Field Interaction and Geometrical Overlap: A New Simplex and Experimental Design Based Computational Procedure for Superposing Small Ligand Molecules

Fabrizio Melani, Paola Gratteri,* Michele Adamo, and Claudia Bonaccini

Department of Pharmaceutical Sciences, University of Florence, Via G. Capponi 9, I-50121 Firenze, Italy

Received October 15, 2002

Alignment of molecules is a crucial and time-consuming step in any 3D-QSAR study. For this reason, the field interaction and geometrical overlap (FIGO) procedure presented in this paper is particularly relevant because it can provide an objective and automatic superposition of ligands through the computation of an appropriate alignment index (AI). Ligand overlay takes place via a simplex optimization of the AI function. Experimental design strategies (full factorial design, D-optimal design) are used to define the starting positions of the superposing molecules. Overlay experiments are carried out to test the performance of the method. Comparison between the results obtained with FIGO and known ligand−receptor X-ray crystallographic data (Protein Data Bank) suggests that FIGO is an effective and reliable computational procedure.

Introduction

Drug design can be approached in one of two different ways: receptor-based or nonreceptor-based. The receptor-based or direct approach depends on the availability of the three-dimensional structure of the target protein in order to simulate in silico and to evaluate the most favorable conditions for interaction of ligands with their binding site. Obviously, the availability of this structure contributes much to the drug design process. The structure of the ligand can be modified in order to find the suitable steric and electronic complementarity with the target protein and to discover the specific interactions necessary or that might improve the binding of ligands to the target.

However, in most cases the 3D structure is not known and rational drug design must be achieved in a nonreceptor or indirect way by developing an empirical model to describe the 3D structure−activity relationships (3D-QSAR) for a data set of bioactive compounds.

To set up a 3D-QSAR model, a suitable set of molecules must be selected. Each molecule is described by proper figures to characterize it effectively. Over the past decade, so-called grid-field descriptors have often been used to identify three-dimensional chemical structures. These are obtained by defining a three-dimensional grid around the ligands and calculating interaction energy between each ligand and an interacting partner (probe) at each grid point (node). When grid-fields are used to quantitatively compare a set of molecules in order to develop a 3D-QSAR model, care must be taken to properly orientate them before calculating the descriptors. Different ligand orientations lead to different values of grid-fields for a given grid node. Therefore, it is evident that 3D descriptors must be derived in a way that is equivalent to the relative spatial arrangement of molecules in the training set (TrS) at the protein binding site. The molecules must be laid on the grid according to their original orientation in the receptor cavity so that the interaction between the ligand and the binding site is optimized. When these conditions are fulfilled, the differences in the ligands bioactivity can be analyzed.

Molecular alignment is often a critical and time-consuming step in 3D-QSAR calculations. When no X-ray crystallographic data are available for ligand−receptor complexes, the researcher has no guidelines and must rely only on his own personal experience for dealing with molecular superpositioning. Therefore, a method providing reliability and speed to this operation would constitute an important improvement to the 3D-QSAR procedure. Recently, examples of alignment-independent applications for developing a virtual receptor site (VRS) have been presented.[1−4] They use a new and promising class of three-dimensional, superposition-free molecular descriptors, GRINDs (grid-independent descriptors),[5] derived from molecular interaction fields (MIFs) produced by the GRID program.[6,7] However, these methods are not yet widely established and most of the descriptors used today in developing 3D-QSAR models require ligands to be preliminarily superposed. For these reasons, in the present study a methodology is presented whereby an automatic alignment procedure is developed to overcome the inconvenience of a purely subjective evaluation. Many recent scientific contributions have been focused on methods for the structural alignments of molecules.[8−33] Most of them have been reviewed by Lemmen and Lengauer,[34] while others have been subsequently presented.[35−38] These proposed methods for small molecule superposition are summarized in Table 1 and mainly differ in the treatment of conformational flexibility, the optimization algorithm used, and the definition of molecular similarity.

Field interaction and geometrical overlap (FIGO) is a new approach whereby the alignment of the molecules occurs via simplex optimization through the superposi-

* To whom correspondence should be addressed. Phone: +39-055-2757303. Fax: +39-055-240776. E-mail: paola.gratteri@unifi.it.

10.1021/jm0210616 CCC: $25.00
Published on Web 03/12/2003

Table 1. Main Recent-Literature Contributions to Computational Methods for the Structural Alignment of Compounds

authors	program	similarity criteria	optimization algorithm/ superposition methodology	superposition mode	ref
Cocchi, M. and De Benedetti, P. G.		molecular electrostatic potential (MEP), size, and shape descriptors	simplex	rigid	8
Cossè-Barbi, A. and Raji, M.		pattern in 3D space	stepwise approach	rigid	9
Dammkoehler et al.	AAA	distance map of pharmacophoric points	combinatorial search	flexible	10
de Caceres et al.	MIPSIM	molecular electrostatic potential	gradient method	semiflexible	32
De Rosa et al.		Euclidean distance in Hi-PCA space		rigid	11
Goldman, B. B. and Wipke, W. T.		geometrically invariant molecular surface descriptors			36
Good et al.	ASP	MEP (Hodgkin function)	simplex	rigid	12
Grant et al.		van der Waals volume described by Gaussian function	analytic first and second derivatives	rigid	13
Handschuh et al.		geometric fit (distance and stereochemical parameters)	genetic algorithm and quasi Newton method	flexible	14
Itai et al.	AUTOFIT	pharmacophoric points	combinatorial matching	flexible	15
Jain et al.	COMPASS	surface description	neural network	semiflexible	16
Jones et al.	GASP	intermolecular conformational energy, volume overlay, intermolecular matching energy	genetic algorithm	flexible	17
Kearsley, S. K. and Smith, G. M.	SEAL	electrostatic and steric terms	rational function optimization (RFO)	rigid	18
Klebe et al.		steric, electrostatic, hydrophobic, and hydrogen-bond interaction fields described by Gaussian function	quasi Newton method	semiflexible	19
Labute et al.		atom properties	modified RIPS (random incremental pulse search) procedure	flexible	38
Lemmen et al.	RigFit	steric occupancy, partial atomic charge, hydrophobicity, hydrogen bond potential described by Gaussian function	quasi Newton method	rigid	20
Lemmen, C. and Lengauer, T.	FLEXS	interaction fields described by Gaussian function	combinatorial matching procedure	flexible	21
Martin et al.	DISCO	pharmacophoric points	combinatorial matching procedure	semiflexible	22
Masek et al.	MSC	physicochemical properties	BFGS (Broyden, Fletcher, Goldfarb, Shanno)	semiflexible	23
McMahon, A. and King, P. M.		electrostatic potential described by Gaussian function	gradient method	rigid	24
McMartin, C. and Bohacek, R. S.	TFIT	inter- and intramolecular energy	Monte Carlo and line search procedure	flexible	25
Mestres et al.	MIMIC	steric and electrostatic fields described by Gaussian function	steepest descent or Newton−Raphson method	semiflexible	26
Miller et al.	SQ	SQ type	simplex	semiflexible	27
Mills et al.	SLATE	distance matrix for H-bonding and aromatic properties	simulated annealing	flexible	35
Nissink et al.	QUASIMODI	electron density described by Gaussian function	simplex	rigid	28
Parretti et al.		steric and electrostatic fields described by Gaussian function	Monte Carlo	rigid	29
Perkins et al.	PLM	surface overlap volume	simulated annealing	semiflexible	30
Petitjean et al.		electronic properties and protonic charge	gradient method	rigid	31
Pitman et al.	FLASHFLOOD	comma descriptors	cluster method	flexible	37
Sheridan et al.		distance geometry of pharmacophore		flexible	33

tion of both MIFs of a set of compounds and the heavy atoms (no hydrogen) of their chemical skeleton. The innovation of the method is the use of the GRID program[6,7] to define the MIF values using several probes including, most importantly, donor- and acceptor-hydrogen-bonding ones. Moreover, experimental design strategies are used in order to properly define the number and initial positions of the vertexes in the simplex. These are the unique aspects of FIGO that successfully combine exploitation of GRID energy function and the experimental design methodology. Although field-fit procedures are well-known in the overlaying of small molecules, the remarkable effectiveness of the force field used by the GRID program has never been applied in this way. GRID program characterizes the superposing molecules with an energy function that includes Lennard Jones (E_{lj}) and electrostatic (E_{el}) terms, as well as an explicit hydrogen bond term (E_{hb}) dependent on the length of the hydrogen bond, the orientation of the probes at the hydrogen-bonding atoms, and the chemical identity of the hydrogen-bonding atoms. Moreover, the use of experimental design strategy to identify the starting positions of the superposing molecules is a new feature in this area that allows superposing molecules to efficiently span the experimental domain of variables (translation and rotation movements of molecules) and enables the modified simplex optimization method[39] to approach the optimum of the studied response.

