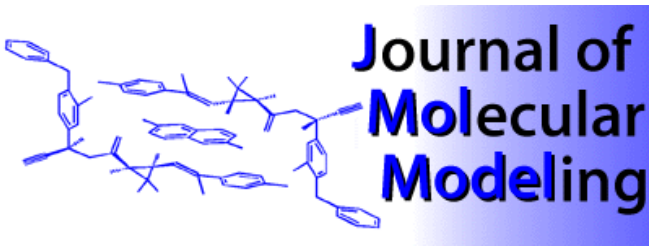
J. Mol. Model. **1997**, *3*, 321 – 324



New Docking Algorithm Based on Fuzzy Set Theory[$]

Thomas E. Exner[1] and Jürgen Brickmann[1,2,*]

[1]Department of Physical Chemistry, Darmstadt University of Technology, D-64287 Darmstadt, Germany
(brick@pc.chemie.th-darmstadt.de)

[2]Darmstadt Center of Scientific Computing, D-64287 Darmstadt, Germany

Received: 19 June 1997 / Accepted: 10 July 1997 / Published: 11 August 1997

Abstract

A new docking algorithm based on shape complementarity is presented. The algorithm is based on Fuzzy Logic strategies. A small number of possible docking configurations is selected using partial ordering of a fuzzy measure of topographical properties between complementary surface patches. The algorithm was tested with the structures of three protein-protein-complexes of serin proteases. In one case the components determined separately by x-ray investigations whereas in the other cases the components obtained from the known complex structure were used.

Keywords: Protein-protein-docking, Shape complementarity, Fuzzy Set Theory

Introduction

The specific recognition of two macromolecules plays a major role in biological activity and specificity. In recent years many docking algorithms predicting the structure of protein - ligand complexes were published. The program DOCK [1,2] generates docking configurations by matching subsets of ligand atoms onto a "negative image" of the protein, consisting of a set of overlapping spheres. Helmer-Citterich and Tramontano [3] introduced an algorithm describing the geometric shape of the surfaces by a 2-dimensional matrix. The search for complementary regions on the surfaces of the protein and the ligand is reduced to the comparison of sub-matrices. Within the algorithms of Connolly [4] and Norel et al. [5,6] the surfaces of the molecules are represented by critical points, describing knobs and holes. These critical points are used to predict the complex structure. Nevertheless, none of them could solve the docking problem satisfactorily mainly due to the fact that one has to know the docking regions in advance in order to obtain reasonable results. In this work a new docking algorithm based on Fuzzy Set Theory [7] is presented. This algorithm uses topographical properties of the "solvent accessible surfaces" [8] for matching the shapes of both molecules.

Basic Principles of Fuzzy Sets

Fuzzy Set Theory may be regarded as a generalization of classical set theory. A fuzzy set A is denoted by an ordered

[$] *Presented at the 11. Molecular Modeling Workshop, 6 -7 May 1997, Darmstadt, Germany*

* *To whom correspondence should be addressed*

set of pairs, the first element of which denotes the element x in the definition space X and the second the degree of membership. The latter is defined by a membership function $\mu_A(x)$, with values lying within the range $0 \le \mu_A(x) \le 1$ between zero and complete membership.

$$A = \{(x, \mu_A(x)) | x \in X\}$$

A more detailed description is given in the literature. [9,10]

Docking Algorithm

For the representation of molecular surfaces the definition of "global curvatures" introduced by Zachmann et al.[11] was used. The global curvatures may be interpreted as average curvatures of the corresponding surface region. These curvatures can be combined to a single quality for convexity increasing continuously through five basic shape descriptors (0 – 4) plus a flatness value (–1) if both curvatures are equal to zero.[12] This quality is labeled as "surface topography index" (STI).[12] Based on the STI values the molecular surface can be subdivided with fuzzy dissimilarity measures into discrete domains which can be labeled by linguistic variables.[6]

