Journal of Computer-Aided Molecular Design, 8 (1994) 751–778
ESCOM

J-CAMD 265

Different approaches toward an automatic structural alignment of drug molecules: Applications to sterol mimics, thrombin and thermolysin inhibitors

Gerhard Klebe*, Thomas Mietzner and Frank Weber

BASF AG, Main Laboratory, Carl-Bosch Strasse, D-67056 Ludwigshafen, Germany

Received 17 March 1994
Accepted 17 May 1994

Key words: Structural alignment; Molecular flexibility; Molecular similarity; Sterol mimics; Thrombin and thermolysin inhibitors

SUMMARY

A relative comparison of the binding properties of different drug molecules requires their mutual superposition with respect to various alignment criteria. In order to validate the results of different alignment methods, the crystallographically observed binding geometries of ligands in the pocket of a common protein receptor have been used. The alignment function in the program SEAL that calculates the mutual superposition of molecules has been optimized with respect to these references. Across the reference data set, alignments could be produced that show mean rms deviations of approximately 1 Å compared to the experimental situation. For structures with obvious skeletal similarities a multiple-flexible fit, linking common pharmacophoric groups by virtual springs, has been incorporated into the molecular mechanics program MOMO. In order to combine conformational searching with comparative alignments, the optimized SEAL approach has been applied to sets of conformers generated by MIMUMBA, a program for conformational analysis. Multiple-flexible fits have been calculated for inhibitors of ergosterol biosynthesis. Sets of different thrombin and thermolysin inhibitors have been conformationally analyzed and subsequently aligned by a combined MIMUMBA/SEAL approach. Since for these examples crystallographic data on their mutual alignment are available, an objective assessment of the computed results could be performed. Among the generated conformers, one geometry could be selected for the thrombin and thermolysin inhibitors that approached reasonably well the experimentally observed alignment.

INTRODUCTION

In drug design it is attempted to correlate the three-dimensional (3D) structure of drug molecules with their biological activity. These molecules evolve their activity through specific binding to a macromolecular receptor. All too frequently we are faced with the situation that no information about the 3D structure of this receptor is available. Any features of the ligand–receptor

*To whom correspondence should be addressed.

interactions that explain the differences in binding affinity of the ligands cannot be discussed straightforwardly. As a consequence, in these cases only relative differences between different ligands themselves can be related to variations in their biological data. A large variety of descriptors can be used to express these differences. Some of them describe global features of the molecules (e.g., total lipophilicity), others depend directly on their 3D structure (e.g., capabilities to form hydrogen bonds). In a relative comparison the latter descriptors require a mutual alignment or superposition of the drug molecules in order to determine their gradual changes within a series.

Several approaches to reveal mutual structural alignments for molecular comparison have been reported (for a recent review, see Ref. 1). The most popular way of aligning molecules follows an iterative adjustment of torsion angles with a subsequent least-squares superposition of key atoms of a pharmacophore (derived for the particular class of compounds under consideration). More complete and objective is the active-analog approach [2,3] which generates plausible molecular geometries showing a common orientation of the pharmacophoric groups. If some hypothesis about the active-site conformation is given, a multiple-flexible fit can be performed. Across a series of drug molecules, the pharmacophoric groups are linked by virtual springs, pulling these groups together. In an optimization, the forces generated by these springs are simultaneously minimized, together with the intramolecular force field of each molecule. Through this procedure, the different molecules are forced to adopt conformations with similar orientation of the pharmacophoric groups.

Other methods of aligning molecules also use information about a predefined pharmacophore. Since hydrogen bonds are thought of as forming specific attachment points, anchoring a ligand to its receptor, hydrogen donor or acceptor sites are often used as prime criteria to reveal a common arrangement in space [4–6].

All these alignment methods concentrate on features that are more or less directly associated and extrapolated from the spatial positions of the individual atoms. However, molecules recognize each other by their surfaces and field properties. Thus, alignment methods that map and compare common shape and field properties appear to be better suited to reveal relevant alignments [1]. Several techniques exploiting field properties to superimpose molecules have been described in the literature [7–15]. However, most of them were reported to be rather slow and computationally intensive. Only rarely, some estimations are given to assess the quality of the obtained alignments [8,16]. This requires some objective reference, preferentially derived from experimental sources.

In the present paper, we report on different approaches to align drug molecules. The first approach is based on modifications of the program SEAL, developed by Kearsley and Smith [16]. In this method, molecules with given geometry are superimposed with respect to their steric and electrostatic properties. Their conformations are kept rigid during the fit procedure. In order to assess the quality of the results obtained by the program, a comparison with structural alignments actually observed by protein crystallography has been performed. With respect to this criterion, adjustable parameters in the program have been optimized to improve the applied alignment function. To consider conformational flexibility during the superposition process, two alternative approaches are described (see below). In a series of molecules, a common orientation with respect to their field and recognition features is searched. The intramolecular force field of each molecule is minimized simultaneously with some restraining conditions, reinforcing the above-mentioned alignment criteria.

Three examples are given below to demonstrate the scope and limitations of the alignment procedures. In the first example, both multiple-fit techniques are applied to a set of transition-state mimics involved in ergosterol biosynthesis. Since a fairly rigid reference structure is known, it can be assumed that the obtained alignments are relevant for the actual geometry at the binding site of the involved enzyme. Both approaches are directly applicable and give comparable results, mainly because the reference and the mimics show evidential similarities in their bonding skeletons.

In a second example, five inhibitors of the enzyme thrombin are aligned. In this case crystallographic data are available to assess the relevance of the obtained results. Since for this example no obvious skeletal resemblance on an atom-by-atom basis between a peptidic reference structure and some non-peptidic sulfonamides is given, only the modified SEAL approach, combined with a conformational analysis through MIMUMBA [17,18], could be successfully applied.

In a third case study, five thermolysin ligands of different size have been investigated by the combined MIMUMBA/SEAL approach. As in the previous case, experimentally determined reference data are available to assess the relevance of the obtained results. In all but one case, alignments were computed that approximate the experimentally observed situation.

Alignment of rigid molecules by maximizing the similarity of their physicochemical properties in space

Kearsley and Smith [16] describe an alignment function for the superposition of two rigid molecules that comprises a double sum over all possible atom pairs between both molecules. Accordingly, they calculate a similarity score A_F for a particular alignment as follows:

$$A_F = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\, e^{-\alpha r_{ij}^2}$$

The first summation i runs over all m atoms of the first molecule; the second considers all n atoms of the second structure. The term w_{ij} itself comprises a summation over several physicochemical properties (w_k, weighting factor):

$$w_{ij} = \sum_{k=1}^{l} w_k\, prop_{ik}\, prop_{jk}$$

In their study, Kearsley and Smith use partial charges and van der Waals radii as properties. They indicate that an atom-based hydrophobic measure should be considered as an additional property and suggest to superimpose relatively neutral parts of the molecules using a relationship involving the squares of the partial charges in a reciprocal relationship [16]. We tried to extend this summation by additional properties to improve the alignment function. The results are discussed below.

In the exponential, r_{ij} is the mutual distance between atoms i and j. An adjustable parameter α determines the attenuation range of this distance dependence. It dictates to what extent the shape and properties of the molecules are considered in the comparison: with small values of α also remote parts of each molecule will influence the alignment. The hypersurface (on which the subsequent minimization is performed) will be rather shallow and smooth, with few minima [16].

Large values of α will result in a strong attenuation of the distance-dependent consideration of molecular similarity. Accordingly, there is less averaging of local feature matches of the molecules being compared. The global molecular similarity becomes less important. Alternatively to the Gaussian-type attenuation in SEAL, a Lorentzian-type functional form can be used. The latter is approximate, but faster to compute.

The program reveals several solutions that are ranked according to the achieved similarity scoring of the alignment function. Kearsley and Smith [16] report convincing CPU times. Since SEAL does not require any predefined atom-by-atom correspondences, it appears to be a promising approach to resolve the alignment problem for a fairly broad range of applications. We adapted the program to a Silicon Graphics workstation. The input/output routines were interfaced to SYBYL [19].

Initial test runs showed that the obtained results strongly depend on the selection of the different adjustable parameters α and w_k. Kearsley and Smith also indicate in their paper that the parameter selection is crucial for the produced alignments and the relative ranking of the different solutions. To evade this unsatisfactory situation, we embarked into systematic parameter studies and modifications of the alignment function. The goal of this study was to examine whether a parameter setting with general validity and an improved ranking of the obtained solutions could be found. Such a task can only be solved if a set of target alignments is given that can be used to assess the quality and reliability of the obtained results. This alignment should approach, as closely as possible, the actually observed binding geometry of various ligands in the binding pocket of a common protein receptor [1]. Binding geometries are experimentally available from X-ray crystallography [20]. Despite the limitation to geometries determined in the crystalline phase and their reduced accuracy due to the limited diffracting power of protein crystals, they are the only reliable source for binding geometry data of drug molecules.

As reference data to optimize and assess the results obtained with different alignment functions in SEAL, a data set of pairs of structurally diverse ligands, binding to the same protein, has been compiled. The considered ligands, binding to thrombin, dihydrofolate reductase, thermolysin, trypsin, endothiapepsin, carbopeptidase and HIV protease, are shown in Fig. 1. The deviations obtained by SEAL with respect to the experimentally observed alignments were measured by the rms deviation between all corresponding atoms, including the hydrogen atoms. During the alignment with SEAL, the ligands were kept in their crystallographically determined conformations. The obtained rms deviations have to be discussed in light of the experimental accuracy of the reference data. The reduced resolution of protein structure determinations is an inherent limitation. The error in spatial position has been estimated to about 0.4 Å [21,22]. Further uncertainty about the relative mutual alignment of the ligands is introduced by the procedure used to superimpose two ligand–protein complexes. Usually, the C^{α} positions are matched by a least-squares fit (as done in this study). However, this process does not appropriately consider any conformational changes of the protein backbone resulting from a ligand-induced fit. If the different protein structure determinations are performed on different crystal forms (ligand–protein complexes not obtained from soaking but from cocrystallization can show this kind of 'polymorphism' [23,24]), some conformational changes of the protein can result from the different packing conditions. Taking all these limitations into account, the accuracy of a mutual alignment known from experiment clearly exceeds the above-mentioned positional error.