The reliability of FIGO alignments was tested by comparing the results obtained with X-ray crystallography data for the ligand−receptor complexes available from the Protein Data Bank.[40,41]

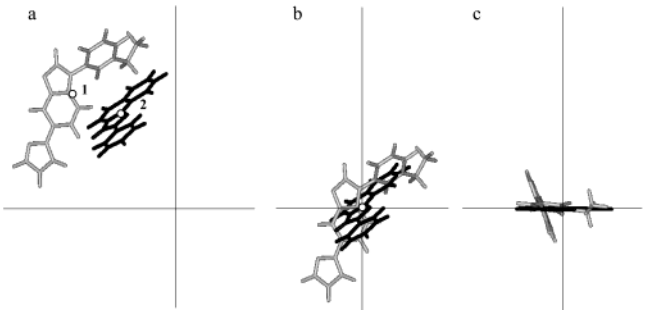


Figure 1. Prealignment steps: (a) starting position of two molecules to overlay , where points 1 and 2 are, respectively, the geometrical centers for the gray and black structures; (b) translation of molecular geometrical centers onto the origin (0, 0, 0); (c) rotation of the molecules to meet the average molecular plane.

Method

FIGO methodology consists of two fundamental steps.

I. Prealignment Procedure. First, the geometrical center of each molecule, i.e., the point that has as coordinates the mean of the coordinates for any atom of the molecule, is moved to coincide with the point (0, 0, 0). Points 1 and 2 in Figure 1a are the geometrical mass centers for the gray and black structures, respectively. The two molecules are depicted before (Figure 1a) and after (Figure 1b) the translation took place.

Second, all molecules are rotated along axes *X*, *Y*, and *Z* so that each of their average molecular plane is made to coincide with the *XZ* plane (Figure 1c). The average molecular plane is define as the plane where the sum of the distances of the molecule atoms is the shortest from the plane.

II. Alignment Sequence. To proceed in this phase, we defined a new alignment index AI, calculated by the formula

$$\mathrm{AI} = \underbrace{\sum_{p=1}^{N}\left(\sum_{\mu=1}^{N_{\mathrm{rif}}}\sum_{\nu=1}^{N_{\mathrm{mol}}}\frac{1}{1+r_{\mu\nu}}e^{-\left|\mathrm{Vrif}_{\mu}-\mathrm{Vmol}_{\nu}\right|}\right)}_{\text{Term } a \text{ relative to probes used}} + \underbrace{\sum_{m=1}^{A_{\mathrm{rif}}}\sum_{n=1}^{A_{\mathrm{mol}}}\frac{1}{1+r_{mn}}w_{\mathrm{a}}}_{\text{Term } b \text{ relative to the heavy atoms}} \quad (1)$$

where definitions are the following.

N: number of probes

N_{rif}: number of selected nodes of the reference structure

N_{mol}: number of selected nodes of the superposing structure

$p = 1 \ldots N$: the pth probe used

$\mu = 1 \ldots N_{\mathrm{rif}}$: the μth node for the reference structure (rif)

$\nu = 1 \ldots N_{\mathrm{mol}}$: the νth node for the superimposed molecule (mol)

Vrif_{μ}: MIF relative to the μth node for the reference structure (rif)

Vmol_{ν}: MIF relative to the νth node for the superposing molecule (mol)

$r_{\mu\nu}$: distance between the μth node of the reference structure and the νth node of the superposing molecule

A_{rif}: number of heavy atoms of the reference structure

A_{mol}: number of heavy atoms of the superposing molecule

$m = 1 \ldots A_{\mathrm{rif}}$: heavy m atom in the reference structure (rif)

$n = 1 \ldots A_{\mathrm{mol}}$: heavy n atom in the superposing molecule (mol)

w_{a}: weight factor of the term b

r_{mn}: distance between the heavy m atom in the reference structure and heavy n atom in the superposed molecule. m and n are characterized by the same atom type.

AI is a measurement of the similarity between structures, and it is formed by two terms: *a* and *b*. The former refers to the different MIFs computed for the superposition, and the latter refers to the heavy atoms of the compounds to be aligned. Thus, as a result of term *a*, the alignment occurs through the superposition of MIFs of equivalent nature, i.e., deriving from acceptor- or donor-hydrogen-bonding or hydrophobic probes, whereas the term *b* focuses on the atom types of the overlaying molecules.

The GRID program is used to calculate MIFs. MIFs represent specific noncovalent interactions calculated for a small chemical group (probe) interacting with a target (e.g., ligands). Ligands are put in a three-dimensional grid (Figure 2a−d), the probe is moved to each grid point (node), and the interaction energy between it and the target is computed. Thus, each grid point is associated with an energy value.

More than one probe can be used for the computation of term *a* of AI. For each probe used, only a limited number of nodes per molecule, equal to the number of its heavy atoms, are considered (parts e and f of Figure 2). The selected nodes are all characterized by negative MIFs values representing attractive interactions. The importance of a node is judged on the basis its MIF values, the larger negative value representing a more significant node. In fact, these nodes are areas where strong molecular interactions with the target occur. Clearly, the value assumed by MIF strongly depends on the kind of probe used. For each probe, the sum of the pairwise difference of the MIFs between each node of molecule **1** (the reference, for example) and each node of molecule **2** (the superposing compound) is computed. The exponential trend of these differences has the effect of enhancing AI value when the differences between MIFs decrease, so nodes with similar MIFs will tend to coincide. On the other hand, the distance *r* between compared nodes plays a crucial role for a good alignment. Two nodes with similar MIFs but far from each other give rise to poor AI, whereas nodes with similar MIFs but near each other lead to good AI. In other word, term *a* of eq 1 leads to privileged situations where nodes with similar MIFs are placed near each other rather than to cases where the same nodes are distant from each other.

The presence of the term *b* in the AI formula is particularly important when functional groups responsible for specific interactions are not present in the overlaying molecules. In these cases MIFs could be insufficient to achieve a good alignment; however, by considering the number and the nature of the heavy atoms, this can be obtained. Moreover, term *b* enables information concerning hindrance and shape of the molecules to be condensed into a few points (corresponding to the number of the heavy atoms). This

