of freedom, where $\Delta\text{Tz} = 1.5$ Å. This step is executed until the probe top polygon reaches the level of the target bottom polygon. All the probe polygons are therefore compared with the target protein polygons.

### *(b-6) X-Y rotations*

In this step, the probe is rotated around the X and Y axis to analyze all the possible orientations with respect to the target.

Rx and Ry are explored stepwise: $-180° \leq \text{Rx} < 180°$, $-90° \leq \text{Ry} < 90°$, with $\Delta\text{Rx} = \Delta\text{Ry} = \Delta R$. This is equivalent to exerting a complete search in the space of all possible target-probe relative orientations, by rotating one of the two peptides by $\Delta R$ steps around the X and Y axis. In the docking experiments described, $\Delta R = 10°$. This value can be changed, when necessary, according to the probe dimensions. The complete exploration of the rotation space around the Z axis (Rz) is exerted continuously within the polygon pairs comparison. At the end of this step, the probe is described into a new set of polygons.

### *(b-7) Target orientation*

The cylindrical symmetry inherent to the kind of description adopted is very convenient to transform a three-dimensional surface matching problem into a simplified two-dimensional polygon comparison but offers a very poor description of the structural element poles.

One of the two elements (the probe) is rotated in all possible orientations and is therefore optimally described at least once. The problem may concern the target element, which remains in the same orientation during the whole procedure. One may overcome this problem describing the target element in more than one set of parallel polygons and repeating the comparison algorithm on the minimal number of polygon sets which one considers necessary. Cases studied to address this problem indicate that, in general, two orthogonal sets of polygons are sufficient to detect a region of geometric complementarity in at least one of the target orientations tested (data not shown).

In all of the test cases discussed below, the target has been described only once in an optimal orientation (with the interaction site parallel to the Z axis). We also verified that the results of the procedure are essentially independent of the orientation of the target in the reference system (a target rotation around the X or Y axis of up to 45° does not affect the results).

## BUMPS Module

SHAPES module evaluates complementarity on the basis of distances, which are intrinsically positive numbers. Consequently, it cannot distinguish between polygon perimeters that diverge or interpenetrate following or preceding a complementary region. Furthermore, a docking solution with a good "local" match may be accompanied by a steric clash in another part of the proposed complex. The solution list of the geometric module may, therefore, include sterically impossible structures, whose components share regions of shape complementarity but also overlap somewhere else.

The BUMPS module analyzes the distance between each atom of the target structural element and each atom of the probe in the complexes proposed by SHAPES and evaluates a "bumps" value.

$$BUMPS = \sum_a \sum_b B_{ab}$$

where

$$B_{ab} = \begin{cases} 0 & \text{if} \quad R_{ab} > R_a + R_b \\ R_a + R_b - R_{ab} & \text{if} \quad R_{ab} \leq R_a + R_b \end{cases}$$

$R_a$ and $R_b$ are the Van der Waals radii of the atoms of the target and of the probe, and $R_{ab}$ is the distance between the atoms.

A complex proposed by the SHAPES module is discarded if the "bumps" value exceeds a certain threshold. The threshold may vary for different docking tests: it should be set high when docking models of proteins that have been crystallized separately, because conformational changes may take place on binding. On the other hand, the threshold could be decreased when docking models derived from cocrystallized structures. The threshold was kept uniformly at 50 Å for all docking experiments (high stringency), except for those with the *common* model structures (110 Å, low stringency). In the analysis of the molecular collisions within models of cocrystallized complexes, one should expect in principle a low or very low bump value. In our procedure, however, we have always taken into account the intrinsic resolution of the SHAPES module (see below).

**TABLE I. "Charge Values" in Different Atom Types Interactions**

| Atom type | (+) | (−) | (Hb-d) | (Hb-a) | (Hb-da) | (0) |
|---|---|---|---|---|---|---|
| Positively charged (+) | −4 | +4 | −3 | +3 | +3 | −2 |
| Negatively charged (−) | | −4 | +3 | −3 | +3 | −2 |
| H-bond donor (Hb-d) | | | −2 | +2 | +2 | −1 |
| H-bond acceptor (Hb-a) | | | | −2 | +2 | −1 |
| H-bond donor/ acceptor (Hb-da) | | | | | +2 | −1 |
| Apolar (0) | | | | | | +1 |

## CHARGES Module

CHARGES provides an estimate of the "electrostatic complementarity" of the solution complexes produced by SHAPES.

Each atom of the two structural elements is assigned to one among six different atom types: positively charged (N$\eta_1$/N$\eta_2$(Arg), N$\zeta$(Lys)), negatively charged (O$\delta_1$/O$\delta_2$(Asp), O$\epsilon_1$/O$\epsilon_2$(Glu)), hydrogen-bond donor (backbone N, N$\epsilon$(Arg), N$\delta_2$(Asn), N$\epsilon_2$(Gln)), hydrogen-bond acceptor (backbone O, O$\delta_1$(Asn), O$\epsilon_1$(Gln)), hydrogen-bond donor or acceptor (O$\gamma$(Ser), O$\gamma_1$(Thr), O$\eta$(Tyr), N$\delta_1$/N$\epsilon_2$(His)), apolar (everything else).