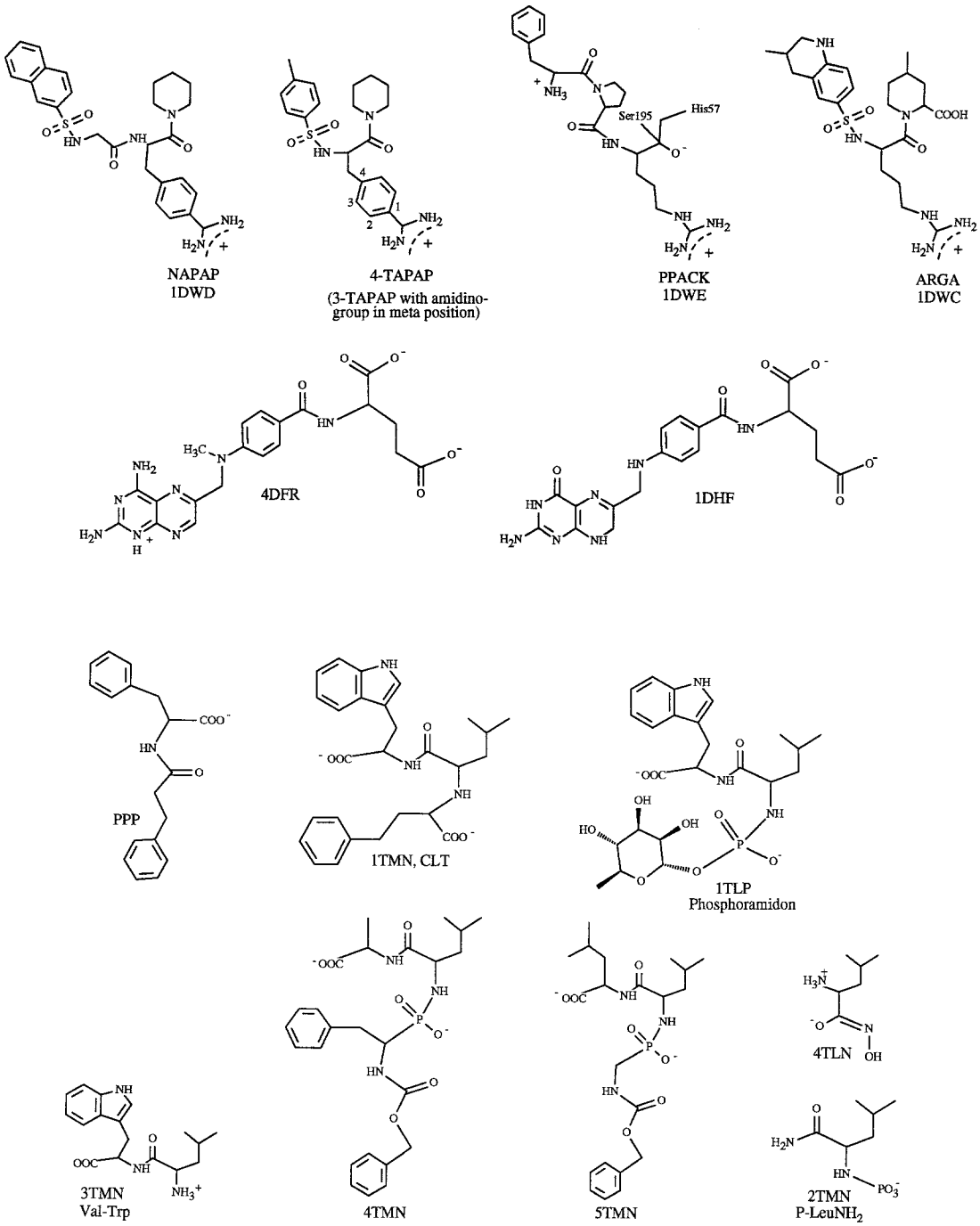

Fig. 1. Molecular formulae of the protein ligands used to optimize the parameterization in SEAL and considered in the case studies of thrombin and thermolysin inhibitors. Each structure is classified by its four-letter PDB reference code name or by an acronym used in the text for the corresponding ligand–protein complex stored in the Brookhaven File [20]. For the thrombin complex with 4-TAPAP [40] and the thermolysin complex with PPP [48], structural data have been used as described in the literature.

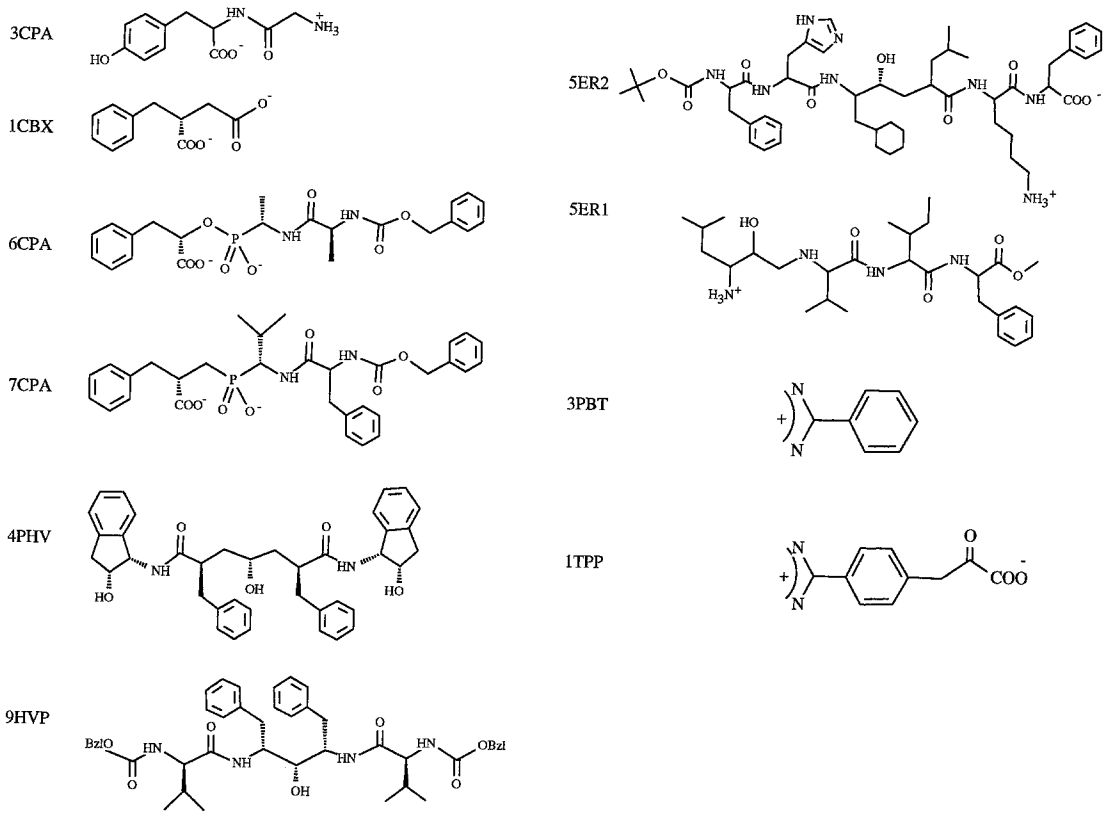

Fig. 1. (continued).

Physicochemical properties in the alignment function

Altogether, we considered four different physicochemical properties in the alignment function. Steric and electrostatic properties were used as the third power of the van der Waals radii or of the partial atomic point charges. Primarily, the latter property has been calculated by the semi-empirical AM1 method [25]. To examine the influence of the applied charge model, alternatively partial charges determined by the force-field program MOMO [26,27] have been used, especially since the AM1 method overemphasizes partial charges of third-row elements, such as phosphorus or sulfur. In addition to the properties originally considered in SEAL, atomic hydrophobicities and refractivities have been taken into account. These two properties describe the lipophilicity and polarizability of molecules.

Atomic hydrophobicities and refractivities have been assigned to the atoms of the molecules according to parameter sets given by Viswanadhan et al. [28]. Some additional hydrophobicities which were missing in the original parameter list have been included using values calculated by the ClogP program [29]. In a subroutine, a topological analysis of the molecules is performed and the parameters are assigned accordingly. Viswanadhan et al. [28] have shown that local correspondences in both parameters are useful criteria to assess the goodness of fit of a geometrical superposition of molecules.

Parameter selection in the alignment function

Since systematic parameter studies are computationally quite intensive, we selected a subset of eight ligand pairs (Table 1) from the reference list in Fig. 1 for the various calibrations and modifications of the alignment function in SEAL. To perform these studies to some extent independent from the actually applied charge model, three pairs of inhibitors of thrombin were considered either with AM1 (A) or MOMO (M) charges, resulting in a total set of 11 pairs (Table 1). In a final run, the parameter setting for the best suited alignment function, emerged from this study, was further optimized with respect to an extended set of 24 ligand pairs (Table 2) from Fig. 1.

TABLE 1
RESULTS OBTAINED FOR THE OPTIMIZATION OF ADJUSTABLE PARAMETERS IN SEAL FOR 11 PAIRS OF PROTEIN LIGANDS, PAIRWISE BINDING TO THE SAME PROTEIN

	Properties: s,e		Properties: s,e,h		Properties: s,e,r			
Ligand pair	rms[a]	Ranking	rms[b]	Ranking	rms[c]	Ranking	rms[d]	$\alpha/w_s/w_e/w_h$
1DWD (A) /	1.665	−1247.8	1.679	−1314.9	1.652	−1443.4	1.475	0.1/40/1/20
1DWE (A)	7.942	−1132.8	3.603	−961.2	7.922	−1308.7		
1DWD (M) /	1.527	−1083.8	1.512	−1150.2	1.521	−1282.4	0.994	0.1/12/8/20
1DWE (M)	5.600	−1000.7	5.617	−1035.6	5.545	−1179.6		
1DWD (A) /	1.508	−1150.2	1.488	−1199.3	1.490	−1326.4	0.958	0.1/40/1/20
4-TAPAP (A)	8.397	−1016.9	8.310	−1030.6	8.408	−1196.8		
1DWD (M) /	1.203	−985.6	1.175	−1038.6	1.161	−1165.0	0.858	0.1/16/1/20
4-TAPAP (M)	7.781	−886.9	7.781	−914.0	7.787	−1061.3		
1DWD (A) /	0.606	−875.4	0.706	−933.4	0.636	−1045.4	0.451	0.1/20/10/20
1DWC (A)	8.858	−823.1	8.814	−840.0	8.864	−985.7		
1DWD (M) /	0.759	−948.7	0.793	−1008.1	0.747	−1117.9	0.700	0.1/40/4/20
1DWC (M)	6.377	−863.4	6.582	−878.5	6.285	−1012.5		
4DFR (A) /	1.210	−890.7	1.110	−911.4	1.035	−1033.0	0.963	0.1/36/2/20
1DHF (A)	4.513	−783.3	5.913	−785.6	5.948	−901.6		
1TMN (A) /	1.342	−762.8	1.262	−821.9	1.418	−908.1	0.868	0.3/40/8/20
5TMN (A)	10.333	−719.2	10.375	−767.8	10.436	−862.8		
1TLP (A) /	0.855	−1007.9	0.758	−1051.2	0.853	−1146.2	0.671	0.2/2/1/20
5TMN (A)	8.242	−850.6	8.339	−851.7	10.237	−912.7		
1TMN (A) /	1.139	−1104.9	1.149	−1158.1	1.130	−1282.4	0.726	0.2/12/8/20
1TLP (A)	8.533	−973.1	8.448	−1018.9	8.043	−1140.3		
5ER2 (A) /	0.947	−1341.1	0.849	−1427.9	0.984	−1587.4	0.699	0.2/2/1/20
5ER1 (A)	10.169	−1326.5	10.346	−1375.7	10.181	−1577.3		

The parameters are α, w_s (steric weight), w_e (electrostatic weight) and w_h (hydrophobic weight). For the thrombin examples (1DWC, 1DWD, 1DWE, 4TAPAP), charges from two different methods (A = AM1, M = MOMO) have been used. In the first row for each example the obtained rms deviation (Å) between calculated and experimentally found alignment is given; the second row refers to the second best solution; the ranking is in arbitrary units. The first set of columns refers to an alignment function solely based on steric and electrostatic properties, the second and third sets include hydrophobicity or refractivity (r) as additional property, and in the last set of columns the rms deviation and the parameter selection (α, w_s, w_e, w_h) of the best possible solution obtained in the studied range are listed.

[a] Parameters: $\alpha/w_s/w_e = 0.2/16/2$.

[b] Parameters: $\alpha/w_s/w_e/w_h = 0.2/16/2/20$.

[c] Parameters: $\alpha/w_s/w_e/w_r = 0.2/16/2/0.05$.

[d] Best possible solution.

The following variations of the alignment function were performed:

(1) determination of an optimal attenuation factor α for a summation over steric (s) and electrostatic (e) terms, together with an adjustment of their relative weights (w_s, w_e);
(2) inclusion of hydrophobicity and refractivity as third summation term;
(3) contribution and discrimination of this third summation term (hydrophobicity or refractivity) compared to the original two-term summation (steric and electrostatic);
(4) introduction of individual distance-dependent attenuation factors α_k for the summation over different physicochemical properties;
(5) influence of the analytical form of the attenuation function (Lorentzian or Gaussian).

In all studies, the 'best' possible alignment for a particular parameter setting within the test data set was determined by:

$$rms_{mean} = \sqrt{\sum_{i}^{N} rms_i^2 / N}$$

where N = the number of considered ligand pairs.