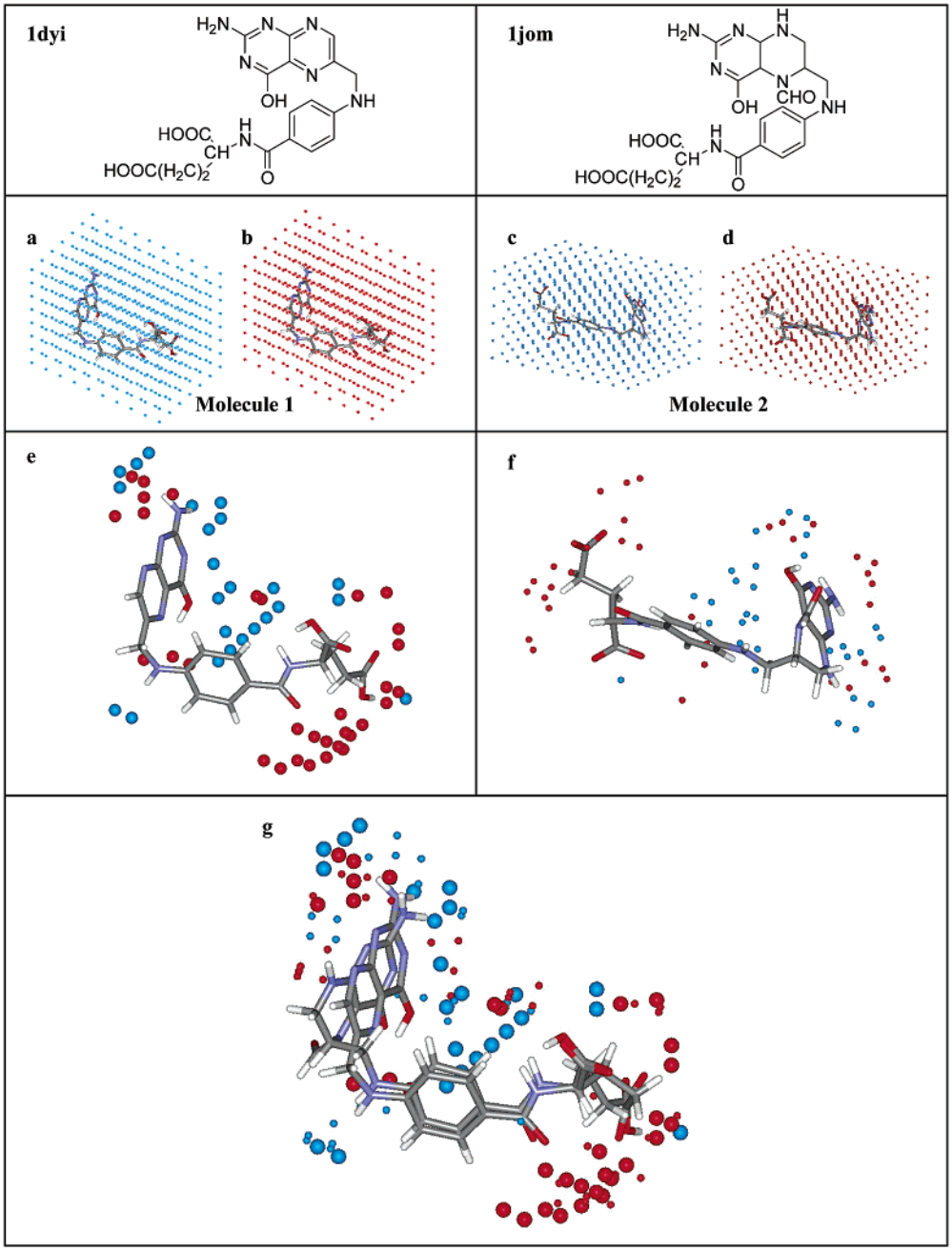


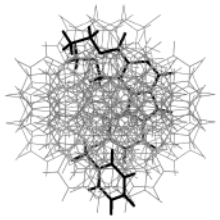
Figure 2. Steps of alignment index (AI) calculation in the superposition of molecules **1** and **2** (PDB code 1dyi and 1jom, respectively). Big and little dots belong to molecular interaction fields (MIFs) of molecules **1** and **2**, respectively. (a) MIFs for molecule **1** computed with an acceptor-hydrogen-bonding probe (O::); (b) MIFs for molecule **1** computed with a donor-hydrogen-bonding probe (N2═); (c) MIFs for molecule **2** computed with an acceptor-hydrogen-bonding probe (O::); (d) MIFs for molecule **2** computed with a donor-hydrogen-bonding probe (N2═); (e) selected nodes (equal to the number of heavy atoms in the molecule) for molecule **1** relative to O:: (blue) and N2═ (red) probes; (f) selected nodes (equal to the number of heavy atoms in the molecule) for molecule **2** relative to O:: (blue) and N2═ (red) probes; (g) final FIGO alignment obtained for molecules **1** and **2**.

signifies that all positive MIFs values in term *a* do not have to be considered. Undoubtedly, these values include steric information spread uniformly around the molecule; however, their inclusion would substantially increase calculation times. The summation of term *b* refers to the sum of the pairwise distance between each heavy atom of molecule **1** and each heavy atom of molecule **2** of Figure 2. It is multiplied by a constant factor w_a, which is the only parameter that can be tuned, and shows the influence that term *b* has on term *a* in AI computation. The procedure sets $w_a = 2$ by default. Figure 2g shows the final alignment obtained for molecules **1** and **2**.

The AI index defined in eq 1 is used as a score function in the optimization process in order to determine the best superposition of the members of the data set onto a molecule chosen as reference. The alignment is readily accomplished through the optimization (maximization) of the AI, which occurs by a simplex procedure according to the Nelder and Mead algorithm.[39]

For the alignment procedure, the most active member of the data set is chosen as the reference. This molecule is kept in the same position during the whole procedure, while the rest of the members are superimposed one at a time in such a way that for each data set member AI is maximized.

Scheme 1. 2^6 FFD + (3×2^3) FFD Experimental Plan and Graphic Representation of the Reference Molecule (Black) and the Whole Set of the 57 Starting Positions for the Superposing Molecule[a]



Exp No	Rx	Ry	Rz	Tx	Ty	Tz
1	-90	-90	-90	-1	-1	-1
2	90	-90	-90	-1	-1	-1
3	-90	90	-90	-1	-1	-1
4	90	90	-90	-1	-1	-1
5	-90	-90	90	-1	-1	-1
6	90	-90	90	-1	-1	-1
7	-90	90	90	-1	-1	-1
8	90	90	90	-1	-1	-1
9	-90	-90	-90	1	-1	-1
10	90	-90	-90	1	-1	-1
11	-90	90	-90	1	-1	-1
12	90	90	-90	1	-1	-1
13	-90	-90	90	1	-1	-1
14	90	-90	90	1	-1	-1
15	-90	90	90	1	-1	-1
16	90	90	90	1	-1	-1
17	-90	-90	-90	-1	1	-1
18	90	-90	-90	-1	1	-1
19	-90	90	-90	-1	1	-1
20	90	90	-90	-1	1	-1
21	-90	-90	90	-1	1	-1
22	90	-90	90	-1	1	-1
23	-90	90	90	-1	1	-1
24	90	90	90	-1	1	-1
25	-90	-90	-90	1	1	-1
26	90	-90	-90	1	1	-1
27	-90	90	-90	1	1	-1
28	90	90	-90	1	1	-1
29	-90	-90	90	1	1	-1
30	90	-90	90	1	1	-1
31	-90	90	90	1	1	-1
32	90	90	90	1	1	-1
33	-90	-90	-90	-1	-1	1
34	90	-90	-90	-1	-1	1
35	-90	90	-90	-1	-1	1
36	90	90	-90	-1	-1	1
37	-90	-90	90	-1	-1	1
38	90	-90	90	-1	-1	1
39	-90	90	90	-1	-1	1
40	90	90	90	-1	-1	1
41	-90	-90	-90	1	-1	1
42	90	-90	-90	1	-1	1
43	-90	90	-90	1	-1	1
44	90	90	-90	1	-1	1
45	-90	-90	90	1	-1	1
46	90	-90	90	1	-1	1
47	-90	90	90	1	-1	1
48	90	90	90	1	-1	1
49	-90	-90	-90	-1	1	1
50	90	-90	-90	-1	1	1
51	-90	90	-90	-1	1	1
52	90	90	-90	-1	1	1
53	-90	-90	90	-1	1	1
54	90	-90	90	-1	1	1
55	-90	90	90	-1	1	1
56	90	90	90	-1	1	1
57	-90	-90	-90	1	1	1
58	90	-90	-90	1	1	1
59	-90	90	-90	1	1	1
60	90	90	-90	1	1	1
61	-90	-90	90	1	1	1
62	90	-90	90	1	1	1
63	-90	90	90	1	1	1
64	90	90	90	1	1	1
65	**0**	**0**	**0**	**0**	**0**	**0**
66	180	0	0	-1	-1	-1
67	180	0	0	1	-1	-1
68	180	0	0	-1	1	-1
69	180	0	0	1	1	-1
70	180	0	0	-1	-1	1
71	180	0	0	1	-1	1
72	180	0	0	-1	1	1
73	180	0	0	1	1	1
74	0	180	0	-1	-1	-1
75	0	180	0	1	-1	-1
76	0	180	0	-1	1	-1
77	0	180	0	1	1	-1
78	0	180	0	-1	-1	1
79	0	180	0	1	-1	1
80	0	180	0	-1	1	1
81	0	180	0	1	1	1
82	0	0	180	-1	-1	-1
83	0	0	180	1	-1	-1
84	0	0	180	-1	1	-1
85	0	0	180	1	1	-1
86	0	0	180	-1	-1	1
87	0	0	180	1	-1	1
88	0	0	180	-1	1	1
89	0	0	180	1	1	1

1 FFD 2^3 | 2 FFD 2^3 | 3 FFD 2^3

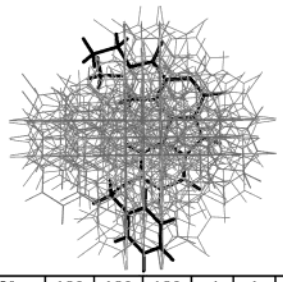
[a] The elided starting positions leading to equal orientations are in gray (compare exp no. 9 equal to exp no. 14). The center of the experimental domain is in bold.

The number and initial positions of the vertexes in the simplex are determined by statistical experimental design. The reason for this is to explore the whole three-dimensional space with a minimum number of starting positions and, as a consequence, a minimum number of simplex vertexes. In this work we refer to two strategies for determining the simplex vertexes: the full factorial (FFD) and the D-optimal designs (Schemes 1 and 2).[42,43]

Scheme 1 shows the 57 (N_1) starting points arising from the fusion of two full factorial designs: (a) a 2^6 FFD in two levels and six variables (three rotations along *X*, *Y*, and *Z* and three translations with respect to *X*, *Y*, and *Z*; the experimental domain for the rotational and translational variables varies in the range of +90° to −90° and from +1 to −1 Å respectively); (b) three 2^3 FFD in two levels and three variables (the translational movements along *X*, *Y*, and *Z* keep the rotational variables constant in each of the three 2^3 FFD). The starting positions having equal orientations have been elided.