The distance between each atom of the target structural element and each atom of the probe structural element is measured. A "charge value" is calculated for atom pairs whose distance is lower than a fixed cutoff (Table I).

$$CHARGES = \sum_a \sum_b C_{ab}$$

where

$$C_{ab} = \begin{cases} 0 & \text{if} \quad R_{ab} > R_{cutoff} \\ A_{K_a K_b} & \text{if} \quad R_{ab} \leq R_{cutoff} \end{cases}$$

and

$$R_{cutoff} = \begin{cases} 4.0\ \text{Å} & \text{if both target and probe} \\ & \quad \text{atoms are apolar} \\ 3.4\ \text{Å} & \quad \text{otherwise} \end{cases}$$

$R_{ab}$ is the distance between atoms a and b; A is a matrix containing the values reported in Table I.

Each complex proposed by the module SHAPES is then discarded if the charge value is below a certain threshold. As already discussed for BUMPS, CHARGES threshold can be varied for different docking experiments: it is set lower (−110, low stringency) when docking *common* models, and higher (−50, high stringency) when docking all other types of structures.

The distances considered and the values assigned to the different potential electrostatic interactions do not mean to mirror real atomic interactions: the values shown in Table I and the cutoff considered were determined empirically by running the CHARGES module on a small database of protein complexes (trypsin/trypsin inhibitor, 3tpi; antibody/lysozyme, 2hfl; triose-phosphate isomerase dimer, 7tim) and on approximately 40 three-dimensional models of mutagenized Rop proteins. They have been set to allow for the correct prediction (in 96% of the cases) of the mutant dimerization phenotype. Atom classification assigned in Table I is similar to others.[6,13]

### Output of the ESCHER Procedure

ESCHER execution time varies with the size of the input structures: on a Silicon Graphics with an R8000 75MH processor, it may vary from a few minutes for two 30 amino acid long α-helices up to a few hours for large domains (more than 300 amino acids each).

The ESCHER procedure output is a list of solutions, each identified by the translations and rotations to be applied to the probe with respect to the target. In all of the docking experiments described, the number of listed solutions has been set to 500. This number has been increased to 5000 for the docking of unbound proteins, with a modest increase in calculation time.

Each solution is associated with the output scores of the three ESCHER modules described: a geometric complementarity score coming from SHAPES, a molecular collisions score from BUMPS, and an evaluation of electrostatic complementarity from CHARGES. The solutions with acceptable BUMPS and CHARGES values are retained and ranked according to their SHAPES score.

When SHAPES identifies an extended region of good complementarity, it usually generates a number of solutions in the same region. Therefore, all the solutions generated are grouped together on the basis of similarity; this is done by taking the solution with the highest geometric complementarity value and defining all other solutions having an rms deviation < 2.5 Å (5.0 Å for low-stringency conditions) from this structure as belonging to one cluster. The grouping continues with the best solution from those remaining and forms the next cluster by using the same procedure. Within each group, solutions are ranked according to their CHARGES value.

The resolution of the ESCHER procedure depends on two major factors: the size of the grouping matrix element (which affects the Tx, Ty, and Rz coordinates) and the stepwise exploration of the Z translation and of the X and Y rotations. The T and R coordinates identifying the solutions proposed should, therefore, be considered more properly as T ± ΔT and R ± ΔR. The expected error is different according to the coordinate considered: $\Delta T_z \sim 0.75$ Å; $\Delta R_x = \Delta R_y \sim 5°$; $\Delta R_z$ varies from 0° to 6°, $\Delta T_x$ and $\Delta T_y$ vary from 0 to 1.5 Å.

In some cases, when docking secondary structure elements, a loose distance constraint was applied to discard solutions not compatible with the primary sequence of the analyzed protein. In the experiment with the two Rop helices, we defined a 14.5 Å distance threshold between the α-carbons of residues 28 and 32; this threshold corresponds to the distance between the α-carbon of the N-terminal and C-terminal residues of a 5-residue long peptide in an extended conformation.

### Construction of the *Common*, *Similar*, and *Backbone* Structural Models

To simulate the experimental cases in which only information on the secondary structure is known, we used three types of structural models in which the amino acid side-chain or backbone atoms position are different from those found in the crystal structure: *common*, *similar*, and *backbone*.

*Common* model structures (Fig. 4a) were obtained from the crystallized element coordinates by substituting the conformation of each amino acid side chain with the most common rotamer in a database of amino acid rotamers[24] (on average, 4.5 rotamers/amino acid). *Common* models can therefore be built on the backbone atom coordinates without any information on side-chain conformation. They display an average rms ~ 0.80 Å with respect to the original crystal structures.

*Similar* model structures (Fig. 4b) have been obtained by replacing each side chain with the most similar rotamer in the rotamer database. These models can be built only when the original crystal structure is known. The *similar* model structures display an average rms ~ 0.30 Å with respect to the corresponding crystal structure.

The *backbone* model structures (Fig. 4c) were built by substituting Rop helices backbone with geometrically regular α-helices ($\phi = -65°$, $\psi = -40°$, $\omega = 180°$), by using the secondary structure option in the Biopolymer module of insightII.[25] Their rms deviation from the starting structures is ~0.55 Å. No energy refinement was applied to the model structures obtained.