At first, the attenuation factor α and the relative weights w_s and w_e in the two-term alignment function considering steric and electrostatic properties were optimized. The attenuation factor α was systematically varied in steps of 0.1 between 0.1 and 0.5, w_s and w_e were optimized in the range between 1 and 40. As best solution ($\alpha = 0.2$, $w_s = 16$, $w_e = 2$), an rms_{mean} deviation of 1.21 Å between experimentally observed and calculated alignment was achieved. The individual results obtained for each example in the test set are given in the first column of Table 1. In all cases, the best solution suggested by SEAL approximates the experimentally observed alignment. The second best solution (Table 1) deviates substantially from this reference in all cases.

In the next step, a third summation term was introduced into the alignment function. Including hydrophobicities revealed an optimized rms_{mean} of 1.18 Å between calculated and observed alignment ($\alpha = 0.2$, $w_s = 16$, $w_e = 2$, $w_h = 20$). Replacing the hydrophobicity summation by refractivities produced a nearly identical agreement (rms_{mean} = 1.19 Å). The optimized relative weight w_r (range: 0.025–0.2) for the refractivity term indicated only a minor contribution to the alignment function. Interestingly, in these optimizations α, w_s and w_e obtained the same values as in the parameter variation excluding a third summation term. The present results show that considering the hydrophobicities and refractivities as additional terms does not significantly improve the obtained mutual alignment of the ligands (Table 1).

In this context, it is interesting to examine whether these additional terms allow for a better discrimination of the various solutions suggested by SEAL. A comparison between the ranking of the best and second best solution clearly shows that including the hydrophobicities in the alignment function improved the relative ranking of the best solution (Table 1). In all test cases, the best solution approached the experimental reference (best solution: rms_{mean} = 1.18 Å; second best solution: rms_{mean} = 7.90 Å). Overall, the discrimination of the best solution compared to the second best one has been improved by a factor δ (see below) of 1.61 compared to the case where no hydrophobicities were considered. A similar analysis including refractivities showed no improvement with respect to a discrimination of the solutions (δ = 0.98). Accordingly, only a summation over hydrophobicities has been further considered in the similarity score A_F, whereas refractivities were no longer studied.

The value δ is given by:

$$\delta = 1/11 \sum_{i}^{11} \frac{\left[(A_F^I - A_F^{II})/A_F^I\right]_{s,e,h}}{\left[(A_F^I - A_F^{II})/A_F^I\right]_{s,e}}$$

where I is the best solution, II is the second best solution, and s, e and h have been defined above.

The different physicochemical properties considered in the alignment function do not necessarily require the same distance-dependent attenuation. Thus, for each individual summation term k, we introduced a separate parameter α_k. Using a three-term summation in A_F, individual attenuation factors α_k were attributed to steric, electrostatic and hydrophobic properties:

$$A_F = -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{l} w_k \, prop_{ik} \, prop_{jk} \, e^{-\alpha_k r_{ij}^2}$$

The systematic variation of three individual attenuation factors α_k (between 0.1 and 0.5) revealed a combination of $\alpha_s = 0.1$, $\alpha_e = 0.2$, $\alpha_h = 0.2$ as best solution, closely followed by a setting of all values of α to 0.2. During this optimization, the relative weights w_k were kept constant at the values indicated above. Choices $\alpha_s = \alpha_e = \alpha_h = 0.1$ or 0.3 resulted in rms_{mean} values that were larger by 0.23 or 0.03 Å, respectively. Since only a minor improvement was obtained using individual attenuation factors α_k, at the expense of a substantial increase in computing time, an overall α has been used in the following.

Finally, for the 11 reference ligand pairs the results obtained with a Lorentzian- and Gaussian-type attenuation function were compared. With the latter form a substantially better agreement with experiment is obtained ($rms_{mean} = 1.01$, with $\alpha = 0.15$, $w_s = 13$, $w_e = 1$, $w_h = 5$). Thus, in the following only Gaussian-type functions have been applied. On average, the required time to compute a mutual alignment of two ligands (e.g. for those shown in Fig. 1) required about 1 min on a Silicon Graphics Indigo RS4000 using Gaussian attenuation functions starting from 20 random orientations.

TABLE 2
RESULTS OBTAINED FOR THE ALIGNMENT OF 24 PROTEIN LIGANDS, PAIRWISE BINDING TO THE SAME PROTEIN

Ligand pair	rms[a]	Ligand pair	rms[a]	Ligand pair	rms[a]
1DWD (A) / 1DWE (A)	1.639	1TMN (A) / 1TLP (A)	1.163	6CPA (A) / 1CBX (A)	0.635
1DWD (M) / 1DWE (M)	1.478	1TMN (A) / 4TLN (A)	0.720	6CPA (A) / 3CPA (A)	1.627
1DWD (A) / 4-TAPAP (A)	1.276	1TMN (A) / 4TMN (A)	0.810	6CPA (A) / 7CPA (A)	0.380
1DWD (M) / 4-TAPAP (M)	0.870	1TLP (A) / 5TMN (A)	0.816	7CPA (A) / 1CBX (A)	0.760
1DWD (A) / 1DWC (A)	0.662	1TLP (A) / 4TMN (A)	1.000	7CPA (A) / 3CPA (A)	1.761
1DWD (M) / 1DWC (M)	0.854	1TLP (A) / 4TLN (A)	7.170[b]	1CBX (A) / 3CPA (A)	1.017
4DFR (A) / 1DHF (A)	1.029	4TMN (A) / 5TMN (A)	1.808	5ER2 (A) / 5ER1 (A)	0.704
1TMN (A) / 5TMN (A)	1.538	4PHV (A) / 9HVP (A)	0.383	3PTB (A) / 1TPP (A)	0.768

The following parameterization was used for SEAL: $\alpha = 0.2$, $w_s = 12$, $w_e = 1$ and $w_h = 10$. The rms deviations listed are those between calculated and experimentally determined alignments.

[a] Best solution using 0.2/12/1/10, $rms_{mean} = 1.112$ Å.

[b] Not considered in rms_{mean}; as best solution from search in an extended range a value of rms = 1.039 Å is obtained with 0.30/12/1/10.

Validation of the parameter selection in SEAL

Considering steric, electrostatic and hydrophobic terms yielded the best agreement between the best solution suggested by SEAL and the experimental references (Table 1, third column). However, the parameter setting corresponding to the lowest rms_{mean} has to be regarded as 'best' compromise for the present reference data set. Selecting for each of the 11 test examples the best possible solution reveals a significantly lower rms_{mean} of 0.89 Å (Table 1, last row). In light of this observation, we felt that our selection of the 'best' parameter setting should be based on a more extended set of ligand pairs. Accordingly, we included 13 additional examples from different proteins in our test set (Fig. 1, Table 2).

After optimizing the adjustable parameters α, w_s, w_e and w_h as described above, this set of 24 test cases (Table 2) revealed deviations between SEAL and experimental alignments that amount to an rms_{mean} = 1.11 Å. A parameter setting of $\alpha = 0.2$, $w_s = 12$, $w_e = 1$ and $w_h = 10$, using a Gaussian-type functional form, was obtained as optimal selection. For one test example (1TLP/4TLN), the SEAL alignment that ranked as best solution (out of 20 trials) did not approach the experimental reference. However, with a larger α a reasonable alignment could be obtained as best solution. In this example, a rather large ligand is compared with one of much smaller size.

In summary, it can be concluded that the SEAL approach is fast enough for general application. In principle, it can solve the alignment problem correctly within the accuracy defined by the experimental references, provided that the 'best' possible parameter selection is known. We did not find a general way to determine such 'best' parameters. Since the approach is highly dependent on their selection, this fact has clearly to be regarded as a major drawback of the method. It leaves some unpleasant uncertainty while applying the approach to problems where no experimental references are known. However, the finally obtained parameter selection gives some confidence in the transferability and general applicability of the approach.

FLEXIBLE MULTIPLE-FIT ALIGNMENT WITH SIMULTANEOUS STRUCTURE OPTIMIZATION

So far molecular flexibility has not been considered during the alignment. Furthermore, the alignment has been determined comparing only two molecules at a time. Two different implementations to tackle the flexibility problem together with a multiple-flexible fit have been performed. Both approaches optimize a set of drug molecules by simultaneously minimizing their internal force field, together with some restraints that try to force molecular portions with similar properties into a common orientation.

Minimizing the internal force field together with distances between common fit centers

In the first approach, common fit centers (atoms or centers of groups) are defined across the set of drug molecules that link these centers by virtual springs. In the initial step, the program performs a rigid superposition of the molecules using the predefined fit centers. Such features are present in many commercially available programs. However, the prerequisite that an equivalent number of fit centers has to be defined in all of the molecules being compared is often limiting, especially if molecules of different size are fitted. As a consequence, we have implemented the algorithm in such a way that not all of the fit centers have to be present in all of the molecules. For example, in a subset of the considered molecules a particular fragment can be missing.

Suppose a set of molecules j = 1,...,J is given and for each of them a particular number of fit centers has been defined. Accordingly, we can ascribe to each molecule j a set of indices I_j that characterizes these fit centers. As an illustration, we give the following example: four fit centers are assigned to the first molecule. For the second one only three centers are assigned and for the third one again only three. However, not all of the three centers in molecules 2 and 3 correspond to each other. Accordingly, in this example the assigned indices I_j would be:

molecule j = 1: $I_1 = \{1,2,3,4\}$
molecule j = 2: $I_2 = \{2,3,4\}$
molecule j = 3: $I_3 = \{1,3,4\}$

Let n(j,k) be the atom number corresponding to index k in molecule j. Now we search for rotation matrices O_j and translation vectors s_j that minimize the expression:

$$f(O,s) = \sum_j \sum_{i<j} \sum_{k \in I_i \cap I_j} w_k \cdot \left\| O_i(x_{n(i,k)} - s_i) - O_j(x_{n(j,k)} - s_j) \right\|^2$$

The w_k describe weighting factors.

In our simultaneous fit, the searches for optimal rotation and translation cannot be separated, since the number of fit centers is not identical in all molecules in the set. As a consequence, we calculate iteratively a correction of the rotations and of a corresponding set of translations according to a method described by Gerber and Müller [30].

The following steps are performed. At first, the most spherical molecule is determined [30], having a center of gravity which is defined by its set of fit centers:

$$\tilde{s}_j = \sum_{k \in I_j} w_k x_{n(j,k)} \Big/ \sum_{k \in I_j} w_k$$

By this center of gravity we define:

$$T_j = \sum_{k \in I_j} w_k (x_{n(j,k)} - \tilde{s}_j)(x_{n(j,k)} - \tilde{s}_j)^T$$

Now we determine, as described by Gerber and Müller [30], the molecule j with the largest linear coefficient of the characteristic polynomial of T_j. This is achieved by maximizing the expression:

$$\left| \sum_{m,n=1}^{3} T_{jmm} T_{jnn} - T_{jmn} T_{jmn} \right|$$

This molecule j is now taken as target molecule. It exchanges position with molecule J intermediately. In the following step, all remaining molecules in the data set are fitted onto this molecule J. For this purpose the program determines for m = j,J the vectors:

$$\hat{s}_m = \sum_{k \in I_j \cap I_J} w_k x_{n(m,k)} \Big/ \sum_{k \in I_j \cap I_J} w_k$$

They correspond to the centers of gravity defined by the common subset of fit centers present in the two molecules j and J. In general, $\tilde{s}_j$ and $\hat{s}_j$ are not identical. We then use the Kabsch algorithm [31,32] to fit the spatial positions of the fit centers in molecule j, $x_{n(j,k)} - \hat{s}_j$, $k \in I_j \cap I_J$ onto those in molecule J, $x_{n(J,k)} - \hat{s}_J$, $k \in I_j \cap I_J$ using the rotation matrix O_j^0.