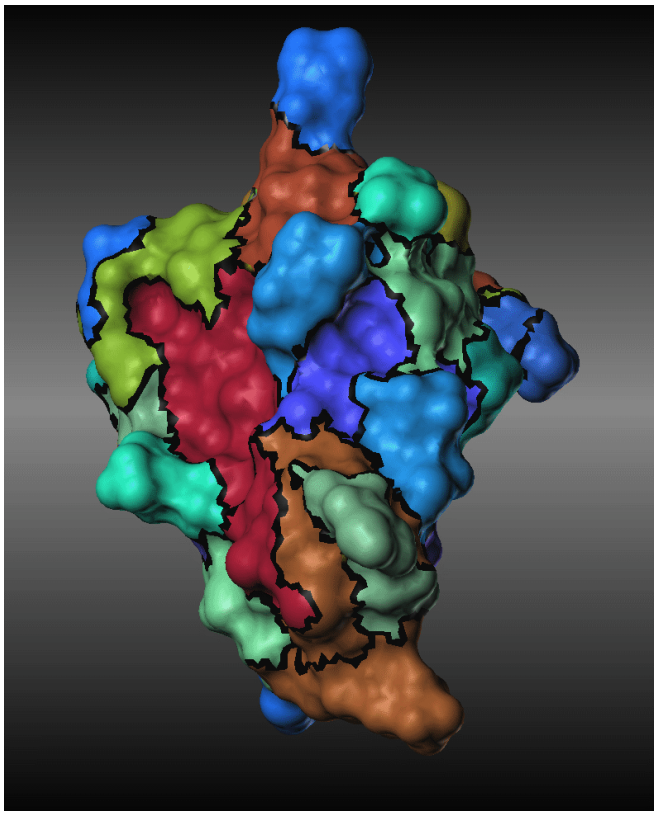


Figure 1. *Segmentation of the surface of PTI (solvent accessible surface). Each domain is color-coded according to the index number of the domain*

Table 1. *Docking results. Numbers of docking configurations suggested by our docking algorithm and rms-values of the best configuration to the structure investigated by X-ray analysis.*

Complex	number of docking configurations	rms-value of best configuration
Trypsin-PTI	28	10.8
Trypsin-PTI (unbounded)	50	
HLE-Ovomucoid	91	2.5
Chymotrypsin-Ovomucoid	80	6.5

In the docking algorithm each domain is represented by a reference point. A normal vector is associated to the domains by making the average of the normal vectors of the domain's surface points. The averages of the STI and the global curvatures of the domain's surface points as well as the sizes of the domains are also determined. The values are used in a Fuzzy Logic based method to calculate the complementarity surfaces of the protein and ligand. Possible docking structures were defined by the position of the reference points of the complementary domains and the reference points of the domain having a common border with these domains.

Results

The algorithm was tested with structures of three complexes of serin proteases with their inhibitors. The selected complexes were trypsin - pancreatic trypsin inhibitor (PTI), a-chymotrypsin - ovomucoid inhibitor and human leukocyte elastase - ovomucoid inhibitor. In the case of the trypsin-PTI complex both the components obtained from the known complex structure and the independently determined structures of the components were used. In the other two examples only the structures of the known complexes were used.

For each protein - ligand complex less than 100 possible docking configurations were discovered and there was at least one of these configurations with a root-mean-square derivation (rms) of less than 10 Å from the original x-ray structure. In future studies this configuration can be used combined with an efficient minimization algorithm to optimize the prediction of the complex structure.

Supplementary material available statement: The used complex structures taken from the Brookhaven Protein Data Bank (http://www.pdb.bnl.gov) are available as supplementary