Scheme 2 highlights the 113 (N_2) starting positions selected from the 1729 candidate points derived from a multilevel six-variable factorial design. According to the D-optimality criterion,[43] this is the best subset. Table 3 provides the levels and the experimental domain for each variable in Scheme 2. Moreover, 2^{6-2} fractional

Scheme 2. Experimental Plan, D-Optimal Design Statistical Parameters, and Graphic Representation of the Reference Molecule (Black) and the 113 Starting Positions of the Superposing Molecule[a]



Exp No	rx	ry	rz	tx	ty	tz
1	-90	-90	-90	-1	-1	-1
2	180	-90	-90	-1	-1	-1
3	180	-90	-90	-1	-1	-1
4	90	90	-90	-1	-1	-1
5	-90	180	-90	-1	-1	-1
6	180	180	-90	-1	-1	-1
7	180	-90	0	-1	-1	-1
8	180	-90	90	-1	-1	-1
9	-90	-90	180	-1	-1	-1
10	0	-90	180	-1	-1	-1
11	-90	180	180	-1	-1	-1
12	-90	180	180	-1	-1	-1
13	180	180	180	-1	-1	-1
14	180	180	-90	0	-1	-1
15	-90	90	180	0	-1	-1
16	0	180	180	0	-1	-1
17	-90	-90	-90	1	-1	-1
18	0	-90	-90	1	-1	-1
19	-90	90	-90	1	-1	-1
20	90	180	-90	1	-1	-1
21	180	180	-90	1	-1	-1
22	-90	-90	0	1	-1	-1
23	-90	180	90	1	-1	-1
24	-90	-90	180	1	-1	-1
25	180	-90	180	1	-1	-1
26	180	-90	180	1	-1	-1
27	180	-90	180	1	-1	-1
28	-90	180	180	1	-1	-1
29	180	180	180	1	-1	-1
30	-90	0	-90	-1	0	-1
31	180	-90	180	-1	0	-1
32	-90	-90	-90	0	0	-1
33	180	90	0	0	0	-1
34	-90	-90	-90	1	0	-1
35	180	180	-90	1	0	-1
36	180	180	180	1	0	-1
37	-90	-90	-90	-1	1	-1
38	180	-90	-90	-1	1	-1
39	-90	180	-90	-1	1	-1
40	-90	180	-90	-1	1	-1
41	180	180	-90	-1	1	-1
42	180	180	-90	-1	1	-1
43	-90	-90	90	-1	1	-1
44	-90	180	90	-1	1	-1
45	-90	-90	180	-1	1	-1
46	180	-90	180	-1	1	-1
47	-90	0	180	-1	1	-1
48	180	180	180	-1	1	-1
49	180	180	180	-1	1	-1
50	180	180	180	-1	1	-1
51	0	-90	-90	0	1	-1
52	90	-90	180	0	1	-1
53	180	-90	-90	1	1	-1
54	180	-90	-90	1	1	-1
55	180	-90	-90	1	1	-1
56	-90	180	-90	1	1	-1
57	-90	180	-90	1	1	-1
58	-90	180	-90	1	1	-1
59	180	90	0	1	1	-1
60	180	180	90	1	1	-1
61	-90	-90	180	1	1	-1
62	-90	-90	180	1	1	-1
63	180	0	180	1	1	-1
64	-90	180	180	1	1	-1
65	0	180	180	1	1	-1
66	-90	180	-90	-1	-1	0
67	180	180	180	-1	-1	0
68	-90	-90	-90	0	-1	0
69	-90	-90	180	0	-1	0
70	90	0	-90	1	-1	0
71	180	180	-90	1	-1	0
72	180	-90	180	1	-1	0
73	-90	180	0	-1	0	0
74	180	-90	180	-1	0	0
75	90	90	90	0	0	0
76	-90	180	180	1	0	0
77	180	0	-90	-1	1	0
78	0	-90	0	-1	1	0
79	0	90	180	-1	1	0
80	90	180	-90	0	1	0
81	180	-90	-90	1	1	0
82	-90	-90	90	1	1	0
83	-90	180	180	1	1	0
84	180	180	180	1	1	0
85	-90	-90	-90	-1	-1	1
86	0	-90	-90	-1	-1	1
87	180	-90	-90	-1	-1	1
88	180	0	-90	-1	-1	1
89	-90	180	-90	-1	-1	1
90	180	180	-90	-1	-1	1
91	-90	180	0	-1	-1	1
92	-90	-90	180	-1	-1	1
93	-90	-90	180	-1	-1	1
94	0	-90	180	-1	-1	1
95	180	90	180	-1	-1	1
96	180	180	180	-1	-1	1
97	180	180	180	-1	-1	1
98	-90	180	-90	0	-1	1
99	180	-90	90	0	-1	1
100	180	0	180	0	-1	1
101	-90	-90	-90	1	-1	1
102	180	-90	-90	1	-1	1
103	180	-90	-90	1	-1	1
104	-90	180	-90	1	-1	1
105	180	180	-90	1	-1	1
106	-90	0	0	1	-1	1
107	180	180	0	1	-1	1
108	-90	-90	180	1	-1	1
109	90	-90	180	1	-1	1
110	-90	180	180	1	-1	1
111	-90	180	180	1	-1	1
112	0	180	180	1	-1	1
113	180	180	180	1	-1	1
114	90	180	-90	-1	0	1
115	90	180	90	-1	0	1
116	-90	-90	180	-1	0	1
117	180	-90	-90	0	0	1
118	180	-90	-90	1	0	1
119	-90	180	-90	1	0	1
120	90	-90	90	1	0	1
121	-90	0	180	1	0	1
122	-90	-90	-90	-1	1	1
123	-90	-90	-90	-1	1	1
124	180	-90	-90	-1	1	1
125	-90	0	-90	-1	1	1
126	180	180	-90	-1	1	1
127	180	180	-90	-1	1	1
128	180	-90	90	-1	1	1
129	180	-90	180	-1	1	1
130	180	-90	180	-1	1	1
131	-90	180	180	-1	1	1
132	-90	180	180	-1	1	1
133	-90	180	180	-1	1	1
134	-90	90	-90	0	1	1
135	180	180	90	0	1	1
136	-90	-90	180	0	1	1
137	-90	180	180	0	1	1
138	-90	-90	-90	1	1	1
139	-90	-90	-90	1	1	1
140	180	90	-90	1	1	1
141	-90	180	-90	1	1	1
142	0	180	-90	1	1	1
143	180	180	-90	1	1	1
144	0	0	90	1	1	1
145	-90	-90	180	1	1	1
146	180	-90	180	1	1	1
147	180	-90	180	1	1	1
148	180	180	180	1	1	1
149	180	180	180	1	1	1
150	180	180	180	1	1	1

D-optimal design statistic			
G-Efficiency	log(Det. of X'X)	Norm. log(Det. of X'X)	Condition Number
75.5892	45.7383	-0.0941748	1.18832

[a] The elided starting positions leading to equal orientations are in gray (compare exp nos. 2 and 3 equal to exp no. 9).

factorial designs were centered on each starting position given by both FFD (N_1) and D-optimal (N_2) designs. The addition of further orientations [$2^{6-2} = 16 \times (N_1 \text{ or } N_2)$] to the already great number of N_1 or N_2 reduces the risk of falling into a local maximum of the AI function and consequently results in a better exploration of the AI function.

The simplex starting points can be seen as orientations assumed by the superposing molecule in the surrounding space. Therefore, each vertex of the sim-

Table 2. List of Ligands Binding to the Same Protein Used in This Study

	PDB code	source	resolution (Å)		PDB code	source	resolution (Å)
		Trypsin				**Dihydrofolate Reductase**	
1	1tnl	bovine (bos taurus) pancreas	1.9	1	1jom	*Escherichia coli*	1.9
2	1tnk	bovine (bos taurus) pancreas	1.8	2	1dds	*Escherichia coli*	2.2
3	1tnj	bovine (bos taurus) pancreas	1.8	3	1dyj	*Escherichia coli*	1.85
4	1tni	bovine (bos taurus) pancreas	1.9	4	1dyi	*Escherichia coli*	1.9
5	1tnh	bovine (bos taurus) pancreas	1.8				
6	3ptb	bovine (bos taurus) pancreas	1.7			**α-Thrombin**	
				1	1ppb	human (Homo sapiens) plasma	1.92
		Steroid		2	1dwd	human (Homo sapiens) plasma	3
1	1dbb	mouse (mus musculus) hybridoma	2.7				
2	1dbj	mouse (mus musculus) hybridoma	2.7			**Endothiapepsin**	
3	1dbk	mouse (mus musculus) hybridoma	3	1	4er1	chestnut blight fungus (endothia parasitica)	2
4	2dbl	mouse (mus musculus) hybridoma	2.9	2	1er8	chestnut blight fungus (endothia parasitica)	2
5	1dbm	mouse (mus musculus) hybridoma	2.7	3	1epo	chestnut blight fungus (endothia parasitica)	2
				4	2er6	chestnut blight fungus (endothia parasitica)	2
		Penicillopepsin		5	1ent	chestnut blight fungus (endothia parasitica)	1.9
1	1apu	fungus (penicillium janthinellum)	1.8	6	3er5	chestnut blight fungus (endothia parasitica)	1.8
2	1apv	fungus (penicillium janthinellum)	1.8	7	3er3	chestnut blight fungus (endothia parasitica)	2
3	1apw	fungus (penicillium janthinellum)	1.8	8	1epq	chestnut blight fungus (endothia parasitica)	1.9
4	1ppk	fungus (penicillium janthinellum)	1.8	9	1epp	chestnut blight fungus (endothia parasitica)	1.9
5	1ppl	fungus (penicillium janthinellum)	1.7	10	1epn	chestnut blight fungus (endothia parasitica)	1.6

Table 3. Factors, Experimental Domain, and Levels for the First-Phase Strategy 2

variables	levels and experimental domain	unit
rx	−90, 0, 90, 180	deg
ry	−90, 0, 90, 180	deg
rz	−90, 0, 90, 180	deg
tx	−1, 0, 1	Å
ty	−1, 0, 1	Å
tz	−1, 0, 1	Å

plex, as well as each orientation, will be associated with a value for the AI. After the simplex run, the value calculated for AI will be maximized as well as the overlap between the superposing molecule and the reference, according to the criteria defining AI.