Subsequently, the iterative simultaneous fit starts by calculating the optimal translation s_j^m (m = index of iteration step), which corresponds to the rotation O_j^{m-1} from the previous step. This is achieved by minimizing the expression $f(O^{m-1},s^m)$ with respect to s^m, keeping the target molecule fixed. Since this problem is quadratic, it can be solved exactly. In a next step, again the rotations are corrected. This is performed by a linearization described by Gerber and Müller [30]. O^m is expressed as $\Omega^m O^{m-1}$. Starting with the coordinates produced by the previous step m–1, Ω^m describes the residual rotation performed on the *m*th iteration level. The function $f(\Omega^m O^{m-1},s^m)$ is optimized in Ω^m only. Again the target molecule is kept fixed in space. All missing fit centers are automatically located in the presently defined center of gravity (spanned by the corresponding fit centers) and they are invariant against any rotations. Consequently, the method described by Gerber and Müller [30] can be directly applied. This iterative procedure converges after a few steps, even with larger corrections Ω^m.

After the rigid simultaneous fit, a force-field minimization of all molecules included in the data set is initiated. This structure optimization is based on the molecular mechanics program MOMO [26,27]. It is performed considering simultaneously additional forces acting between corresponding fit centers. These forces are derived from a potential which increases with the sum of the squared distances between the fit centers. Its initial value can be assigned by the user; a default value is set to 120 kJ/mol. The energy contribution that arises from the potential associated with the additional forces should diminish toward the end of the calculation. This is achieved by increasing its slope during the optimization.

This application of additional restraining forces is based on the following assumptions. In the beginning of the optimization, they introduce some perturbation into the system under consideration. They can give rise to conformational transitions. During the subsequent force-field minimization, the applied perturbation is relaxed in a step-wise fashion. The increasing slope of the distance-dependent potential is achieved by successively ascending the power of the sum of the squared distances, starting from one up to a value of four. During the force-field minimization, following a Quasi-Newton method (adapted from MOMO [26]), the additional forces are optimized together with the intramolecular force field until no significant change in the total energy is obtained and the residual gradient passes a predefined lower limit. The implemented algorithm allows one to assign weights to the individual fit centers. Through this step, the various groups in the molecules which are driven into a common spatial orientation can be considered as having deviating importance. Besides a superposition of atoms as fit centers, the program also allows one to define centers of atomic groups as fit centers. To achieve this assignment, the geometric center of a particular group is calculated together with the normal defined by a best plane through the atoms of the group. In the optimization the group centers are superimposed together with the normal vectors.

Minimizing the internal force field considering the alignment function of SEAL

The above-described multiple-flexible fit alignment requires predetermined atom-by-atom correspondences. This assignment can involve some arbitrary assumptions. In order to avoid such

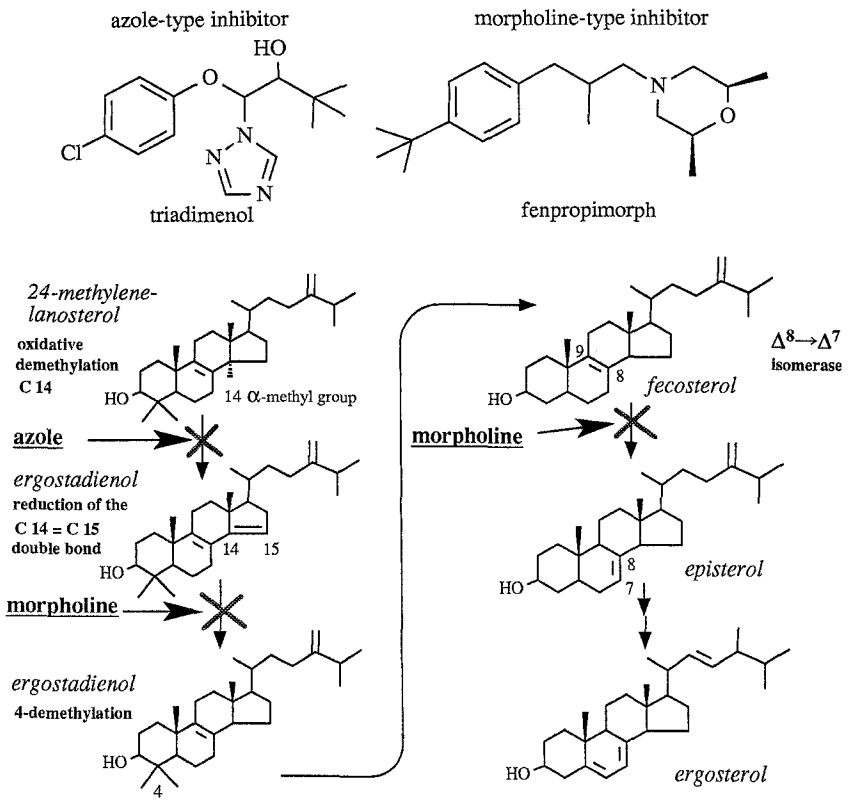

Fig. 2. Biochemical transformation pathway of lanosterol to ergosterol. Triadimenol and fenpropimorph are typical azole- and morpholine-type inhibitors, respectively. The latter compounds are assumed to inhibit Δ^{14}-sterol reductase or $\Delta^{8}\rightarrow\Delta^{7}$-isomerase.

prejudicial simplifications, we performed an alternative implementation of a restraint molecular-fit condition into a force-field optimization. The alignment function applied in SEAL (see above) to achieve a common orientation of molecules has been adapted. To allow for a fast flexible fit, the structure optimization implemented into MIMUMBA [18] has been used. This program performs a geometry optimization in torsion angle space considering van der Waals interactions together with empirically derived 'potentials'. These 'potentials' are based on statistically evaluated torsional fragments retrieved from the Cambridge Structural Database. In the optimization step, the initially generated geometries are minimized to remove unfavorable steric contacts between adjacent atoms in a molecule, simultaneously keeping the torsion angles as close as possible to the most frequently occurring values in experimental structures. It has been shown [18] that local minima produced by MIMUMBA fall close to geometries actually observed at the binding site of proteins. In the program TORSEAL, an alignment of molecules starting from a given orientation (which is normally obtained from a rigid prefit in SEAL) is structurally optimized by minimizing the following potential:

$$\mathrm{Pot} = \mathrm{Pot}_{\mathrm{vdW}} + \mathrm{Pot}_{\mathrm{Tor}} + \sum_{i}\sum_{j} w_{ij} e^{-\alpha r_{ij}^{2}}$$

MIMICS FOR THE TRANSITION STATE IN ERGOSTEROL BIOSYNTHESIS: A CASE STUDY BASED ON MULTIPLE-FLEXIBLE FIT ALIGNMENTS

Inhibitors of ergosterol biosynthesis

The application of a multiple-flexible fit alignment requires at least one fairly rigid molecule in the data set or several molecules of varying skeletons showing rigidity in complementary spatial regions. In the following example a rather rigid structure is considered, i.e., the assumed transition state involved in the enzymatic transformation of ergosterol in biosynthesis. Both alternative multiple-flexible fit alignments have been applied to this example.

Inhibitors of the ergosterol biosynthesis are potent fungicides [33]. Two enzymes, Δ^{14}-sterol reductase and $\Delta^{8}\rightarrow\Delta^{7}$-isomerase, are involved in the ergosterol transformation (Fig. 2). The reduction of the Δ^{14} double bond is assumed to be initiated by a protonation step at C15 (see Fig. 2) [34]. The intermediate allyl carbocation is mesomerically stabilized. Subsequently, a hydride ion is transferred stereospecifically onto C14 from the α-face. The isomerase reaction pathway is assumed to proceed via a very similar intermediate. Initially, the C8-C9 double bond is protonated at C8 from the α-face [35]. Subsequently, a hydrogen is abstracted from C7 to yield the isomerized double bond C7-C8.

Morpholine fungicides (e.g. fenpropimorph, Fig. 2) inhibit both enzymes [33]. Due to their basic character, they are protonated under physiological conditions. Their inhibiting potency is ascribed to their structural and electronic similarity with the intermediate carbocation [36] (Fig. 3). In the reductase, the positive charge should be delocalized between C8, C9 and C14; in the isomerase it should be located at C8. However, various molecular frameworks can be imagined to mimic the geometry of this cationic transition state [33].

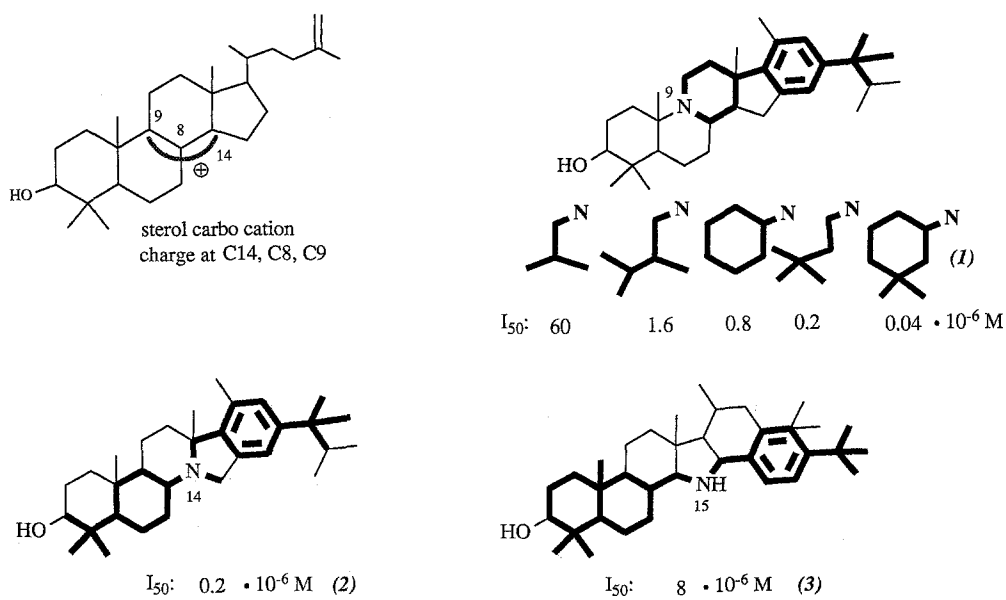

Fig. 3. Chemical formula of the assumed cationic sterol transition state in the enzymatic reduction and isomerization step. Three different classes of mimics, formally orienting a positively charged nitrogen toward the charged portion in the sterol reference cation, are shown with their I_{50} values for the inhibition of Δ^{14}-reductase from *Ustilago maydis*: 4-*t*-butylphenylpiperidines (**1**), *t*-butylbenzyl-(*t*-butyl)cyclohexylamines (**2**), and 4-*t*-butylbenzylmethylaminodecalins (**3**).

Similar volume requirements and a positive charge are exhibited by a protonated nitrogen present in the different types of sterol mimics shown in Fig. 3. The class of 4-*t*-butylphenylpiperidines formally orients a nitrogen toward the 9-position of the sterol skeleton. The potency of these derivatives increases with growing size of the side chain at nitrogen (see Fig. 3; I_{50} values of Δ^{14}-reductase from *Ustilago maydis* [33]). For the structural comparison the 3,3-dimethylcyclohexyl derivative (**1**) has been used. The second class of amines bears a *t*-butylbenzyl and a 4-*t*-butylcyclohexyl substituent at nitrogen (**2**). These groups formally orient a basic nitrogen onto the 14-position of the sterol moiety. The third class of inhibitors formally locates a nitrogen toward the 15-position of the sterol skeleton. In these decalin-methyl-4-*t*-butylbenzyl amines (**3**) this position is adjacent to the atoms bearing the charge in the sterol allyl cation.

Multiple-flexible fit by minimizing distance between common fit centers

Assuming a similar orientation for all three different types of compounds at the binding site, the sterol reference cation and three representative inhibitors were subjected to a multiple fit. According to the approach based on the alignment of common fit centers, various centers were selected as shown in Fig. 4. A simultaneous minimization of the restraining forces and the internal molecular force fields has been performed. The residual energy contribution of the restraining forces amounts to 3.8 kJ/mol toward the end of the calculation (initial value 120 kJ/mol). The obtained fit is shown in Fig. 5. The relative distances between the various fit centers in the different molecules, measured from their common geometric centers, are listed in Table 3.