### Structural Elements Preparation

All of the target and probe structural elements docked by ESCHER in this work were obtained by splitting single-protein structures into two parts. The splitting procedure consisted of three steps. In the first step, each amino acid was defined as belonging to either the target or the probe structural elements. In the second, the protein main chain was

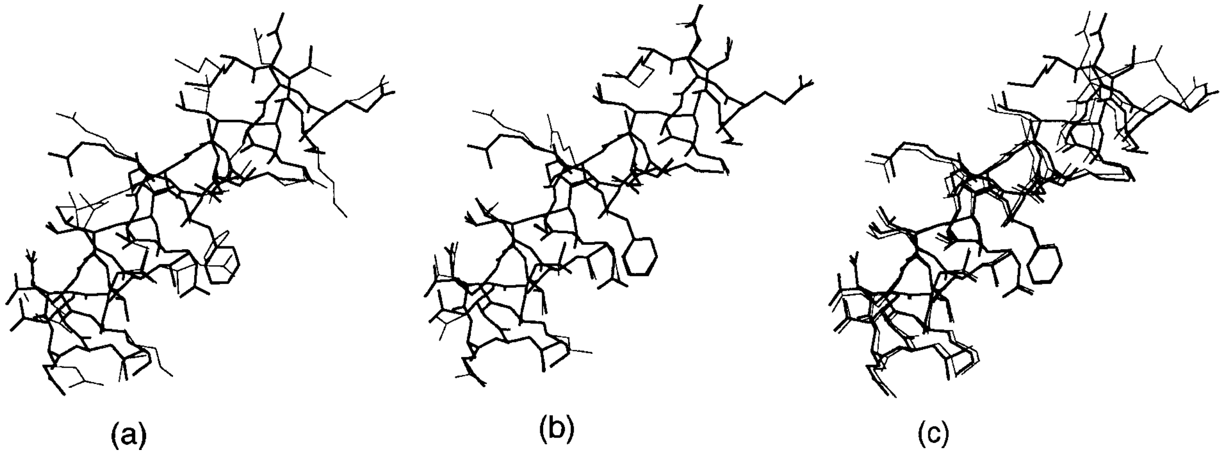

Fig. 4. Rop helix 1 (a1–a30) crystallographic structure is shown in bold. It is superposed to the *common* model structure (**a**), to the *similar* model structure (**b**), and to the *backbone* model (**c**).

cut at the peptide bond between target and probe amino acids, generating a number of peptide fragments. In the last step, the target and probe structural elements were defined as the merged ensemble of their fragments. When the protein had to be split into two domains, the target and probe were defined with automated criteria,[26–29] when the protein had to be split into secondary structure elements, the separation was performed according to secondary structure assignment (see Table 4).

The smallest structural element was designed as the probe. The starting orientation of the target has always been chosen so that the interface was approximately parallel to the Z axis. In each docking experiment, the two parts were positioned in the worst possible relative orientation: the probe was rotated 5° around the X axis and 5° around the Y axis away from the crystallographically defined orientation (SHAPES explores the rotation space around the X and Y axis by 10° steps).

### Protein Structures Analyzed

Twenty-three proteins of known structure were used for our docking experiments from the Protein Data Bank (PDB).[30] For test runs with protein domains, we used the following: porcine pancreatic elastase (3est); papain (9pap); penicillopepsin (1ppl); papain (1ppn); bovine liver rhodanese (1rhd); thermolysin (4tln); bovine ribonuclease A (1rnd); yeast guanylate kinase (1gky); bovine gamma a-crystallin (2gcr); a cytochrome C2 (2c2c); a bacterial holo-D-glyceraldehyde-3-phosphate dehydrogenase (1gd$_1$); a bacterial trypsin (1sgt); rat mast cell protease II (3rp2); a bacterial dihydrofolate reductase (3dfr); alpha-lactalbumin (1alc); a protozoan calmodulin (1clm); yeast phosphoglycerate kinase (3pgk); a bovine prothrombin fragment (2pf2). These molecules were selected for our tests because the definition of their domains were derived by means of automated procedures and were present in the literature.[26–29] For test runs with unbound components, we chose one protease/inhibitor complex (cocrystallized structure: 1cho, corresponding structures crystallized separately: 5cha/2ovo); and two antibody/antigen complexes: 1) the antibody moiety of the 3hfm structure versus the lysozyme unbound structure 6lyz; 2) the antibody moiety of the 2hfl structure versus 6lyz. In these two last cases, solutions were compared with the 3hfmy and 2hfly molecules. For test runs with secondary and supersecondary structure elements, we used the Rop protein (PDB code: 1rop[31]); the light chain of Hy-Hel10 (3hfm[32]) and lysozyme (6lyz[33]).

## RESULTS

In an initial series of experiments (data not shown), ESCHER has been successfully applied to the reconstruction of protein complexes starting from the crystal structure of their separated components: trypsin/trypsin inhibitor (3tpi); antibody/lysozyme (2hfl); Rop and triose-phosphate isomerase dimers from their separated monomers (1rop and 7tim).

### Docking of Protein Domains

We have applied ESCHER to the reconstruction of two-domain proteins starting from their separated domains. The proteins chosen represent a good database for the study of protein domain interaction: some of them have an extended buried surface and a large number of domain contacts, whereas others show only a limited number of interdomain contacts and a very small buried surface (Table II).

The target and probe domains were prepared as described in Methods, starting from the deposited PDB structure and following automatic criteria[26–29] for the target and probe definition.