The program flowchart is shown in Chart 1, while Chart 2 outlines the whole process.

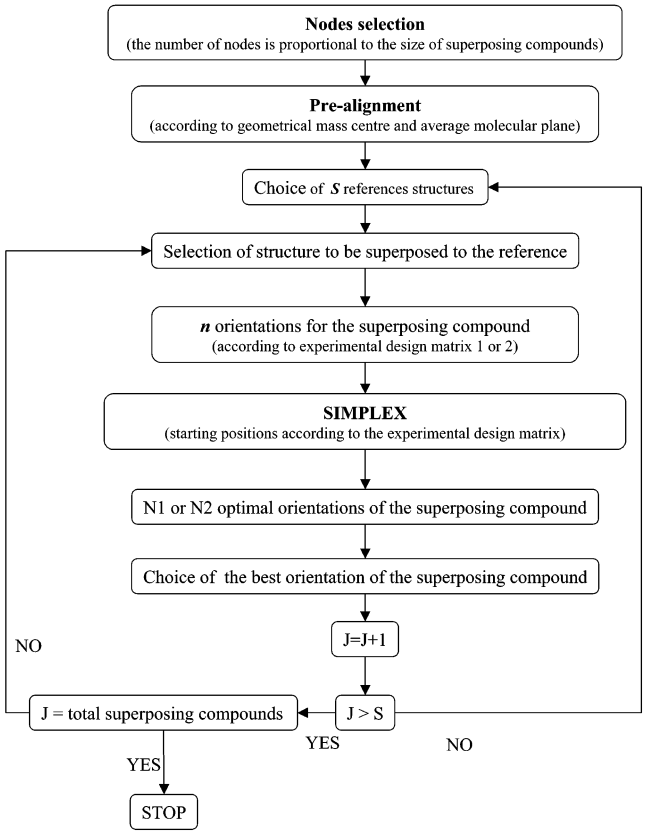
Results and Discussion

FIGO methodology steps in to determine the best superposition between two molecules, according to the criteria defining the alignment index AI. Nevertheless, the procedure can also be applied to superpose a set of molecules. In these cases, a reference structure is needed for the pairwise alignment between the reference and any other molecule of the set. As a consequence of the successive pairwise alignment to the same reference structure, the overall set is clearly superposed.

The relevance of alignments performed by FIGO methodology was tested using seven well-suited data sets for which crystal structures of the receptor−inhibitor complexes were available (Table 2). Studies carried out on these data sets pointed out some crucial aspects to be carefully considered during the application of the proposed procedure. Bearing in mind the method presented in the Method section, they can be summarized by the following points: (1) choice of the experimental design for the definition of the starting orientations of the molecule to be superposed onto the reference; (2) choice of the w_a parameter (eq 1, term *b*); (3) choice of the reference structure; (4) MIF types for the calculation (eq 1, term *a*).

FIGO methodology was tested using all the ligands in the same conformation because they came from the

Chart 1. Flow Chart of FIGO Procedure



crystal structure but with different starting orientations. Each component within the testing data sets was selected one at a time to be the reference compound, and the other members of the data set were superposed. To evaluate the quality of the alignment obtained, the orientation of each ligand after the FIGO procedure (FIGO-aligned ligand) was compared to the orientation presented by the same ligand in the crystallographic structure superposition, the latter obtained using the common protein backbone atoms (crystal-aligned ligand). The lower is the root-mean-square (rms) deviation between the FIGO-aligned and the crystal-aligned ligands the better is the quality of FIGO superposition.

Chart 2. Steps of FIGO Procedure

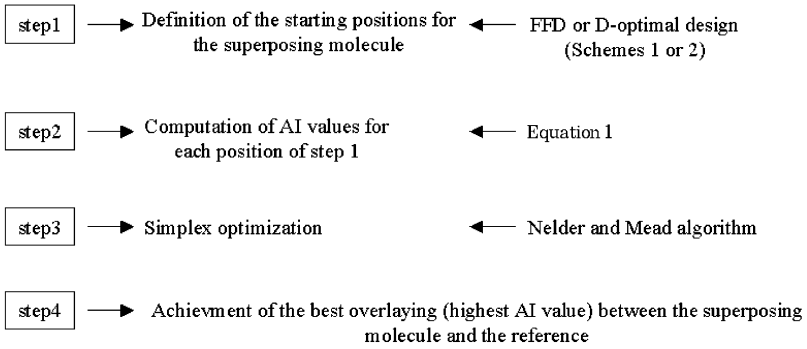
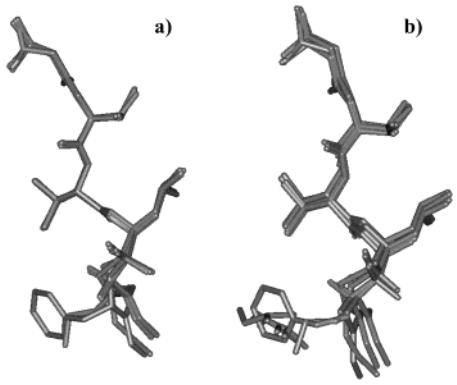


Table 4. Numerical and Graphical Results of the Penicillopepsin Data Set: (a) FIGO Alignment; (b) Alignment in the Crystal



	rms, reference 1		rms, reference 2		rms, reference 3		rms, reference 4		rms, reference 5	
	FFD	DOPT	FFD	DOPT	FFD	DOPT	FFD	DOPT	FFD	DOPT
Penicillopepsin Data Set, $w_a = 1$										
1			0.192	0.187	0.135	0.137	0.258	0.258	1.404	1.402
2	0.643	0.641			0.224	0.224	0.312	0.306	1.424	1.429
3	0.653	0.651	0.219	0.219			0.259	0.257	1.417	1.416
4	0.648	0.651	0.266	0.267	0.225	0.223			1.377	1.380
5	0.703	0.700	0.354	0.355	0.359	0.355	0.175	0.175		
average value:	0.662	0.661	0.258	0.257	0.236	0.235	0.251	0.249	1.405	1.407
Penicillopepsin Data Set, $w_a = 2$										
1			0.172	0.172	0.124	0.123	0.249	0.246	1.414	1.411
2	0.667	0.664			0.126	0.127	0.308	0.306	1.436	1.437
3	0.676	0.674	0.123	0.123			0.268	0.269	1.424	1.423
4	0.625	0.621	0.277	0.277	0.248	0.247			1.368	1.365
5	0.687	0.683	0.360	0.361	0.365	0.364	0.173	0.173		
average value:	0.663	0.660	0.233	0.233	0.216	0.215	0.250	0.249	1.411	1.409

Experimental evidence coming from the testing data sets indicates that the best results of the FIGO procedure were accomplished using the full factorial design strategy and $w_a = 2$ parameter (points 1 and 2). The data set in Table 4 refers to the application of both FFD and D-optimal strategies and two values of the w_a parameter when the penicillopepsin data set was taken as an example. As can be seen, the lowest average rms value is obtained with $w_a = 2$ and 1apw as the reference structure. In this case, both experimental design strategies lead to almost equivalent results.

As far as the reference compound within a data set is concerned, the requirements for its selection are high affinity and a reduced conformational freedom (point 3). In other words, a potent and rigid compound, or at least conformationally constrained, would be an ideal candidate as a reference structure. Whenever the reference structure presents a notable conformational freedom, it would be more appropriate to use a hyperstructure (HS) as reference. HS can be imagined as a molecule with the number of atoms and nodes equal to the sum of those in the individual molecules participating in its formation. The utilization of HS instead of a single reference structure should be considered when dealing with very structurally different compounds having similar biological activities or when no fine structural and electronic features of the receptor-binding mode are known.