Multiple-flexible fit based on the SEAL alignment function

For comparison purposes, a multiple fit with TORSEAL, using the alignment function imple-

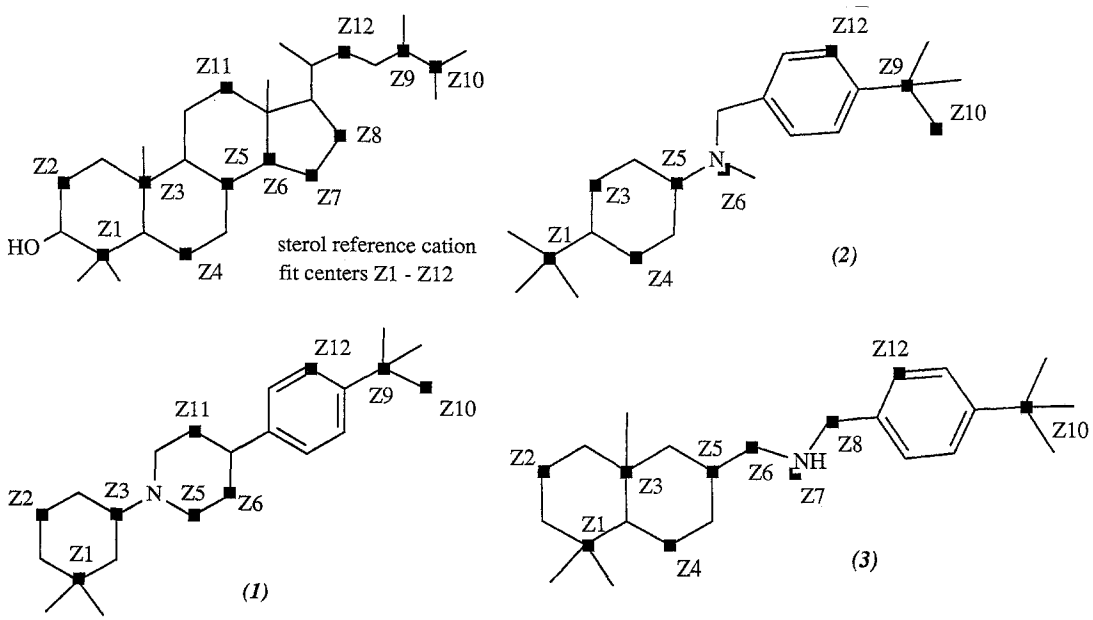

Fig. 4. Definition of fit centers in the sterol reference and the three mimics. These centers were used in a multiple-flexible fit to minimize the internal force field of the molecules, together with the mutual distances between these centers. Corresponding centers have the same numbering.

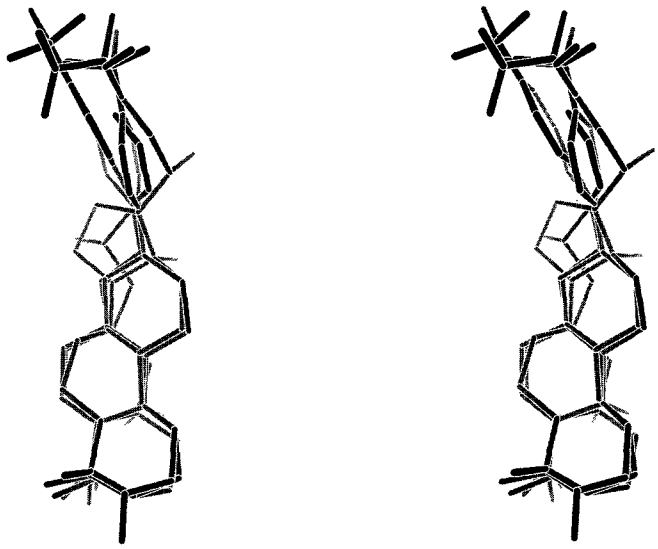

Fig. 5. Stereodiagram of the structural alignment of the sterol reference and the three mimics obtained from the multiple-flexible fit, based on a simultaneous minimization of the internal force field (MOMO) of the molecules together with the mutual distances between corresponding fit centers. Drawings were performed with the program SHADEMOL [52].

mented into SEAL (parameter setting as obtained from the set of 24 ligand pairs), has been performed. The results from the previously described multiple fit were used as starting geometries; charges were calculated according to the AM1 method [25]. The minimization revealed an alignment as shown in Fig. 6. Whereas the method with common fit centers tends to sharply superimpose identical parts of the molecular skeletons, the second approach seems to consider more appropriately the global shape of the molecules. Since no experimental reference data are avail-

TABLE 3
RELATIVE DISTANCES BETWEEN THE VARIOUS FIT CENTERS IN THE STEROL REFERENCE AND THE THREE MIMICS

	Sterol	Mimic **1**	Mimic **2**	Mimic **3**
Z1	0.11	0.27	0.19	0.46
Z2	0.45	0.23	–	0.52
Z3	0.13	0.27	0.13	0.46
Z4	0.52	–	0.10	0.60
Z5	0.15	0.09	0.13	0.18
Z6	0.15	0.37	0.35	0.64
Z7	0.47	–	–	0.47
Z8	0.61	–	–	0.61
Z9	0.29	0.26	0.11	–
Z10	0.34	0.32	0.20	0.53
Z11	0.11	0.11	–	–
Z12	0.75	0.32	0.30	0.71

The relative distances (Å) have been measured from the common geometric centers, obtained in the multiple-flexible fit approach using virtual springs.

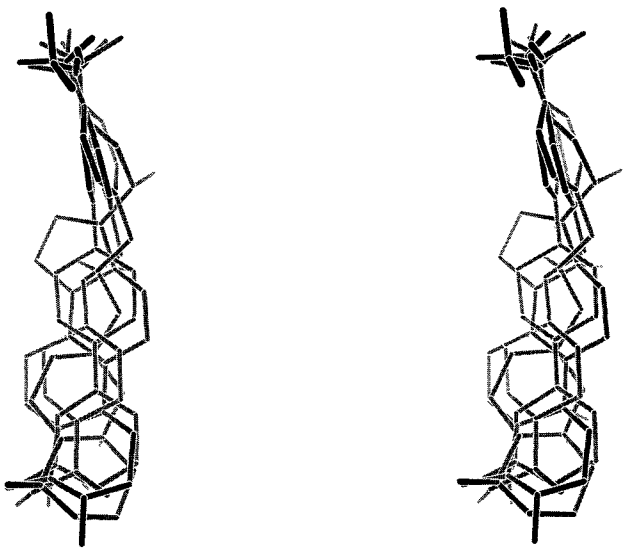

Fig. 6. Stereodiagram of the structural alignment of the sterol reference and the three mimics obtained from the multiple-flexible fit, based on a simultaneous minimization of the torsion angles (MIMUMBA) in the molecules together with the alignment condition used in SEAL.

able to assess the obtained alignments, both solutions appear to be of equivalent relevance. To quantify the differences between both alignments, normalized similarity scores among the four molecules in both mutual superpositions have been calculated directly from the alignment func-

TABLE 4
NORMALIZED SIMILARITY SCORINGS OF TWO ALTERNATIVE ALIGNMENTS OF THE STEROL REFERENCE AND THREE MIMICS

		Mimic **1**		Mimic **2**		Mimic **3**	
		TORSEAL	MOMO	TORSEAL	MOMO	TORSEAL	MOMO
Sterol	Total	0.854	0.830	0.882	0.870	0.890	0.836
	Steric	0.916	0.893	0.930	0.922	0.945	0.896
	Electro	0.522	0.550	0.615	0.618	0.571	0.501
	Hydro	0.729	0.664	0.767	0.722	0.778	0.704
Mimic **1**	Total			0.917	0.869	0.937	0.852
	Steric			0.970	0.938	0.954	0.892
	Electro			0.634	0.577	0.886	0.681
	Hydro			0.870	0.735	0.885	0.779
Mimic **2**	Total					0.902	0.770
	Steric					0.952	0.853
	Electro					0.633	0.366
	Hydro					0.824	0.613

The similarity scores were obtained by a multiple-flexible fit, minimizing mutual distances between corresponding fit centers (Fig. 5) or optimizing the alignment condition from SEAL (Fig. 6). All mutual scorings between the aligned molecules are given: for each pair, in the first column the alignment is scored according to the results obtained by TORSEAL and in the second column the results according to the multiple-flexible fit method (MOMO) are listed.

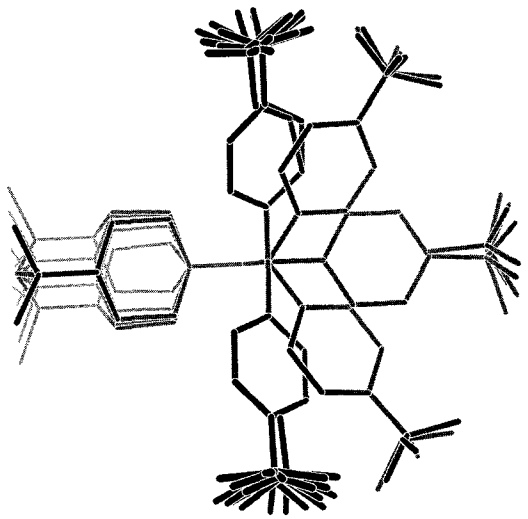

Fig. 7. Spatial distribution of 47 conformers of mimic **1**, generated by MIMUMBA.

tion A_F used in SEAL:

$$A_F^{norm} = \frac{A_F}{\left[\left(A_F^{self}\right)_{mol\,1} \cdot \left(A_F^{self}\right)_{mol\,2}\right]^{1/2}}$$

In Table 4, the obtained scorings are given. A comparison of the two alignments shows that, except for one case, all mutual scorings based on the results obtained with TORSEAL indicate a higher similarity. It has to be admitted that this comparison is clearly biased to favor the SEAL approach.

The two procedures described so far require a reasonable hypothesis about an initial conformation and possible correspondences in terms of equivalent fit centers or a reasonable preorientation of the molecules in the second approach. To combine the search for an appropriate alignment with a more general consideration of conformational flexibility, the following procedure has been applied. Independently, for the three sterol mimics (**1–3**) a conformational analysis has been performed using MIMUMBA. In the energy range selected for the rings (only chair conformations, with the largest substituent equatorial) and open-chain fragments, 47, 8 and 64 conforma-

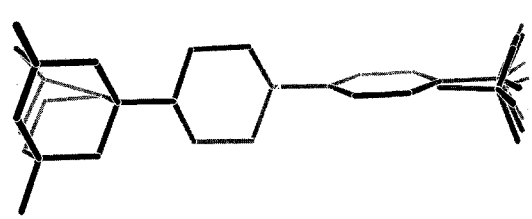

Fig. 8. Spatial distribution of eight conformers of mimic **2**, generated by MIMUMBA.

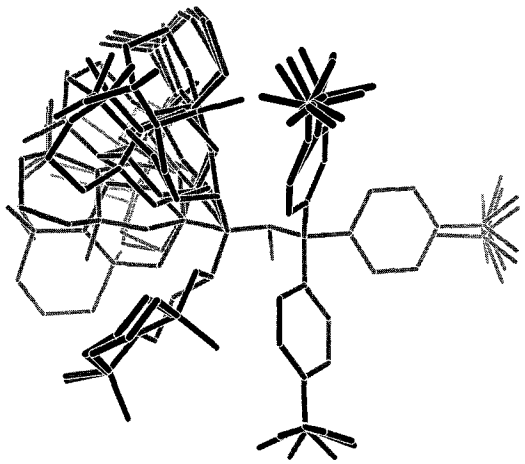

Fig. 9. Spatial distribution of 64 conformers of mimic **3**, generated by MIMUMBA.

tions were generated for **1**, **2** and **3**, respectively. To illustrate the distribution in space, superpositions of these conformers are shown in Figs. 7–9. The sterol cation has been used for the subsequent molecular comparison in a fixed conformation.