In Table II, the results obtained with protein domains are shown: for each domain pair, the num-

**TABLE II. Protein Domains Results**

| PDB code | Target | Probe | $\phi \times 10^{-3}$ | Buried surface (Å$^2$) | Rank (rms*) of the first correct solution† | Rank (rms) of the best solution | Number of solutions proposed |
|---|---|---|---|---|---|---|---|
| 1clm | 5–88 | 89–147 | 1.1 | 263 | — | | 8 |
| 3pgk | 189–404 | 1–188, 405–415 | 1.1 | 639 | — | | 10 |
| 1gky | 1–32, 82–186 | 33–81 | 3.3 | 468 | — | | 1 |
| 2pf2 | 63–145 | 1–62 | 4.2 | 457 | — | | 7 |
| 2gcr | 1–81 | 89–169 | 5.3 | 532 | — | | 3 |
| 3dfr | 1–33, 109–162 | 34–108 | 12.1 | 932 | — | | 1 |
| 1rnd | 1–49, 80–103 | 50–79, 104–124 | 14.5 | 726 | — | 8 (3.3) | 19 |
| 1ppl | 1–192, 304–323 | 193–303 | 19.1 | 1603 | 1 (2.4) | 1 (2.4) | 1 |
| 1rhd | 1–160 | 161–293 | 20.4 | 2001 | 1 (1.5) | 1 (1.5) | 1 |
| 2c2c | 63–95 | 96–112 | 20.8 | 861 | 7 (2.0) | 7 (2.0) | 18 |
| 4tln | 1–151 | 152–316 | 21.6 | 1807 | 1 (1.6) | 1 (1.6) | 3 |
| 1alc | 38–104 | 1–37, 105–122 | 21.7 | 972 | 1 (1.8) | 1 (1.5)‡ | 31 |
| 1gd1 | 1–148 | 149–332 | 22.0 | 2186 | 1 (2.0) | 2 (1.5)‡ | 11 |
| 9pap | 1–16, 113–208 | 17–112, 209–212 | 23.7 | 1387 | 2 (2.0) | 2 (2.0) | 6 |
| 1ppn | 1–18, 111–212 | 19–110 | 28.0 | 1567 | 1 (2.1) | 2 (1.5) | 4 |
| 1sgt | 16–28, 69–80, 121–234 | 29–68, 81–120 | 31.2 | 1583 | 1 (1.5) | 1 (1.5) | 9 |
| 3est | 30–121, 234–245 | 16–29, 121–233 | 38.9 | 2186 | 1 (1.7) | 1 (1.7) | 4 |
| 3rp2 | 1–27, 123–216, 221–230 | 28–122, 231–243 | 47.6 | 2382 | 1 (1.6) | 3 (1.4)‡ | 5 |

*All atoms rms is calculated by superposing the target domain of the docking solution complex onto the target domain of the protein crystal structure.

†Rank and rmsd of the first solution having an rmsd below 2.5 Å.

‡Identifies solutions not presenting the best CHARGES value of their cluster (see Methods).

ber of proposed solutions is reported together with the ranking of the correct solution and its all-atom root-mean-square deviation (rmsd) from the crystallographic structure. In 9 of 18 docking runs, the correct solution was found and ranked first; in two cases, the correct solution ranked among the first few solutions; in the remaining seven cases, ESCHER was not able to find the correct solution within the first 500 solutions proposed by the SHAPES module. In all but one of these seven cases (the bacterial dihydrofolate reductase, 3dfr), the number of atomic contacts among the two domains is generally low. It is interesting to note that 1) the number of solutions proposed by the SHAPES module can be increased up to many thousands (if necessary), with a modest increase in calculation time and 2) that, on the other hand, the number of solutions that escape the BUMPS and CHARGES filters is generally very low: in case of a really blind test, this would allow further investigation of few solutions by more sophisticated methods.

There is no evident correlation between ESCHER efficiency and buried surface area in the domain-domain interface. A strong correlation could be detected between ESCHER efficiency and a value $\phi$, defined as the product of the ratios between the buried and the whole surface of each domain:

$$\phi = \frac{b_1 \cdot b_2}{tot_1 \cdot tot_2}$$

where $b_1$ and $b_2$ are the two domain buried surfaces and $tot_1$ and $tot_2$ are the two domains total surfaces. This value has no physical significance but can be used to predict ESCHER ability in rebuilding a complex of known structure.

In the test cases that we analyzed, when $\phi > 19 \times 10^{-3}$, ESCHER ranked the correct solution among the first few. For lower $\phi$ values, the correct solution was not found among the first 500 solutions proposed by SHAPES.

### Docking of Protein Complexes Starting From Unbound Components

We chose three cases: a protease-protease inhibitor and two antibody-antigen complexes. The results of the docking procedure were evaluated by superimposing the docked (originally unbound) molecules onto their crystal structures in the PDB complexes. These complexes (Table III, second column) were used as reference systems to calculate the rmsd from the C$\alpha$ of the proposed solutions.

The protease-protease inhibitor complex was run by using the entire protein surfaces, and the procedure proposed only three solutions. A solution with an rmsd of 1.9 Å with respect to the defined reference system scored third.