MIFs employed to carry out the superposition affect the overlay results (point 4). As a general rule, MIFs derived from donor- and acceptor-hydrogen-bonding probes are always used because hydrogen bonds play a crucial role in determining the specificity of ligand−macromolecule interactions. On the other hand, when hydrophobic interactions are determinant for the binding of the ligands, it is important to use MIFs arising from hydrophobic probes (DRY, CH3). Table 5 sum-

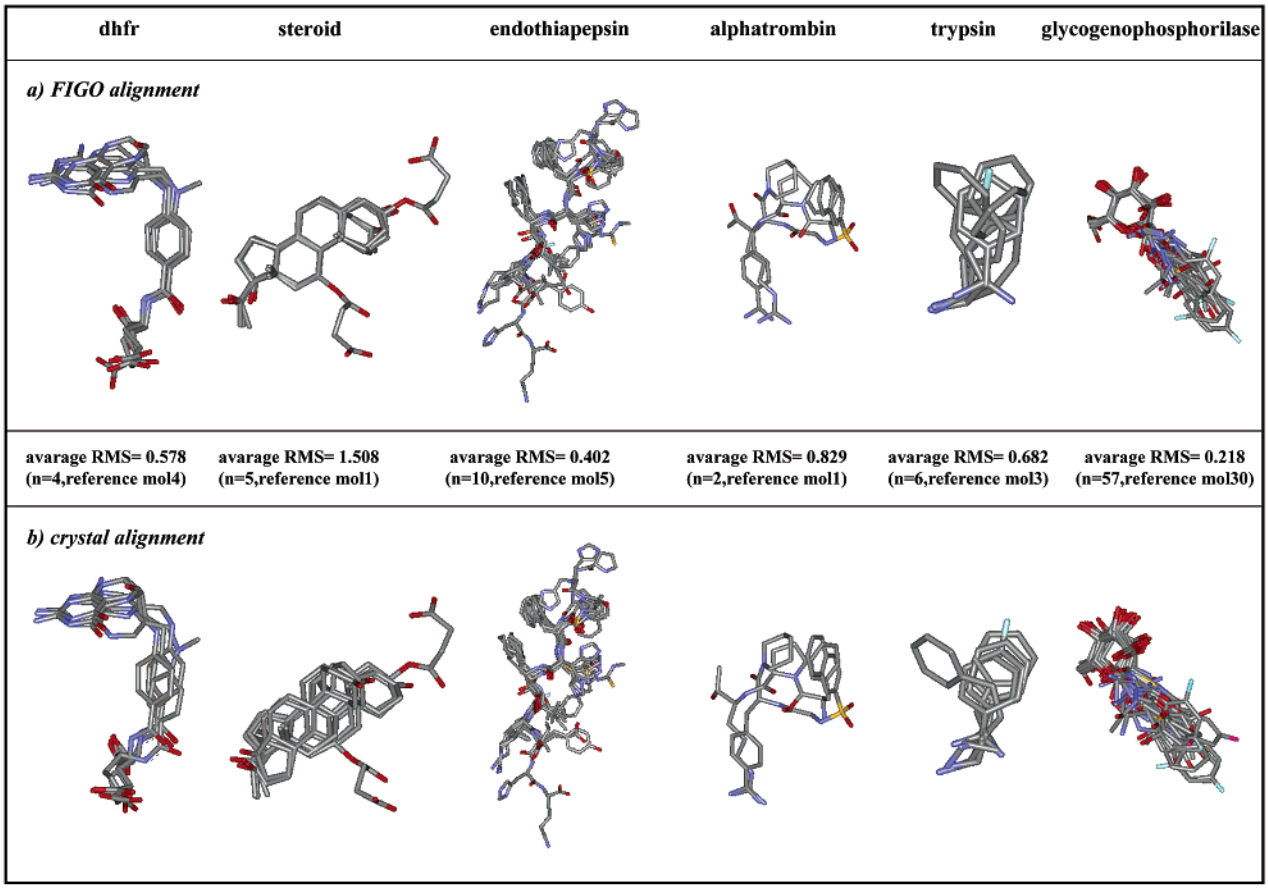


Figure 3. Superposition of molecules within each data set. The average rms deviation is also reported. (a) Best alignment calculated by FIGO; (b) alignment in the crystal.

Table 5. Some of the Probes Used in MIFs Calculations with GRID Program[6,7]

grid probes	description	no. of H bonds accepted	no. of H bonds donated	H-bond type
O	sp^2 carbonyl oxygen	2	0	sp^2 carbonyl oxygen bonded to one atom and with two lone pairs, e.g., in aldehyde or ketone or amide groups.
O::	sp^2 carboxyl oxygen	2	0	explicit resonating sp^2 oxygen with two lone pairs in carboxyl acid anion or pyridine oxide (C_5H_5NO).
OH	phenol or carboxyl OH	1	1	oxygen atom bearing one acidic hydrogen in phenols or carboxyl −COOH.
N2=	sp^2 amine NH_2 cation	0	2	cationic sp^2 N bonded to two hydrogens and with one other nonrotatable (double) bond. Always flat. No lone pair. NOT guanidinium NH_2.
N1	neutral flat NH, i.e., amide	0	1	planar nitrogen bonded to one H and two other atoms. No lone pair. e.g., amide, guanidinium, methylaniline, and some aromatic heterocycles.
C3	methyl CH_3 group	0	0	sp^3 aliphatic carbon atom bonded to three hydrogen atoms.
DRY	hydrophobic probe	2	2	This type of hydrogen bond is exclusively used in the GRID program for the hydrophobic probe, which identifies places where hydrophobic atoms on the surface of a target molecule will make favorable interactions with hydrophobic atoms on another molecule. It is a distinguishing characteristic of hydrophobic interactions that they only occur when both molecules are immersed in water. This hydrophobic probe may be regarded as a modified water probe. Like water, it must be able to donate and accept hydrogen bonds and must be electrically neutral.

marizes the chemical properties of some of the main probes used.

When the above-mentioned points were carefully followed during application of the FIGO procedure to each testing data set, the superpositions shown in Figure 3 and Table 4 were obtained.

In all cases, the calculated alignment (part a) is consistent with that observed in the crystal (part b), the

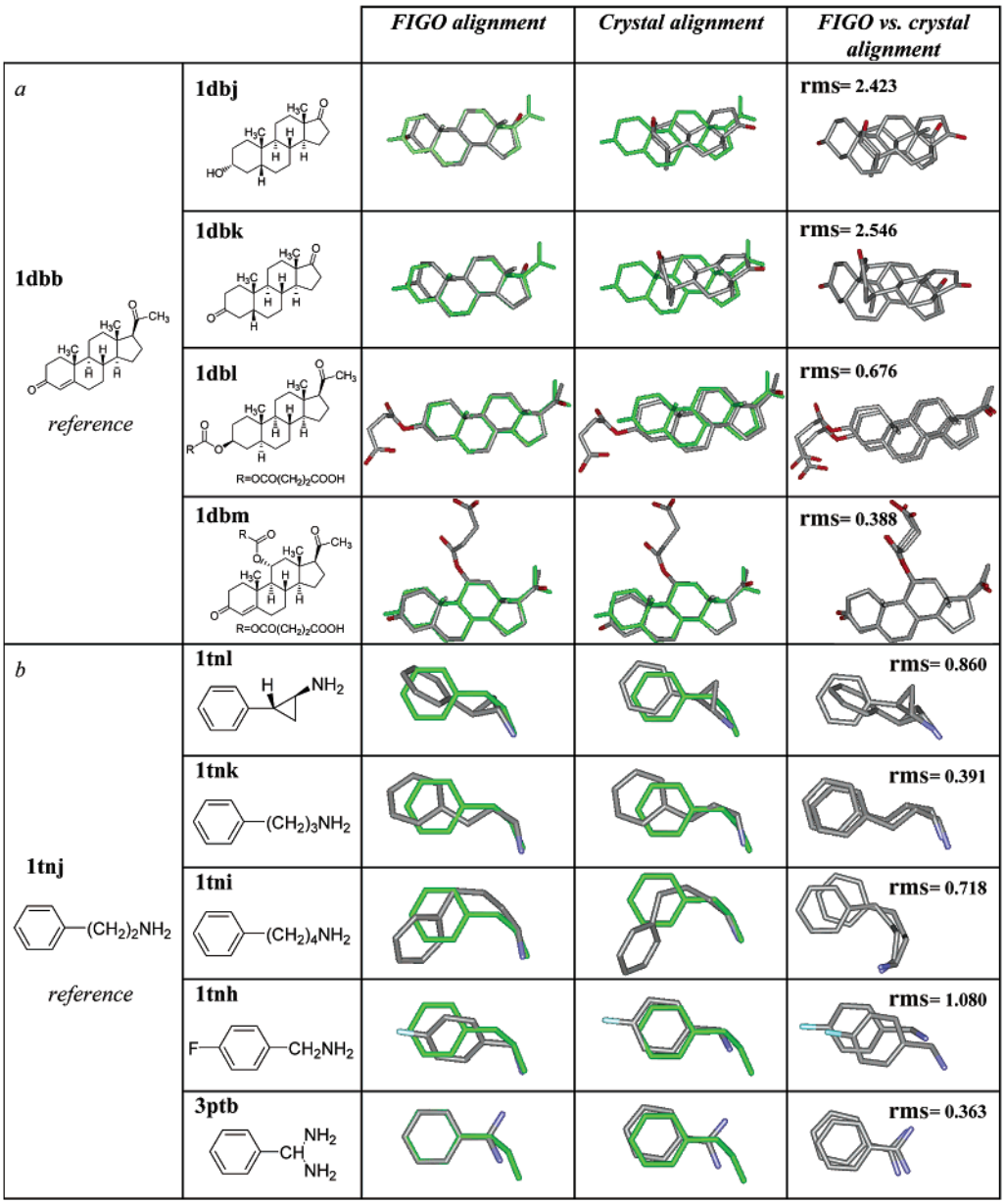


Figure 4. Pairwise representation between the reference (green) and the superposing molecule from the overlay of (a) five steroidic compounds (PDB codes 1dbb, 1dbj, 1dbk, 1dbl, 1dbm); (b) six molecules belonging to the trypsin data set (PDB codes 1tnj, 1tnl, 1tnk, 1tni, 1tnh, 3ptb).