For each of the three mimics, a separate comparison with the sterol reference has been performed. Each conformer from the sets generated by MIMUMBA was subjected to SEAL. In a first approximation, for all conformers the same charges have been used. In the present study, the two best alignment solutions suggested by the program were further considered; however, additional alignments can easily be included. In subsequent runs these solutions were submitted

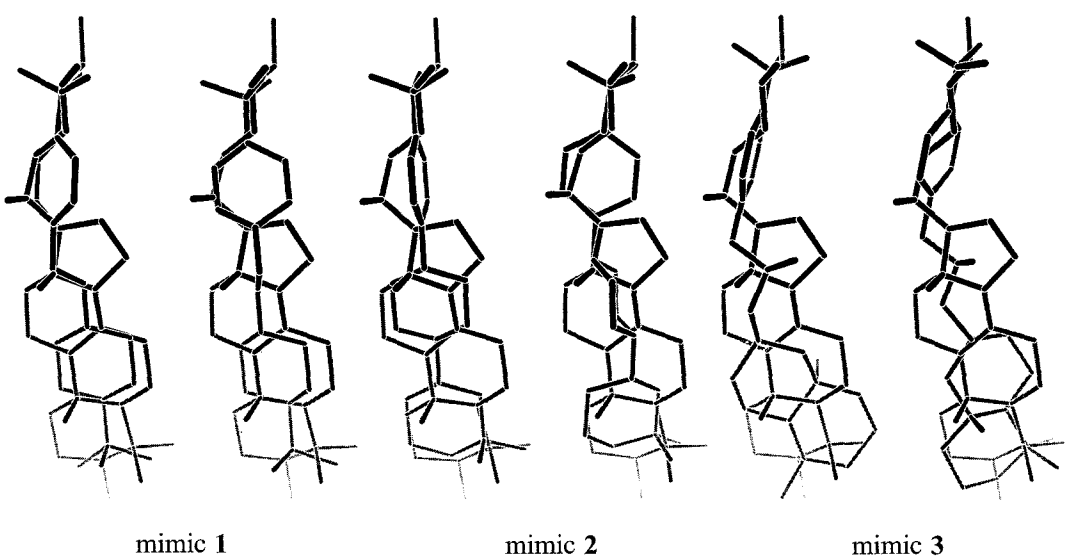

Fig. 10. Structural alignment of the sterol reference with each of the three mimics. The conformers selected by SEAL as best (left) and second best (right) alignment are shown together with the reference.

to TORSEAL. The conformation of the sterol cation was kept rigid, but the torsion angles of the mimics were optimized in order to maximize the structural similarity with the reference. The achieved normalized similarity scores revealed in this step are given in Table 5. For mimics **1** and **2**, the best solution from the rigid SEAL alignment also corresponded to the best solution after relaxing their conformations. In the case of the third mimic, the second best solution suggested by SEAL is ranked higher after conformational relaxation. A more detailed analysis of the SEAL solutions shows that for **1** and **2** the two selected conformers differ in the orientation of the terminal phenyl ring, and in **1** in the central cyclohexane moiety (Fig. 10, left and center). For **3**, the two selected conformers differ by a 180° flip in the orientation of the decalin portion (Fig. 10, right). As a consequence, in the best SEAL solution, the axial methyl group of **3** at the bridge head between both fused rings is oriented in a direction opposite to the 19-methyl group in the sterol cation. In the second SEAL solution both methyl groups are aligned in the same direction. This superposition reveals a higher similarity scoring after conformational relaxation. In a final optimization run with TORSEAL, the reference and the three mimics (for **1** and **2** the conformers from the best solution, for **3** that from the second best solution) were simultaneously relaxed. The resulting alignment (Fig. 6) is identical to that obtained while starting TORSEAL from the multiple-flexible fit alignment using virtual springs between predefined centers (Fig. 5). The results suggest that the combined MIMUMBA/SEAL approach is appropriate to resolve the flexibility and alignment problem fairly automatically, with as few prejudices as possible.

Inhibitors of thrombin

In a second case study, five inhibitors of thrombin have been compared. Thrombin is a trypsin-like serine protease involved in the blood coagulation cascade [37,38]. It plays a key role in thrombosis, hemostasis and wound healing. Accordingly, thrombin inhibition might offer an attractive means of antithrombotic therapy. Apart from large natural inhibitors such as hirudin, the small tripeptide D-Phe-Pro-Arg-chloromethylketone (PPACK, Fig. 1) is known to bind irreversibly to thrombin. It is covalently connected to Ser^{195} and His^{57}. Its guanidinium group forms a chelate-type hydrogen bond with Asp^{198} and Gly^{213}. The structure of this inhibitor bound to thrombin has been determined by protein crystallography [39,40]. The coordinates of PPACK have been extracted from the ligand–protein complex (1DWC) [39] by breaking the bonds between the protein and the inhibitor. In the subsequent structural comparison with non-peptidic and reversibly binding inhibitors, it has been used in this experimental conformation.

NAPAP, argatroban (ARGA) and 3- and 4-TAPAP are non-peptidic inhibitors of thrombin

TABLE 5
NORMALIZED SIMILARITY SCORINGS BETWEEN THE STEROL REFERENCE AND EACH OF THE THREE MIMICS

		Mimic **1**		Mimic **2**		Mimic **3**	
		Conf1	Conf2	Conf1	Conf2	Conf1	Conf2
Sterol	Total	0.855	0.849	0.889	0.872	0.883	0.891
	Steric	0.917	0.918	0.937	0.924	0.943	0.948
	Electro	0.531	0.491	0.621	0.606	0.523	0.554
	Hydro	0.728	0.704	0.774	0.736	0.766	0.777

The scorings were based on the SEAL alignment obtained for conformers that give the best and second best ranking in SEAL.

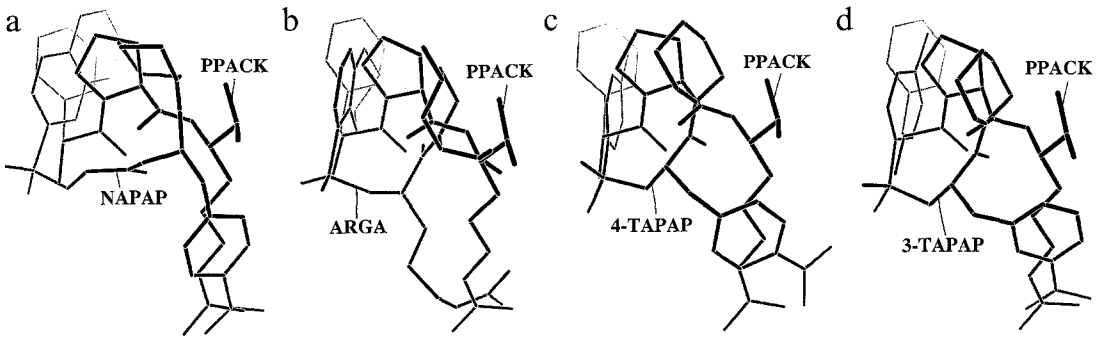

Fig. 11. Mutual structural alignment of the peptidic, covalently binding thrombin inhibitor D-Phe-Pro-Arg (PPACK) with the inhibitors NAPAP (a), argatroban (ARGA, b), 4-TAPAP (c), and 3-TAPAP (d), as determined by protein crystallography (for structural formulae see Fig. 1). As evident from the diagram, no clear 'atom-by-atom' correspondences are present between the different inhibitors and the reference compound.

(Fig. 1). For all sulfonamides the binding geometry is known from protein crystallography (1DWD and 1DWE [39]; 1PPH and 1PPC [40,41]). In the case of 3-TAPAP (1PPH), only the binding geometry in the active site of trypsin is known. However, the structure of the binding site in the latter serine protease is closely related to that of thrombin. In order to obtain a reasonable alignment among all inhibitors, the protein structures of human thrombin (1DWC, 1DWD, 1DWE [39]) and the complex of 4-TAPAP with bovine thrombin [41] were aligned (C^{α}-atoms). The trypsin structure containing 3-TAPAP (1PPH [41]) was accordingly superimposed with the trypsin complex of NAPAP (1PPC [41]). Subsequently, the two trypsin complexes were aligned to the thrombin complexes using the coordinates of NAPAP (1DWD, 1PPC) in both complexes. Since the naphthyl side chain binds to both enzymes in conformations differing by a 180° flip, this portion was excluded from the fit. Apart from this moiety, the geometry of NAPAP is nearly identical in both enzymes. It is obvious that the mutual alignment of PPACK, NAPAP and ARGA (all from human thrombin studied by the same group) will be defined with higher reliability compared to that of 4-TAPAP (thrombin from different species, different protein research groups) and especially that of 3-TAPAP (from a different, but related enzyme). In Fig. 11, the mutual alignments of the peptidic reference structure PPACK with each of the four non-peptidic sulfonamides are shown as evident from protein crystallography.

From these diagrams it is obvious that a direct atom-by-atom superposition of the different ligands is hardly given. All five inhibitors are fairly flexible and – without additional constraints – the active-site geometry cannot be determined in the absence of the protein structure. Accordingly, as already mentioned, the peptidic reference was assumed rigid. For the four non-peptidic ligands, separate conformational analyses were performed using MIMUMBA [17]. The program generated, in a predefined energy window, 94, 148, 146 and 144 conformers for NAPAP, ARGA, 3-TAPAP and 4-TAPAP, respectively. This required 2.57, 2.25, 1.85 and 1.72 s, respectively, per conformer on a Silicon Graphics Indigo RS4000. For all four sulfonamides these sets of conformers have been submitted to a pairwise alignment with the 'rigid' peptidic reference compound D-Phe-Pro-Arg, using SEAL (which starts from random orientations). Atomic charges were determined with the AM1 method [25] and were assumed identical for all conformers. As in the previous example, the best solutions from SEAL were further optimized in TORSEAL,

TABLE 6
RMS DEVIATIONS BETWEEN THE EXPERIMENTALLY DETERMINED CONFORMATION AND ORIENTATION OF FOUR THROMBIN INHIBITORS

Inhibitor	rms (TORSEAL) including H	rms (TORSEAL) without H	rms (SEAL) including H
NAPAP	0.808	0.795	1.162
ARGA	1.210	1.007	1.212
3-TAPAP	1.542	1.348	1.690
4-TAPAP	1.163	1.135	1.225

The conformers were generated by MIMUMBA and selected and aligned by SEAL. The rms values (Å) are given considering all atoms or the non-hydrogen atoms only. For comparison, the deviations before and after the conformational relaxation in the final step are given.

relaxing the torsion angles of the sulfonamides. The finally obtained conformations and relative alignments of the four sulfonamides with respect to PPACK have been compared with their experimentally determined binding geometries. The achieved rms deviations, including all atoms or omitting the hydrogen atoms, are given in Table 6. In order to quantify the changes in the final conformational relaxation, the rms deviations before this relaxation are also given.