In the immunoglobulin-lysozyme docking experiment, the entire antigen was docked onto the complementary determining region (CDR) of the immunoglobulin. The 3hfm experiment ended with only two

**TABLE III. Unbound Proteins Results**

| | Reference for complex pdb structure | Rank (rms*) of the first correct solution§ | Rank (rms) of the best solution | Number of solutions |
|---|---|---|---|---|
| 2ovo/5cha | 1cho | 3 (1.9) | 3 (1.9) | 3 |
| 2hfl/6lyz† | 2hfl | — | — | 0 |
| 3hfm/6lyz† | 3hfm | 2 (1.0) | 2 (1.0) | 2 |
| 2hfl/6lyz‡ | 2hfl | 11 (1.12) | 11 (1.12) | 11 |
| 3hfm/6lyz‡ | 3hfm | 1 (1.0) | 1 (1.0) | 1 |

*The rmsd of the unbound molecules with their corresponding reference molecules is obtained after superposing the Cα atoms of the unbound ligand (receptor) on those of the complexed reference ligand (receptor).

†Antibody CDRs versus entire lysozyme molecule. We selected a subset of the antibody solvent accessible surface, whose distance from the antigen is less than 6 Å in the cocrystallized structure; regions of geometric complementarity were evaluated if sharing at least 1 $Å^2$ with the defined subset.

‡Antibody CDRs versus the lysozyme antibody binding site. We selected a subset of the lysozyme solvent accessible surface, whose distance from the antibody is less than 6 Å in the cocrystallized structure; complementarity has been evaluated for regions sharing at least 1 $Å^2$ with this defined subset. In this case, approximately half of the antigen molecule has been considered in the docking test.

§Rank and rmsd of the first solution having an rmsd below 2.5 Å.

proposed solutions, the second one displaying an rms of 1.0 Å with respect to its reference in the crystal complex. No docking solution was found for the 2hfl complex. When restricting the search area on the antigen to approximately half of the lysozyme molecule, including the epitope, the procedure ended with 11 solutions: the eleventh solution shows 1.12 Å rms from the lysozyme crystal structure in the complex. This result shows that the correct solution ranks lower than 5000th in the SHAPES scoring list, when the entire lysozyme molecule is used as probe protein; however, in this case, ESCHER proposed no solution. When restricting the search area on the probe protein, the correct solution was in the SHAPES scoring list and could be successfully selected from the BUMPS and CHARGES modules.

ESCHER results with unbound components from protein-protein complexes can be compared with others.[8,10] In the protease-inhibitor complex, Fischer et al. obtained 8490 solutions: the best solution (also the first correct solution) ranked 214th with an rms of 1.32 Å. In the same case, the PUZZLE procedure[8] proposed only four solutions: the best solution ranked second with an rms of 7.3 Å (4.9 Å when considering only the contacting region).

In the antibody-antigen complex, Fischer et al. obtained 8957 solutions: the best solution ranked 2436th with 1.28 Å rms from the crystallographic position; the best ranking correct solution ranked 1497th with rms = 1.97 Å. Using the PUZZLE procedure, we had only three solutions: the correct solution scored first with an rms of 5.4 Å (4.2 Å when considering only the contacting region).

**TABLE IV. Structural Elements Definition**

| | PDB code | Target | Probe | $\phi \times 10^{-3}$ |
|---|---|---|---|---|
| 1α/3α | 1ROP | a1–a30,<br>b1–b56 | a31–a56 | 28 |
| 1α/αβ | 6LYZ | 1–86,<br>103–129 | 87–102 | 18 |
| 1α/1α | 3HFM | L1–L27,<br>L60–L77 | L28–L59,<br>L78–L108 | 30 |
| 1α/1α | 1ROP | a1–a30 | a31–a56 | 20 |

The described results show that generally a low rms can be obtained in a long list of possible solutions and that the choice of selecting only a few solutions, including the correct one, is usually paid in terms of higher rmsd. Both procedures, the geometry-based suite[10] and PUZZLE,[8] generally require four to five times less cpu time than ESCHER.

## Docking of Secondary and Supersecondary Structural Elements

The experiments described in the previous section show that 1) ESCHER can reconstruct whole proteins starting from their separated crystal domains when the interdomain contact area is sufficiently large with respect to the total surface area and 2) ESCHER is able to dock unbound protein components whose structures display differences with respect to their conformation in the crystallographic complex.

Next, we explored the feasibility of docking secondary and supersecondary structure elements, where the $\phi$ value is generally higher than $19 \times 10^{-3}$ (Table IV).

We chose three docking cases, in which the docking partners were as follows: 1) a secondary structure element separated from its structural environment and the remaining part of the protein. We selected alpha helices extracted either from Rop, a four-helix bundle protein (1α/3α), or from the alpha and beta protein lysozyme (1α/αβ); 2) two supersecondary structures forming a protein domain: the beta sheets components of an immunoglobulin variable domain (1β/1β); and 3) two secondary structure elements: the two helices forming the Rop monomer (1α/1α).