major features of the crystal overlay being preserved. Figure 4 shows greater details with regard to the superpositions of each molecule in the data sets and the appropriate reference structure when the steroid and trypsin data sets were taken as examples. Although the steroid data set gave a less convergent alignment (Figures 3 and 4a), it should be pointed out that the average rms deviation (1.508) for this data set falls within the X-ray crystallography resolution. It can be presumed that while FIGO superposition involves both MIF and heavy atoms of the data set molecules, thus explaining the almost perfect overlay of the gonane system, the same moiety could freely fit into a large pocket of the enzyme with no residues in proximity. Therefore, a rigorous superposition of the gonane system is not strictly required for the activity, whereas superpositions of the molecules belonging to the trypsin data set predicted by FIGO (Figures 3 and 4b) compare well with the crystal setup as confirmed by both the average rms value (0.682) and the rms deviations for each data set molecule shown in Figure 4b.

FIGO is an ideal application in cases where the crystal structure of protein−ligand complexes is not available. Even in this case, verification of the test results was carried out by comparing the FIGO-aligned ligand to the crystal-aligned ligand. For this purpose, conformers derived from the conformational analysis carried out on the ligands belonging to the testing data sets were considered. Subsets of conformers were derived by the application of an appropriate clustering algorithm, and then the FIGO procedure was applied to cluster representatives. Figure 5 refers to the results obtained from the flexibility treatment of the dihydrofolate reductase and α-thrombin data sets taken as examples. The former data set is formed by four molecules. Taking compound 1dyi (PDB code) in its crystallographic conformation as the reference (Figure 5A), seven cluster representatives for each of the other

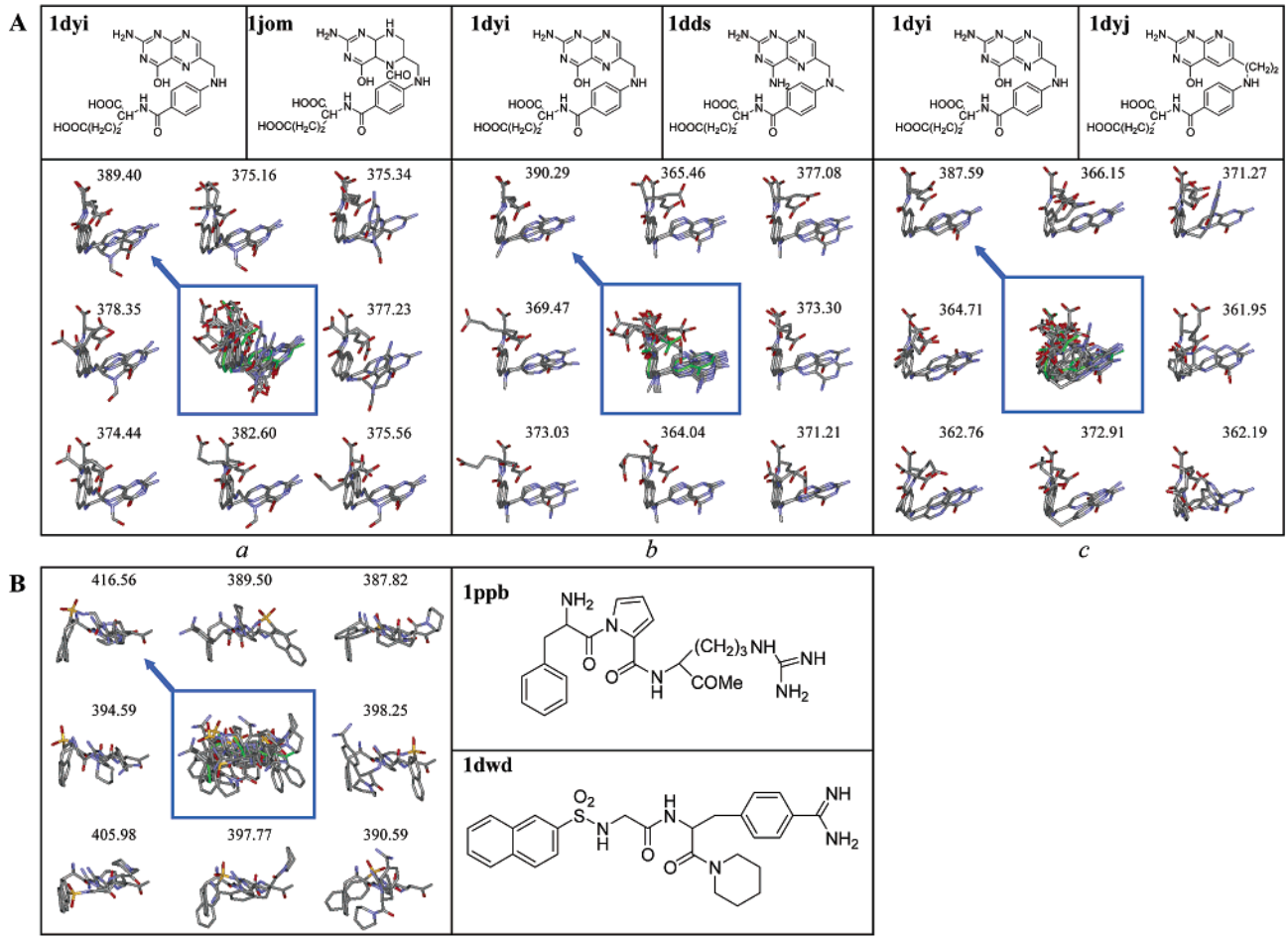


Figure 5. FIGO conformational flexibility treatment. (A) Dihydrofolate reductase data set. Conformers of molecules (a) 1jom (PDB code), (b) 1dds (PDB code), and (c) 1dyj (PDB code) are superposed to the reference structure 1dyi (PDB code, in green). Arrows indicate the superposition with the highest AI value. (B) α-Thrombin data set. Conformers of molecule 1dwd (PDB code) are superposed to the reference structure 1ppb (PDB code, in green). The arrow indicates the superposition with the highest AI value.

three compounds (PDB codes 1jom, 1dds, 1dyj) were taken for the superposition, together with their crystallographic conformations. The number of cluster representatives here is limited because of graphical needs; application of the procedure to a higher number of conformers allows the same kind of conclusions. Overlays of the reference 1dyi and all eight conformers of 1jom, 1dds, and 1dyj are shown in the central portions of parts a, b, and c of Figure 5A, respectively. The alignments between 1dyi and each conformer, together with the corresponding AI values, are illustrated around the central figures. The highest AI values are always associated with the crystallographic conformations of 1jom, 1dds, and 1dyi, which indicates the reliability of FIGO in choosing the appropriate binding conformer orientation. The same conclusions are reached when the components of the other data sets in Table 2 are submitted to this flexibility treatment (data shown only for α-thrombin data set, Figure 5B). In the case of cluster representatives, the AI index has been used as a superposition evaluation parameter because comparison involves different conformers of the same molecule having the same number and type of atoms.

The results presented in this study confirm that FIGO is a semiflexible superposing procedure because it explicitly considers conformers derived from a conformational analysis performed on all studied molecules.

Conclusion

The results collected so far suggest that FIGO is an innovative and promising methodology capable of replacing subjective evaluations for superposition criteria. The ability of the joint MIFs/simplex/experimental design approach has been proved by experimental evidence to be qualitatively very close to reality, the average rms distances being inferior to X-ray crystallography resolution.

Studies are in progress to evaluate the quality performance offered by FIGO when either a single molecule or a hyperstructure is used as the reference. This would give insights into the real effect of the reference structure selection step on the quality of the results.

At present, experience gained with different data sets shows that the best choice is represented by the most biologically active and less flexible compound.

Finally, we are currently considering the possibility of new applications for FIGO methodology in other related 3D-QSAR contexts. For example, FIGO could be successfully joined to a procedure for the extraction of the bioactive conformation of each ligand of the data set. Studies are in progress in order to apply the genetic algorithm procedure to the problem of conformer selection. The reason for this additional feature is straightforward. The selection of molecules in their biologically

active conformation will furnish a detailed picture of the unknown receptor cavity taken from the ligand point of view. It is quite possible that application of such a methodology could help resolve scientific problems that would otherwise be left to subjective solutions.