In all cases, the best solution obtained agrees quite satisfactorily with the experimentally observed situation (Table 6). The smallest deviations are obtained for those complexes (PPACK/ARGA, PPACK/NAPAP, PPACK/4-TAPAP) where the alignment is more reliably defined. They fall into the same range as found for the alignment of ligands with given rigid conformation (Table 1). This underlines the relevance of the conformations obtained by MIMUMBA, especially since only minor conformational changes are observed in the final torsional relaxation, which only slightly improved the rms deviations (Table 6). For PPACK/3-TAPAP a higher rms deviation is obtained. Supposedly, this is also due to the less well defined experimental reference compound. In Fig. 12, the four sulfonamides are shown in their experimentally determined geometry (XS), together with a calculated conformation as obtained after alignment to the given

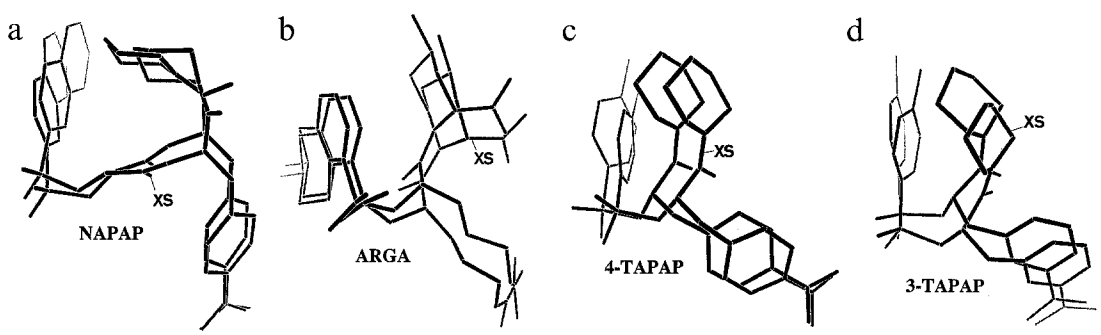

Fig. 12. Structural comparison between the experimentally determined conformation and spatial orientation (XS) of the thrombin inhibitors NAPAP (a), ARGA (b), 4-TAPAP (c), and 3-TAPAP (d) and a conformer generated by MIMUMBA which revealed the best solution in a mutual structural alignment with the receptor-bound conformation of the tripeptide PPACK. This best solution has been relaxed in torsion angle space, simultaneously considering the alignment condition implemented in SEAL. From this diagram the deviations between the alignment known from experiment (Fig. 11) and the computed alignment with respect to PPACK are evident.

orientation of PPACK. The results indicate that the combined MIMUMBA/SEAL approach can successfully resolve the conformational searching and alignment problem. The approach is not limited to structures with similar binding skeletons; peptidic and non-peptidic ligands can be compared successfully.

Inhibitors of thermolysin

Five inhibitors of thermolysin for which the binding-site geometry is known from protein crystallography were selected for a comparative alignment study using the combined MIMUMBA/SEAL approach. Thermolysin is an endopeptidase from a family of extracellular neutral proteases with zinc in the catalytic center [42]. Ligand data were processed as described in the thrombin example. Sets of conformers were generated by MIMUMBA (CLT: 551, Val-Trp: 200, P-Leu-NH_2: 64, PPP: 171; see below). After sorting these conformers according to their energy ranking, the 150 most favorite were submitted to SEAL. The ligand phosphoramidon (Fig. 1), a phospho derivative of the dipeptide Leu-Trp, was chosen as rigid reference compound. The phosphate group is substituted by a rhamno-pyranosyl moiety. Its receptor-bound conformation has been crystallographically determined (1TLP [43]). Due to an intramolecular hydrogen bond between the terminal sugar moiety and the peptide carboxy terminus, the ligand adopts a cyclic conformation. The hydrophobic side chains of the leucine and tryptophan residues orient 'away' from the ring center.

In a first comparison, multiple conformations of the inhibitor *N*-(1-carboxy-3-phenylpropyl)-Leu-Trp (CLT, 1TMN [44], Fig. 1) were aligned with SEAL onto the given conformation of

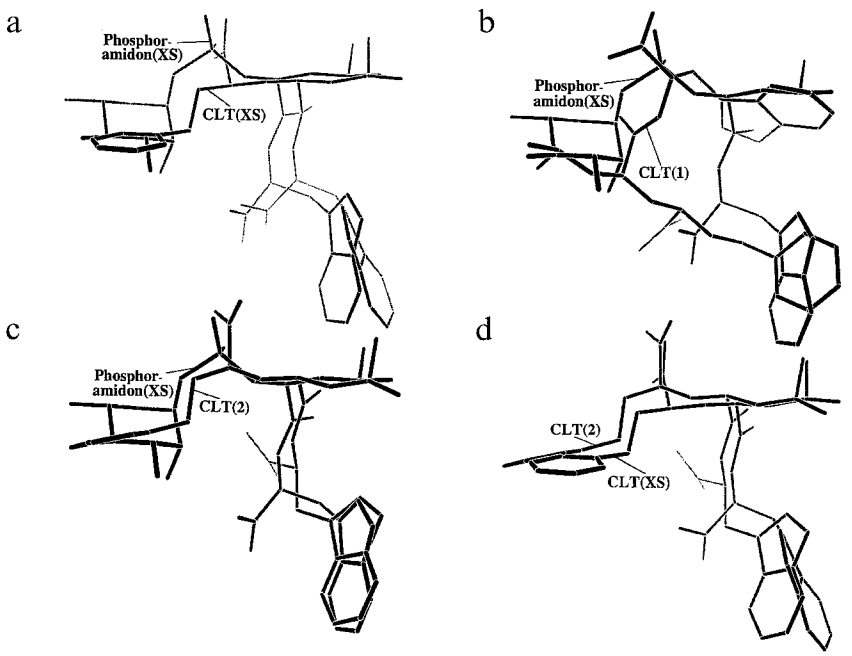

Fig. 13. Different mutual alignments of phosphoramidon with CLT (for structural formulae see Fig. 1): experimentally observed alignment (a), best solution (b, CLT(1)) and second best solution (c, CLT(2)) from the SEAL/MIMUMBA approach. In (d) the orientation and conformation of CLT observed in the crystalline ligand–protein complex (XS) is superimposed with the second best solution (CLT(2)) from the computed alignments.

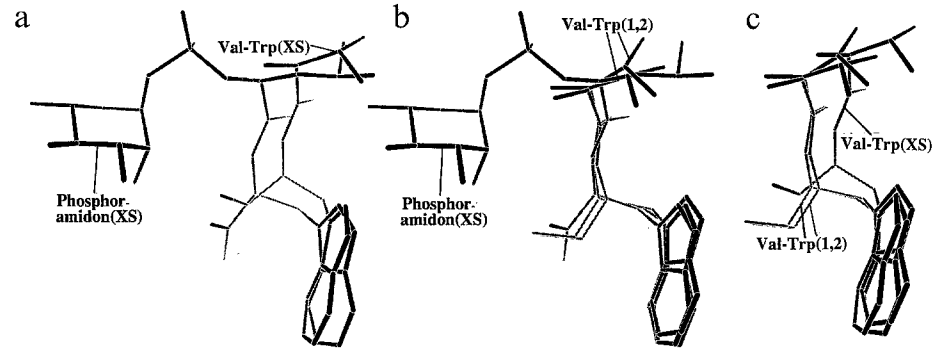

Fig. 14. Different mutual alignments of phosphoramidon with Val-Trp: experimentally observed alignment (a) and best and second best solution (b, Val-Trp(1,2)) from the SEAL/MIMUMBA approach. In (c) the orientation and conformation of Val-Trp observed in the crystalline ligand–protein complex (XS) is superimposed with the two best solutions from the computed alignments.

phosphoramidon. Both ligands are of comparable size. The two best solutions suggested by SEAL were further optimized in torsion angle space. They display completely different solutions (Fig. 13). The alignment ranked best in SEAL (Fig. 13b) superimposes the terminal phenylpropyl moiety of CLT onto the indole group of phosphoramidon and its tryptophan portion is oriented toward the leucine residue in phosphoramidon. The backbone of CLT 'bridges over' the intramolecular hydrogen bond in phosphoramidon. Simultaneously, the two carboxy groups of CLT are brought in the neighborhood of the phosphate and carboxy groups in phosphoramidon. The second best solution suggested by SEAL (Fig. 13c) approaches the experimentally determined alignment (Fig. 13a); however, an rms deviation of 2.23 Å is found (Fig. 13d). The conformation generated by MIMUMBA deviates in the region of the C-terminus of CLT compared to the X-ray structure (the rms deviation in a direct atom-by-atom superposition was 1.22 Å [45]).

The dipeptide ligand Val-Trp is the reaction product of the enzyme reaction (3TMN [46]). This ligand is substantially smaller than phosphoramidon. The four best solutions obtained by SEAL (Figs. 14b,c) all superimpose Val-Trp onto phosphoramidon in a way that approaches the experi-

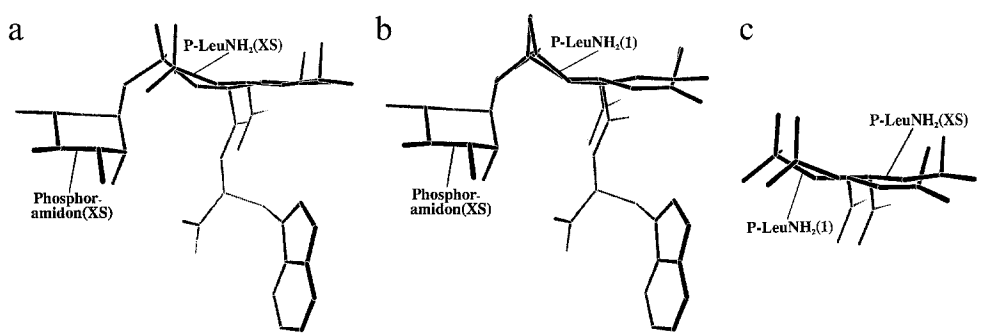

Fig. 15. Different mutual alignments of phosphoramidon with P-Leu-NH_2: experimentally observed alignment (a) and best solution (b, P-Leu-NH_2(1)) from the SEAL/MIMUMBA approach. In (c) the orientation and conformation of P-Leu-NH_2 observed in the crystalline ligand–protein complex (XS) is superimposed with the best solution (P-Leu-NH_2(1)) from the computed alignments.

mentally observed alignment (Figs. 14a,c, rms deviation 1.38 Å). *N*-phosphoryl-leucinamide (P-Leu-NH_2, 2TMN [47]) is correctly aligned with phosphoramidon by the combined MIMUMBA/SEAL approach (Fig. 15). The obtained geometry deviates from the experimentally given alignment by an rms value of 0.94 Å.

As a final example, the ligand β-phenylpropionyl-phenylalanine (PPP) has been studied. The two best solutions suggested by SEAL are shown in Figs. 16b,d. Both differ from the experimentally observed alignment [48] (Fig. 16a). PPP possesses hydrophobic groups at both terminal ends. Polar amide and carboxy groups are located in the central part of the molecule. In the X-ray structure the amide carbonyl oxygen binds to zinc, thus corresponding to the PO_2 portion in phosphoramidon. The phenylpropionyl moiety matches with the rhamnose ring and the benzyl portion of phenylalanine in PPP with the leucine side chain in phosphoramidon. The carboxylate of PPP binds to the protein in a similar way as the peptide group between the leucine and tryptophan residues of phosphoramidon.

In the best solution obtained by SEAL (Fig. 16b), the phenylalanine side chain of PPP orients toward the isopropyl group of phosphoramidon, and its phenylpropionyl portion superimposes with the indole group of phosphoramidon. The carboxylates of both inhibitors are oriented in equivalent spatial regions. This alignment would fail to coordinate the zinc atom in the protein. The second best solution (Fig. 16d) also matches the two phenyl groups of PPP with the two hydrophobic side chains in phosphoramidon; however, the result is a conformation that orients the carboxy group toward the PO_2 unit in the latter inhibitor.