We performed docking experiments starting either from the crystal structures or from approximate molecular models. Three different types of models (Fig. 4) were utilized as described in Methods: 1) *similar* models are derived from crystal structures by substituting each amino acid rotamer with the most similar rotamer from a small database of amino acid rotamers.[24] Their rmsd from the crystal structures is approximately 0.30 Å. 2) In *backbone* models, backbone dihedral angles are approximated to

**TABLE V. Secondary Structure Results (High Stringency*)**

| | Model structure | Rank (rms) of the first correct solution† | Rank (rms) of the best solution | Number of solutions |
|---|---|---|---|---|
| 1α/3α | crystal | 1 (1.4) | 1 (1.3)‡ | 8 |
| 1α/αβ | crystal | 1 (1.2) | 1 (1.0)‡ | 18 |
| 1β/1β | crystal | 1 (1.8) | 1 (1.4)‡ | 9 |
| 1α/1α | crystal | 1 (2.3) | 5 (1.5) | 24 |
| 1α/3α | similar | 1 (1.4) | 1 (1.4) | 2 |
| 1α/αβ | similar | 1 (1.1) | 1 (0.9)‡ | 11 |
| 1β/1β | similar | 1 (1.4) | 1 (1.3)‡ | 10 |
| 1α/1α | similar | 2 (1.8) | 2 (1.7)‡ | 18 |
| 1α/3α | backbone | 1 (1.7) | 1 (1.3)‡ | 8 |
| 1α/1α | backbone | 1 (2.1) | 1 (1.6)‡ | 23 |
| 1α/3α | common | — | 1 (4.1)‡ | 2 |
| 1α/αβ | common | 1 (1.0) | 1 (1.0) | 13 |
| 1β/1β | common | 2 (2.3) | 1 (1.5)‡ | 7 |
| 1α/1α | common | — | 2 (8.1) | 10 |

*High-stringency conditions: highly selective BUMPS and CHARGES thresholds (50 and −50, respectively).

†Rank and rmsd of the first solution having an rmsd below 2.5 Å.

‡Solutions not presenting the best CHARGES value of their group.

**TABLE VI. Secondary Structure Results (Low Stringency*)**

| | | Rank (rms) of the best ranking correct solution† | Rank (rms) of the best solution |
|---|---|---|---|
| 1α/3α | common | 2 (3.8) | 77 (2.4) |
| 1α/αβ | common | 1 (3.0) | 4 (1.0) |
| 1β/1β | common | 1 (3.1) | 7 (1.5) |
| 1α/1α | common | 84 (4.0) | 84 (4.0) |

*Low-stringency conditions: nonselective BUMPS and CHARGES thresholds (110 and −110, respectively); a solution is considered correct when its backbone rmsd from the crystal position is lower than 5 Å; solutions are grouped, as described in Methods, whenever their difference is lower than 5 Å.

†Rank and rmsd of the first solution having an rmsd below 5 Å.

geometrically regular alpha-helical angles (rms $\sim$ 0.55 Å from the crystal structure in the described cases). 3) In *common* models, amino acid rotamers are substituted with the most common rotamer from the rotamer database (rms $\sim$ 0.80 Å from the crystal structure). Each pair of these crystal or model structures has been used in an ESCHER docking simulation. Target and probe were prepared as described above (for target and probe definition, see Table IV). Results are reported in Tables V and VI.

When the crystallographic structures are used, the correct solution always scores first. When the *similar* model structures are used, the correct solution scores first in three of the four tested cases. The 1α/1α correct solution scored second in a total of 18 solutions proposed.

*Backbone* models were built only for the reconstruction of the Rop four helix bundle: the correct solution scored first in both the 1α/1α and 1α/3α docking experiments.

*Common* structural models have been subjected to the BUMPS and CHARGES modules under two different conditions, described as low and high stringency in Methods. A larger number of molecular collisions and a worse electrostatic complementarity were tolerated in the low-stringency condition, whereas high-stringency conditions corresponded to the ones used in the crystal, *similar,* and *backbone* experiments.

When high-stringency conditions are used (Table V), ESCHER produces a low rms solution in the first ranking position in the 1α/αβ and 1β/1β test cases. In the other two cases tested, no correct solution was obtained. When using low-stringency conditions (Table VI), the correct solution scores among the first ones in three of the four tested cases. In low-stringency conditions, a solution is considered correct when its backbone rmsd from the crystal position is lower than 5 Å.

Analysis of the solutions obtained with low-stringency conditions reveals that, in the 1α/1α and in the 1β/1β cases, *common* model structures display a substantially lower degree of electrostatic complementarity compared with the corresponding crystal structure; in the 1α/3α case, both molecular clashes and lowering of electrostatic complementarity contribute to the lack of low rms solutions in the proposed solutions.

We report the results obtained under both conditions for discussion.

Finally, a control test was performed in which a 30 amino acid long Rop helix is docked onto the lysozyme domain 2, which normally shelters a 15 amino acid helix. ESCHER geometric module predicted the formation of a Rop/lysozyme "mosaic" protein, with a good geometric complementarity. However, the electrostatic module gave a very poor score of the electrostatic complementarity between the two protein portions and discarded it.

In all the docking runs described, the target and the probe were treated as if they were independent molecules; nevertheless, distance constraints could be applied, thus discarding physically impossible solutions. In the 1α/3α and 1α/1α cases, when using *similar* or *common* model structures, a number of proposed solutions was antiparallel to the correct ones. They were, therefore, discarded because amino acids close in the primary sequence were too far in the proposed solution.

## DISCUSSION

ESCHER is a new rigid body docking algorithm that was developed after PUZZLE.[8] ESCHER differs from PUZZLE in the simplified description of the protein surface as well as in the matching and scoring criteria. Both procedures rely on a geometric approach, but ESCHER has been endowed with a molecular clashes (BUMPS) and an electrostatic (CHARGES) module, which are used as a selective filter on the geometrically favorable solutions. The CHARGES electrostatic module allows a lower resolution definition of the protein surfaces and less stringent criteria of geometric complementarity. Typically, when docking unbound components, thousands of solutions are proposed by the geometric module, but only a few, including the correct one, survive screening with the electrostatic module. This illustrates the importance of adding this module to the procedure.