Materials and Software

The FIGO method was implemented using Visual Basic programming language version 6.0 (Microsoft Corporation). The calculations were performed on a 900 MHz K7Athlon personal computer. CPU calculation time required for the alignment varies according to the number of superposing compounds and the number of atoms in all the molecules. Usually, superposition of 10 compounds requires 20 min to achieve an optimal alignment.

Data sets used for testing the FIGO procedure were retrieved from the Protein Data Bank[40,41] and are set out in Table 2. An additional data set of 57 glycogenophosphorylase inhibitors based on the glucose scaffold was kindly supplied by the Laboratory for Chemometrics, Department of Chemistry, Perugia. X-ray crystallographic results of all ligands complexed to glycogenophosphorilase were available.[44,45]

Hydrogen atoms of data set compounds were added, and charges were loaded from AMBER[46] by the Building module of the InsightII 2000[47] program. These structures were refined by keeping their heavy atoms at fixed positions and then minimizing their energy.

MIFs value were computed with the GRID program.[6,7]

Acknowledgment. This work received financial support from the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR). The authors thank Dr. Mauro Adamo (Oxford University) for his support and useful suggestions.

References

(1) Benedetti, P.; Mannhold, R.; Cruciani, G.; Pastor, M. GBR Compounds and Mepyramines as Cocaine Abuse Therapeutics: Chemometric Studies on Selectivity Using Grid Independent Descriptors (GRIND). *J. Med. Chem.* **2002**, *45*, 1577−1584.

(2) Gratteri, P.; Romanelli, M. N.; Bonaccini, C. Grid independent descriptors (GRIND) in the study of σ-receptor subtype selectivity. *Drugs Future* **2002**, *27* (Suppl. A), 257.

(3) Fontaine, F.; Pastor, M.; Sanz, F. Incorporating Shape into the Grid-INdependent Descriptors: Applications in 3D-QSAR and Bioactive Conformation Searching. *Drugs Future* **2002**, *27* (Suppl. A), 232.

(4) Gnerre, C.; Thull, U.; Gaillard, P.; Carrupt, P. A.; Testa, B.; Fernandes, E.; Silva, F.; Pinto, M.; Wolfender, J. L.; Hostettmann, K.; Cruciani, G. Natural and Synthetic Xanthones as Monoamine Oxidase Inhibitors: Biological Assay and 3D-QSAR. *Helv. Chim. Acta* **2001**, *84*, 552−570.

(5) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. Grid Independent Descriptors (GRIND). A Novel Class of Alignment-Independent Three-Dimensional Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3242.

(6) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(7) *GRID*, version 19; Molecular Discovery Ltd. (20 A Bearkeley Street): Mayfair, London, England, 2001.

(8) Cocchi, M.; De Benedetti, P. G. Use of the Supermolecule Approach To Derive Molecular Similarity Descriptors for QSAR Analysis. *J. Mol. Model.* **1998**, *4*, 113−131.

(9) Cossé-Barbi, A.; Raji, M. Discrete pattern recognition by fitting onto a continuous function. *J. Comput. Chem.* **1997**, *18*, 1875−1892.

(10) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. Sampling conformational hyperspace: Techniques for improving completeness. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 491−499.

(11) De Rosa, M. C.; Berglund, A. A New Method for Predicting the Alignment of Flexible Molecules and Orienting Them in a Receptor Cleft of Known Structure. *J. Med. Chem.* **1998**, *41*, 691−698.

(12) Good, A. C.; Hodgkin, E. E.; Richards, W. G. The Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188−191.

(13) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comput. Chem.* **1996**, *17*, 1653−1666.

(14) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220−232.

(15) Itai, A.; Tomioka, N.; Yamada, M.; Inoue, A.; Kato, Y. Molecular Superposition for Rational Drug Design. In *3D QSAR in Drug Design. Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 200−225.

(16) Jain, A. N.; Dietterich, T. G.; Laterop, R. H.; Chapman, D.; Critchlow, R. E.; Bauer, B. E.; Webster, T. A.; Lonzano-Perez, T. J. Compass: A shape-based machine learning tool for drug design. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 635−652.

(17) Jones, G.; Willet, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532−549.

(18) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.

(19) Klebe, G.; Mietzner, T.; Weber, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35−49.

(20) Lemmen, C.; Hiller, C.; Lengauer, T. RigFit: A new approach to superimposing ligand molecules. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 491−502.

(21) Lemmen, C.; Lengauer, T. Time-efficient flexible superposition of medium-sized molecules *J. Comput.-Aided Mol. Des.* **1997**, *11*, 357−368.

(22) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its applications to dopaminergic and benzodiazepine agonist. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83−102.

(23) Masek, B. B.; Merchant, A.; Matthew, J. R. Molecular Shape Comparison of Angiotensin II Receptor Antagonists. *J. Med. Chem.* **1993**, *36*, 1230−1238.

(24) McMahon, A. J.; King, P. M. Optimization of Carbó molecular similarity index using gradient methods. *J. Comput. Chem.* **1997**, *18*, 151−158.

(25) McMartin, C.; Bohacek, R. S. Flexible matching of test ligands to a 3D pharmacophore using a molecular superposition force field: Comparison of predicted and experimental conformations of inhibitors of three enzymes. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 237−250.

(26) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. MIMIC: A Molecular-Field Matching Program. Exploiting Applicability of Molecular Similarity Approaches. *J. Comput. Chem.* **1997**, *18*, 934−954.

(27) Miller, M. D.; Sheridan, R. P.; Kearsley, S. K. SQ: A Program for Rapidly Producing Pharmacophorically Relevent Molecular Superpositions. *J. Med. Chem.* **1999**, *42*, 1505−1514.

(28) Nissink, J. W. M.; Verdonk, M. L.; Kroon, J.; Nietsner, T.; Klebe, G. Superposition of molecules: Electron density fitting by application of Fourier transforms. *J. Comput. Chem.* **1997**, *18*, 638−645.

(29) Parretti, M. F.; Kroemer, R. T.; Rothman, J. H.; Richards, W. G. Alignment of molecules by the Monte Carlo optimization of molecular similarity indices. *J. Comput. Chem.* **1997**, *18*, 1344−1353.

(30) Perkins, T. D. J.; Mills, J. E. J.; Dean, P. N. Molecular surface-volume and property matching to superpose flexible dissimilar molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 479−490.

(31) Petitjean, M. Geometric Molecular Similarity from Volume-Based Distance Minimization: Application to Saxitonin and Tetrodotoxin. *J. Comput. Chem.* **1995**, *16*, 80−90.

(32) de Caceres, M.; Villà, J.; Lozano, J. J.; Sanz, F. MIPSIM: similarity analysis of molecular interaction potentials. *Bioinformatics* **2000**, *16*, 568−569.

(33) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. The Ensemble Approach to Distance Geometry: Application to the Nicotinic Pharmacophore. *J. Med. Chem.* **1986**, *29*, 899−906.

(34) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232

(35) Mills, J. E.; de Esch, I. J.; Perkins, T. D.; Dean, P. M. SLATE: a method for the superposition of flexible ligands. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 81−96.

(36) Goldmann, B. B.; Wipke, W. T. Quadratic shape descriptors. 1. Rapid superposition of dissimilar molecules using geometrically invariant surface descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 644−658.

(37) Pitman, M. C.; Huber, W. K.; Horn, H.; Krämer, A.; Rice, J. E.; Swope, W. C. FLASHFLOOD: A 3D field-based similarity search and alignment method for flexible molecules. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 587−612.

(38) Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schmidt, J. M. Flexible alignment of small molecules. *J. Med. Chem.* **2001**, *44*, 1483−1490.
(39) Nelder, J. A.; Mead R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308.
(40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(41) Web site: http://www.biochem.ucl.ac.uk/bsm/pdbsum/.
(42) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experiments*; Wiley: New York, 1978.
(43) Mitchell, T. J. An Algorithm for the Construction of "D-Optimal" Experimental Designs. *Technometrics* **1974**, *16*, 203−210.
(44) Martin, J. L.; Veluraja, K.; Ross, K.; Johnson, L. N.; Fleet, G. W. J.; Ramsden, N. G.; Bruce, I.; Ochard, M. G.; Oikonomakos, N. G.; Papageorgiou, A. C.; Leonidas D. D.; Tsitoura, H. S. Glucose Analogue Inhibitors of Glycogen Phosphorilase: The Design of Potential Drugs for Diabetes. *Biochemistry* **1991**, *30*, 10101−10116.
(45) Watson, K. A.; Mitchell, E.; Johnson, L. N.; Son, J. C.; Bichard, C. J. F.; Ochard, M. G.; Fleet, G. W. J.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M.; Papageorgiou, A. C. The Design of Inhibitors of Glycogen Phosphorilase: A Study of α and *β* C-Glucosides and 1-Thio-*β*-D-Glucose Compounds. *Biochemistry* **1994**, *33*, 5745−5758.
(46) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.
(47) *InsightII 2000*; Accelrys Inc. (www.accelrys.com).

JM0210616