The question can be raised why an alignment approaching the experimentally observed situation is not detected. Among the conformations generated by MIMUMBA, one conformer could

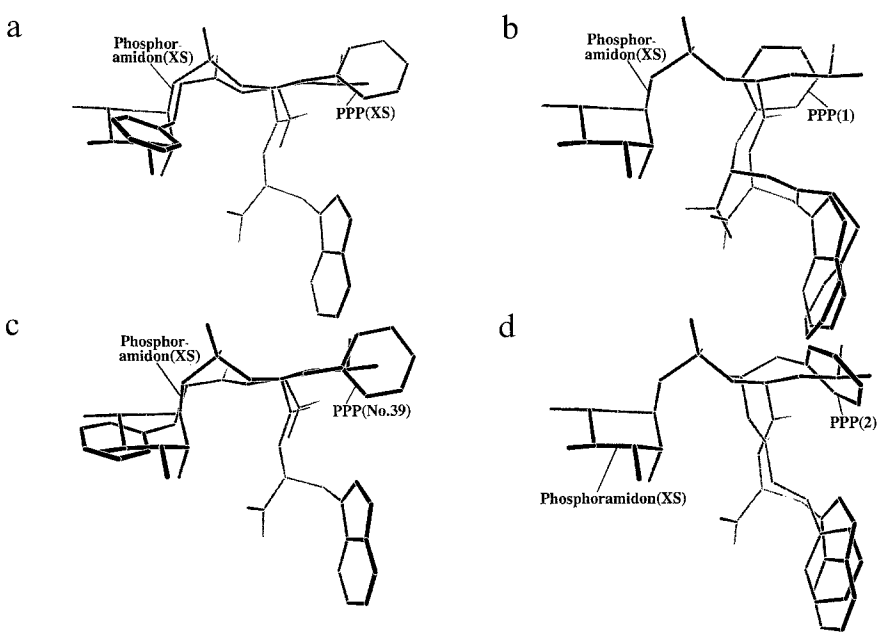

Fig. 16. Different mutual alignments of phosphoramidon with PPP (for structural formulae see Fig. 1): experimentally observed alignment (a), best solution (b, PPP(1)) and second best solution (d, PPP(2)) from the SEAL/MIMUMBA approach. In (c) the conformer generated by MIMUMBA (no. 39) which falls closest (rms deviation: 0.69 Å) to the geometry observed in the crystalline ligand–protein complex has been aligned with SEAL onto phosphoramidon.

be found (no. 39) that deviates only by an rms value of 0.69 Å from the geometry observed in the protein. Aligning this conformer with SEAL onto phosphoramidon reveals as fourth best solution (Fig. 16c) an orientation which approximates the experimental reference (rms is 1.63 Å). This solution is scored by SEAL substantially worse compared to the two previously described alignments.

SUMMARY AND CONCLUSIONS

The alignment function in the program SEAL, developed by Kearsley and Smith [16] has been extended and improved. Apart from steric and electrostatic properties, additional physicochemical parameters have been tested. The relative weight and the distance-dependent attenuation of the different contributions were optimized with respect to test data sets of 24 pairs of low-molecular-weight ligands binding to common proteins. For these ligands the actual orientation and, accordingly, their relative alignment have been determined by protein crystallography. Considering steric, electrostatic and hydrophobic properties allows one to reproduce the experimentally observed alignment with a mean rms deviation of 1.11 Å.

In order to incorporate 'local' conformational flexibility in an alignment procedure, two approaches were applied. Both require some assumptions about possible correspondences (distribution of 'fit centers' or initial preorientation) of the molecules. This could be an obvious guess based upon a putative pharmacophoric pattern or, more objective, the result of a previous rigid SEAL alignment. In the first approach, across the set of molecules common fit centers are distributed and pairwise linked by virtual springs. After an initial rigid reorientation, the distances between these centers are minimized together with the intramolecular force field of the molecules. Alternatively, the alignment function in SEAL has been introduced as additional term into a force field-type optimization in torsion angle space. Whereas the first approach requires some assumptions about possible correspondences (fit centers), the second approach automatically considers these similarities through the properties in the alignment function. Since the latter procedure does not require any fit centers directly associated with the molecular framework, also structures with strongly deviating bonding skeletons can be successfully compared and aligned. Both alignment procedures have been applied to a set of ergosterol biosynthesis inhibitors.

As mentioned above, to perform the alignment both approaches require some initial assumptions about possible conformations and putative correspondences between the ligands. Thus, only a restricted conformation and alignment space is searched. To circumvent these important limitations, the alignment procedure in SEAL has been combined with a conformational search using MIMUMBA. The latter method generates conformers with geometries relevant in a molecular environment. It covers conformation space rather efficiently. Two case studies, on thrombin and thermolysin inhibitors, were performed. For these examples, structural information about the mutual alignment of the ligands is experimentally evident from protein crystallography. The combined MIMUMBA/SEAL approach allowed us to select from sets of up to 150 conformers one structure which showed, in all but one case, after alignment with a common reference structure of given conformation, a geometry and relative orientation approximating the experimentally observed situation.

The latter method is the more objective and global one. On a Silicon Graphics Indigo RS4000, its computational effort requires about 1–3 s per conformer in MIMUMBA and about 30 to 90

min (depending on the size of the molecules and the number of considered conformations) for selecting the best alignment within SEAL. Since these procedures should be ideally suited for parallelization, a tremendous speed-up is to be expected on a multiprocessor device.

The most critical steps in the MIMUMBA/SEAL approach are the completeness in conformational searching with MIMUMBA and the quality and ranking of the best solution in SEAL. Using up to 150 conformers for the comparison might be insufficient for ligands with many rotatable bonds. Difficult to handle are also cases where a ligand binds in a conformation with torsional fragments adopting geometries ranked as unfavorable in the initial screening. On the basis of a reliable scoring, the SEAL approach has to select as best solution one which approximates the experimental situation. To improve this critical ranking of the various alignment solutions, perhaps additional physicochemical properties can help to further discriminate between these solutions.

The superposition of molecules of comparable size with groups at the terminal ends, discriminating with respect to their physicochemical properties (e.g. the thrombin case), will be more easy to handle than examples including molecules of different size (e.g. the thermolysin case). Most complicated are cases where the ligands being compared show a roughly symmetric pattern of physicochemical properties on their surfaces (e.g., extended molecules with hydrophobic groups at the terminal ends and polar hydrogen bond-forming groups in the center, like PPP).

However, in the absence of the protein receptor structure the most critical step in an alignment study is a plausible hypothesis on the receptor-bound conformation of, at least, one ligand. This problem is solved if a rigid reference structure is available (e.g. the sterol example). If a set of fairly rigid ligands with conformational restrictions in complementary parts of the overall bonding skeleton connecting the pharmacophoric groups is given [49], the problem can be approached by conformational analyses with MIMUMBA and subsequent alignments with SEAL.

REFERENCES

1 Klebe, G., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 173–199.
2 Marshall, G., Annu. Rev. Pharmacol. Toxicol., 27 (1987) 193.
3 Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.D. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds.) Computer-Assisted Drug Design, Vol. 112, American Chemical Society, Washington, DC, 1979, pp. 205–222.
4 Kato, Y., Itai, A. and Iitaka, Y., Tetrahedron, 43 (1987) 5229.
5 Kato, Y., Inoue, A., Yamada, M., Tomioka, N. and Itai, A., J. Comput.-Aided Mol. Design, 6 (1992) 475.
6 Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P., J. Comput.-Aided Mol. Design, 7 (1993) 83.
7 Herrmann, R.B. and Herron, D.K., J. Comput.-Aided Mol. Design, 5 (1991) 511.
8 Manaut, M., Sanz, F., Jose, J. and Milesi, M., J. Comput.-Aided Mol. Design, 5 (1991) 371.
9 Clark, M., Cramer III, R.D., Jones, D.M., Patterson, D.E. and Simeroth, P.E., Tetrahedron Comput. Methodol., 3 (1990) 47.
10 Good, A.C., Hodgkin, E.E. and Richards, W.G., J. Chem. Inf. Comput. Sci., 32 (1992) 188.
11 Dean, P.M., In Johnson, M.A. and Maggiora, G.M. (Eds.) Concepts and Applications of Molecular Similarity, Wiley, New York, NY, 1990, pp. 211–238.
12 Namasivayam, S. and Dean, P.M., J. Mol. Graphics, 4 (1986) 46.
13 Chau, P.L. and Dean, P.M., J. Mol. Graphics, 5 (1987) 97.
14 Dean, P.M. and Chau, P.L., J. Mol. Graphics, 5 (1987) 152.

15 Dean, P.M., Callow, P. and Chau, P.L., J. Mol. Graphics, 6 (1988) 28.
16 Kearsley, S.K. and Smith, G.M., Tetrahedron Comput. Methodol., 3 (1990) 615.
17 Klebe, G. and Mietzner, T., J. Comput.-Aided Mol. Design, 8 (1994) 583.
18 Klebe, G. and Mietzner, T., In Jones, D.W. and Katrusiak, A. (Eds.) Correlations, Transformations and Interactions in Organic Crystal Chemistry, IUCr Series, Oxford University Press, Oxford, 1994, pp. 135–158.
19 SYBYL, Version 6.05, Tripos Associates, St. Louis, MO, 1993.
20 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rogers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
21 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T., Proteins, 4 (1988) 31.
22 As a rule of thumb the approximate positional error amounts to about one sixth of the resolution (G. Vriend, personal communication).
23 For HIV protease, structure determinations have been reported which show the protein in five different space groups [24]. Across these different crystal forms the enzyme structure is conserved, with rms differences of approximately 0.7 Å.
24 Appelt, K., Perspect. Drug Discov. Design, 1 (1993) 23.
25 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
26 Beck, H. and Egert, E., Force-Field Program MOMO, University of Göttingen, Göttingen, 1988.
27 Smith, A.E. and Lindner, H.J., J. Comput.-Aided Mol. Design, 5 (1991) 235.
28 Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
29 ClogP program, Release 3.51, Medicinal Chemistry Software, Pomona College, Claremont, CA, 1987.
30 Gerber, P. and Müller, K., Acta Crystallogr., A30 (1987) 426.
31 Kabsch, W., Acta Crystallogr., A32 (1976) 922.
32 Kabsch, W., Acta Crystallogr., A34 (1978) 827.
33 Akers, A., Ammermann, E., Buschmann, E., Götz, N., Himmele, W., Lorenz, G., Pommer, E.H., Röhl, F., Sauter, H., Siegel, H., Zipperer, B. and Zipplies, M., Pestic. Sci., 31 (1991) 521.
34 Akhtar, M., Rahimtula, A.D. and Wilson, D.C., Biochem. J., 117 (1970) 539.
35 Wilson, D.C., Rahimtula, A.D. and Akhtar, M., Biochem. J., 114 (1969) 71.
36 Rahier, A., Schmitt, P., Huss, B., Benveniste, P. and Pommer, E.H., Pestic. Biochem. Physiol., 25 (1986) 112.
37 Fenton II, J.W., Semin. Thromb. Hemost., 14 (1988) 234.
38 Liu, L.W., Vu, T.K.H., Esmon, C.T. and Coughlin, S.R., J. Biol. Chem., 266 (1991) 16977.
39 Banner, D.W. and Hadvary, P., J. Biol. Chem., 266 (1991) 20085.
40 Brandstetter, H., Turk, D., Hoeffken, H.W., Grosse, D., Stürzebecher, J., Martin, P.D., Edwards, B.F.P. and Bode, W., J. Mol. Biol., 226 (1992) 1085.
41 Turk, D., Stürzebecher, J. and Bode, W., FEBS Lett., 287 (1991) 133.
42 Matthews, B.W., Acc. Chem. Res., 21 (1988) 333.
43 Weaver, L.H., Kester, W.R. and Matthews, B.W., J. Mol. Biol., 114 (1977) 119.
44 Monzingo, A.F. and Matthews, B.W., Biochemistry, 23 (1984) 5724.
45 Recently we reported [17] on a conformational analysis of CLT using MIMUMBA, which revealed a geometry in closer agreement with the experimental observation. In this search more torsion angles were restrained and accordingly, in a smaller set of conformers an appropriate conformer was found. In the present study, the geometry reported in Ref. 17 is also found. However, the relative ranking classifies this conformer higher in energy. Accordingly, it is not present in the set of the 150 most favored geometries.
46 Holden, H.M. and Matthews, B.W., J. Biol. Chem., 263 (1988) 3256.
47 Tronrud, D.E., Monzingo, A.F. and Matthews, B.W., Eur. J. Biochem., 157 (1986) 261.
48 Kester, W.R. and Matthews, B.W., Biochemistry, 16 (1977) 2506.
49 An example of such a data set is the ACE case studied by Mayer et al. [50]. However, no experimental reference to assess the relevance of the obtained computed alignment is available. For ACE also deviating alignments have been reported [50,51].
50 Mayer, D., Naylor, C.B., Motoc, I. and Marshall, G.R., J. Comput.-Aided Mol. Design, 1 (1987) 3.
51 Hausin, R.J. and Codding, P.W., J. Med. Chem., 33 (1990) 1940.
52 Hahn, M. and Wipke, W.T., Tetrahedron Comput. Methodol., 1 (1988) 81.