The application of ESCHER to a database of two-domain proteins indicates that its performance correlates with the ratio between the buried surface area upon domain-domain complex formation and the total surface ($\phi$). When starting from the crystallographic coordinates of the target and probe domains with $\phi$ greater than $19 \times 10^{-3}$ (see Table II), the correct docking solution always scores among the first few.

ESCHER has been applied to the unbound components of a small database of protein-protein complexes. The low resolution description of the protein surface permits recognition of correct protein interfaces even when side-chain movements upon binding substantially alter the geometric and electrostatic complementarity of the interface.

In comparison with other published docking procedures,[8,9,10,34] ESCHER generally requires longer computational times but can propose low rms solutions among a very restricted number of possible choices. This allows one to rank the small number of possible solutions by applying more sophisticated computational methods.

We have also explored the feasibility of applying ESCHER to the reassociation of secondary and supersecondary structure elements within a protein structure. We simulated a helix-helix and a helix-three helices interaction within a four-helix bundle (Rop); a helix and the remaining part of the protein in an alpha-beta protein (lysozyme) and two beta-sheets within a beta protein (an immunoglobulin variable domain).

Initially, these docking experiments were performed with secondary structure elements extracted from the crystallographic structure, without altering the conformation of the side chains and maintaining a rigid alpha-carbon backbone. In this ideal situation, in which the crystallographically determined complementarity is strictly maintained, ESCHER proved to be very effective. In a realistic docking experiment, however, the conformation of the side chains in the complex is not known. Thus, to explore the versatility of ESCHER and to simulate the assembly of a three-dimensional model from the definition of its secondary structure, we applied the ESCHER procedure to the docking of secondary structure elements after changing the conformation either of the amino acid side chains or of the backbone. The results obtained with the *similar* and *backbone* models are very satisfactory. With *common* models (Table V), the results were satisfactory in half of the cases tested: changing the rotamer conformations to the most common ones found in the database considerably lowered the geometric complementarity of target and probe elements. It was, therefore, necessary to lower the stringency of the BUMPS and CHARGES selective filters (Table VI) to achieve acceptable results in three of the four docking experiments.

The performance obtained in the *common* $1\alpha/1\alpha$ case deserves more discussion: this approach does not seem to be applicable to elements, such as single alpha helices or beta strands. Work is in progress to see whether, in these difficult cases, our procedure could be supported and complemented by different methods, such as the use of backbone-dependent rotamer libraries[35,36] or the identification of correlated mutations.[37]

The relatively low resolution of the best solutions in the experiments performed with *common* models, however, are more appropriately to be compared with resolutions typical of methods that predict structure from sequence rather than with docking methods. This confirms that, especially when considering small docking elements, the side-chain conformation is not irrelevant to the definition of the geometric and electrostatic complementarity. The success of the experiments with the *similar* models, however, reassures that, by exploring combinatorially a small library of side-chains conformations, the probability of predicting the "correct solution" is very high.

Other applications for the ESCHER procedure can be envisaged. The structure of a growing number of large proteins is approached by experimentally determining the conformation of their separate domains. We propose ESCHER to represent a useful tool to assist in the task of assembling the complete protein starting from the crystallographic structures of its domains.

It is often found that the structure of newly discovered proteins can be reliably predicted by homology modeling. Alignment of the two primary sequences, however, may reveal that the protein of unknown conformation contains peptide insertions with respect to the protein in the database. Whenever the secondary structure of the inserted peptide can be predicted, ESCHER can help in docking the orphan peptide onto the "core" structure determined by homology modeling.

ESCHER has already been exploited as a tool to predict the formation of protein-protein complexes. In one application, starting from the coordinates of the crystallized proteins, we have predicted the conformation of the complex between the TP7 antibody and its cognate antigen, *Thermus aquaticus* DNA polymerase.[38]

Furthermore, ESCHER has been used as a tool to design and evaluate Myc mutants that can homodimerize (manuscript in preparation). Work is in progress to address the possibility of speeding up the procedure on parallel (or even massively parallel) machines.

## ACKNOWLEDGMENTS

The authors thank Dr. A. Tramontano for helpful support, advice, and discussion and Dr. A. Lahm for critically reading the manuscript. We acknowledge the expert computer support of M. Pasquariello and M. Pizzelli at Centro di Calcolo e Documentazione of the University of Rome "Tor Vergata." This work has been supported by the Supercomputing Resource for Molecular Biology, European Union Human Capital and Mobility Programme, Access to Large Scale Facilities grant, Contract ERBCHGECT940062, and by the CNR programme "Development of innovative algorithms for dynamic and static simulation of molecular recognition."