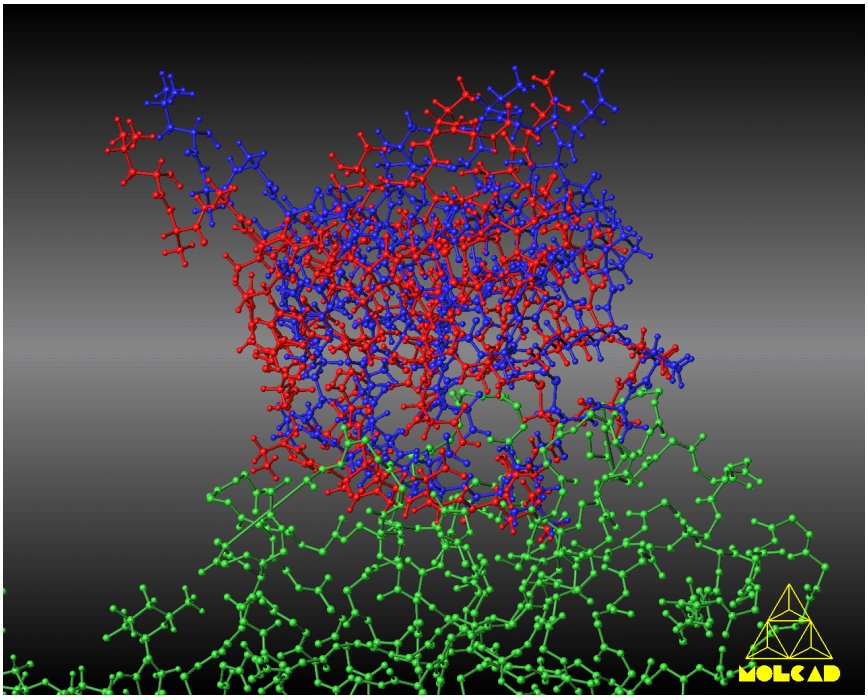
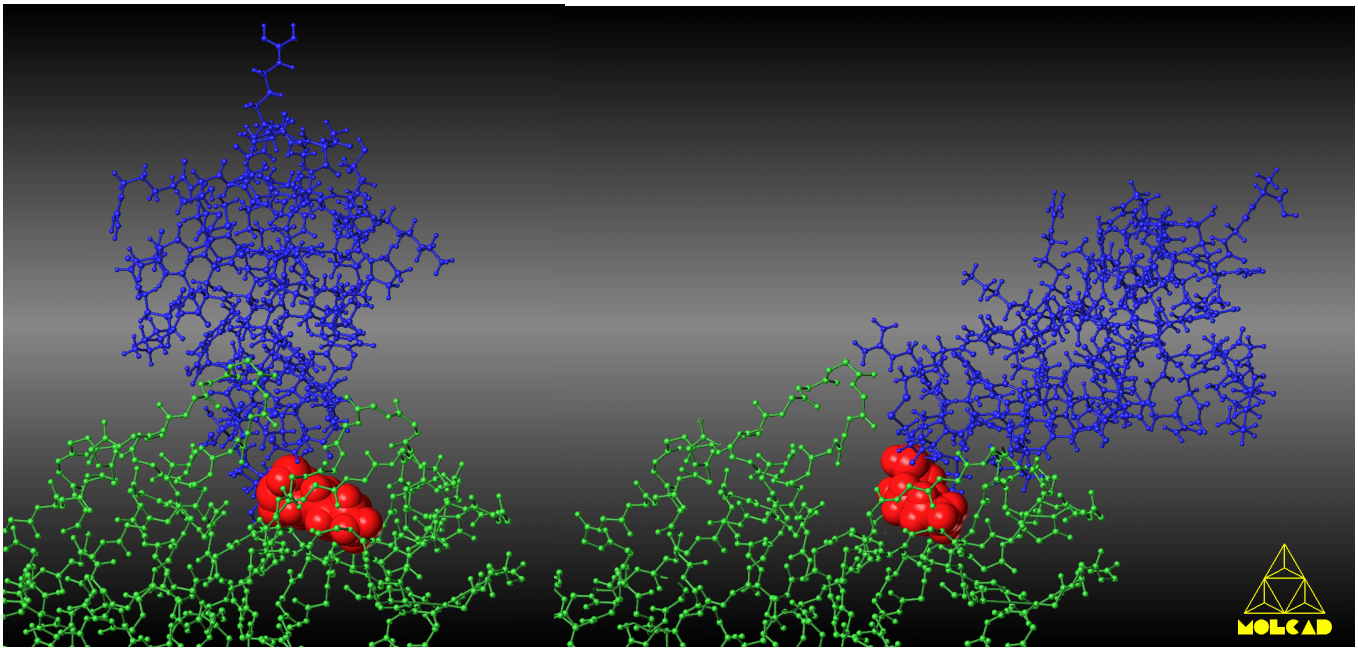


Figure 2. *Docking configuration of the HLE - ovomucoid complex (balls and sticks). Green: backbone of HLE from PDB-file 1PPF. Blue: configuration of the ovomucoid inhibitor form PDB-file 1PPF. Red: predicted configuration of ovomucoid inhibitor.*



(a) (b)

Figure 3. *Docking configuration of the trypsin - PTI complex (balls and sticks). Green: backbone of trypsin. Blue: PTI. Red: binding region of PTI (amino acid Lysin 15).*
(a): configuration from PDB-file 2PTC
(b): predicted configuration of the unbounded components from the PDB-files 2PTN and 4PTI

material. The entries are 1PPF for the HLE-ovomucoid, 1CHO for α-chymotrypsin-ovomucoid, 2PTC for trypsin-PTI complex, and 2PTN and 4PTI for trypsin and PTI, respectively.

References

1. Oshiro, C.M.; Kuntz, I.D. and Dixon, J.S. *J. Comp.-Aided Mol. Des.* **1995**, *9,* 113-130.
2. Gschwend, D.A. and Kuntz, I.D. *J. Comp.-Aided Mol. Des.* **1996**, *10,* 123-132.
3. Helmer-Citterich, M. and Tramontano, A. *J. Mol. Biol.* **1994**, *235,* 1021-1031.
4. Connolly, M.L. *Biopolymers* **1986**, *25,* 1229-1247.
5. Lin, S.L.; Nussinov, R.; Fischer, D. and Wolfson, H.J. *Proteins* **1994**, *18,* 94-101.
6. Norel, R.; Lin, S.L.; Wolfson, H.J. and Nussinov, R. *Biopolymers* **1994**, *34,* 933-940.
7. Zadeh, L.A. *Fuzzy sets; Information and Control* **1965**, *8,* 338-353.
8. Connolly, M.L. *Science* **1983**, *221,* 709-713.
9. Zimmermann, H.-J. *Fuzzy Set Theory and Its Applications*; second ed.*;* Kluwer: Boston, **1991**.
10. Brickmann, J. in *Fuzzy Logic in Chemistry*: Rouvray, D. H., Ed.; Academic: San Diego, **1997**.
11. Zachmann, C.-D.; Heiden, W.; Schlenkrich, M. and Brickmann, J. *J. Comp. Chem.* **1992**, *1,* 76-84.
12. Heiden, W. and Brickmann, J. *J. Mol. Graphics* **1994**, *12,* 106-115.