## REFERENCES

1. Cherfils, J., Janin, J. Protein docking algorythms: Simulating molecular recognition. Curr. Opin. Struct. Biol. 3:265–269, 1993.
2. Strynadka, N.C., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B.K., Kuntz, I.D., Abagyan, R., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson, A., Duncan, B., Rao, M., Jackson, R., Sternberg, M., James, M.N. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. Nat. Struct. Biol. 3:233–239, 1996.
3. Shoichet, B., Kuntz, I.D. Protein docking and complementarity. J. Mol. Biol. 221:327–346, 1991.
4. Bacon, D.J., Moult, J. Docking by least-squares fitting of molecular surface patterns. J. Mol. Biol. 225:849–858, 1992.
5. Lenhof, H.P. An algorithm for the protein docking problem. In: "Bioinformatics: From Nucleic Acids and Proteins to Cell Metabolism." Schomburg, D., Lessel, U. (eds.). GBF. Weinheim, 1995.
6. Jiang, F., Kim, S.H. "Soft docking": Matching of molecular surface cubes. J. Mol. Biol. 219:79–102, 1991.
7. Walls, P.H., Sternberg, J.E. New algorythm to model protein-protein recognition based on surface complementarity. J. Mol. Biol. 228:277–297, 1992.
8. Helmer-Citterich, M., Tramontano, A. PUZZLE: A new method for automated protein docking based on surface shape complementarity. J. Mol. Biol. 235:1021–1031, 1994.
9. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. USA 89:2195–2199, 1992.
10. Fischer, D., Lin, S.L., Wolfson, H.L., Nussinov, R. A geometry-based suite of molecular docking processes. J. Mol. Biol. 248:459–477, 1995.
11. Cherfils, J., Duquerroy, S., Janin, J. Protein-protein recognition analyzed by docking simulation. Proteins 11:271–280, 1991.
12. Jackson, R.M., Sternberg, M.J.E. A continuum model for protein-protein interactions: Application to the docking problem. J. Mol. Biol. 250:258–275, 1995.
13. Sobolev, V., Wade, R.C., Vriend, G., Edelman, M. Molecular docking using surface complementarity. Proteins 25:120–129, 1996.
14. Rosenfeld, R., Vajda, S., De Lisi, C. Flexible docking and design. Annu. Rev. Biophys. Biomol. Struct. 24:677–700, 1995.
15. Rarey, M., Kramer, B., Lengauer, T., Klebe, G. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 261:470–486, 1996.
16. Sippl, M.J., Hendlich, M., Lackner, P. Assembly of polypeptide and protein backbone conformations from low energy ensembles of short fragments. Development of strategies and construction of models for myoglobin, lysozyme and thymosin b-4. Protein Sci. 1:625–640, 1992.
17. Monge, A., Friesner, R.A., Honig, B. An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure. Proc. Natl. Acad. Sci. USA 91:5027–5029, 1994.
18. Monge, A., Lathrop, E.J.P., Gunn, J.R., Shenkin, P.S., Friesner, R.A. Computer modeling of protein folding: Conformational and energetic analysis of reduced and detailed protein models. J. Mol. Biol. 247:995–1012, 1995.
19. Cohen, F.E., Richmond, T.J., Richards, F.M. Protein folding: Evaluation of some simple rules for the assembly of helices into tertiary structures with myoglobin as an example. J. Mol. Biol. 132:275–288, 1979.
20. Cohen, F.E., Kuntz, I.D. Prediction of the three-dimensional structure of human growth hormone. Proteins 1:162–166, 1987.
21. Callaway, D.J.E. Solvent-induced organization: a physical model of folding myoglobin. Proteins 20:124–138, 1994.
22. Connolly, M.L. Solvent-accessible surface of proteins and nucleic acids. Science 221:709–713, 1983.
23. Connolly, M.L. Analytical molecular surface calculation. J. Appl. Crystallogr. 16:548–558, 1983.
24. Ponder, J.W., Richards, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. 193:775–791, 1987.
25. Dayringer, H.E., Tramontano, A., Sprang, S.R., Fletterick, R.J. Interactive program for visualization and modelling of proteins, nucleic acids and small molecules. J. Mol. Graphics 4:82–87, 1986.
26. Holm, L., Sander, C. Parser for protein folding units. Proteins 19:256–268, 1994.
27. Siddiqui, A.S., Barton, G.J. Continuous and discontinuous domains: An algorithm for the automatic generation of reliable protein domain definitions. Protein Sci. 4:872–884, 1995.
28. Sowdhamini, R., Blundell, T.L. An automatic method involving cluster analysis of secondary structures for the identification of domains in proteins. Protein Sci. 4:506–520, 1995.
29. Swindells, M.B. A procedure for detecting structural domains in proteins. Protein Sci. 4:103–112, 1995.
30. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., 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. 112:535–542, 1977.
31. Banner, D.W., Kokkinidis, M., Tsernoglou, D. Structure of the ColE1 Rop protein at 1.7 Å resolution. J. Mol. Biol. 196:657–675, 1987.
32. 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 hy/hel10 fab-lysozyme complex. Proc. Natl. Acad. Sci. USA 86:5938–5942, 1989.
33. Diamond, R. Real-space refinement of the structure of hen egg-white lysozyme. J. Mol. Biol. 82:371–391, 1974.
34. Vakser, I.A. Protein docking for low-resolution structures. Protein Eng. 8:371–377, 1995.
35. Dunbrack, R.L., Karplus, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J. Mol. Biol. 230:543–574, 1993.
36. Dunbrack, R.L., Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat. Struct. Biol. 1:334–340, 1994.
37. Gobel, U., Sander, C., Scheneider, R., Valencia, A. Correlated mutations and contact in proteins. Proteins 18:309–317, 1994.
38. Murali, R., Helmer-Citterich, M., Sharkey, D.J., Scalice, E., Daiss, J., Murthy, K. Structural Studies on a Inhibitory Antibody Against *Thermus aquaticus* DNA Polymerase Suggest Mode of Inhibition (submitted).