

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

Colin McMartin* and Regine S. Bohacek

Pharmaceuticals Division, CIBA-GEIGY Corporation, Summit, NJ 07901, U.S.A.

Received 18 January 1995

Accepted 26 April 1995

Keywords: Template fitting; Pharmacophore matching; Flexible molecular superposition; Drug design; Superposition force field

Summary

A computer procedure TFIT, which uses a molecular superposition force field to flexibly match test compounds to a 3D pharmacophore, was evaluated to find out whether it could reliably predict the bioactive conformations of flexible ligands. The program superposition force field optimizes the overlap of those atoms of the test ligand and template that are of similar chemical type, by applying an attractive force between atoms of the test ligand and template which are close together and of similar type (hydrogen bonding, charge, hydrophobicity). A procedure involving Monte Carlo torsion perturbations, followed by torsional energy minimization, is used to find conformations of the test ligand which minimize the internal energy of the ligand and the superposition energy of ligand and template. The procedure was tested by applying it to a series of flexible ligands for which the bioactive conformation was known experimentally. The 15 molecules tested were inhibitors of thermolysin, HIV-1 protease or endothiapepsin for which X-ray structures of the bioactive conformation were available. For each enzyme, one of the molecules served as a template and the others, after being conformationally randomized, were fitted. The fitted conformation was then compared to the known binding geometry. The matching procedure was successful in predicting the bioactive conformations of many of the structures tested. Significant deviation from experimental results was found only for parts of molecules where it was readily apparent that the template did not contain sufficient information to accurately determine the bioactive conformation.

Introduction

Medicinal chemists frequently use 3D pharmacophores as targets for structure-based drug design. In order to design novel molecules which accurately mimic the pharmacophore, it is necessary to compare proposed structures to the target. A useful procedure would accept a design molecule and superimpose it on the pharmacophore, so that groups of similar functionality would be aligned. The result could then be visually examined to see which groups were aligned well. It would also be helpful to know the internal energy of the superimposed molecule relative to a global minimum energy. In this way, molecules which match the template without internal strain could be selected. If necessary, the design could then be modified to improve the match or lower the strain energy.

We have developed a computer method that uses a superposition force field, which makes it very simple to carry out such comparisons. This method differs from a number of published template matching methods [1–7] in which the user selects atoms to be superimposed and the program searches for conformations which give the best superimposition of the selected atoms. These methods, based on user-defined atom pairs, have been successfully applied to small molecules. Although such methods are clearly useful, they are not always easily applied to complex molecules. One problem is that the atoms to be matched must be specified in advance by the user. For large molecules, where many arrangements of matching atoms may need to be considered, this can be time-consuming and it is often impossible to specify the matches in advance. Particularly difficult problems can occur when

*To whom correspondence should be addressed.

matching hydrophobic regions of molecules. For example, an aliphatic isopropyl side chain can often match well to a phenyl, but the match cannot be defined in terms of pairs of atoms.

In addition, these fitting procedures do not directly take into account conformational strain energy. Bioactive conformers usually differ from energy minimized conformations of the unbound molecule. Although the geometry may differ appreciably from an energy minimized geometry, the strain energy is unlikely to be large. When a computer search finds a match which involves very high strain energy, this should not be accepted. To solve this problem, it is necessary to have a procedure which allows some strain to occur and evaluates the trade-off between strain energy and the extent to which the structures are well matched.

The use of a general superposition force field solves the problem of specifying the atoms to be superimposed. The superposition potential automatically introduces attractive forces between atoms in the test molecule and atoms of the template which have similar properties and are close to each other. As the test molecule is moved to minimize the energy, the attractive forces automatically change to reflect the current overlap of the test ligand and the pharmacophore. Smith [8] has successfully applied a method based on this principle to the problem of superimposing two molecules in fixed conformations. In this rigid match application, one of the molecules is rotated and translated to minimize the superposition energy. In order to extend this method, to allow flexible fitting of the test molecule, we use a superposition force field and carry out conformational searching and energy minimization using internal torsional degrees of freedom of the molecule to be fitted.

By searching for conformations which co-minimize superposition and internal energy, it is possible to flexibly fit the test molecule in a manner which takes internal strain directly into account. Like the superposition force field, the concept of energy co-minimization is not new. It has previously been successfully applied in situations where the atom pairs, which are specified explicitly by the user, were constrained to superimpose [9]. By combining a superposition force field, energy co-minimization and a powerful conformational search procedure [10,11], we have developed a template fitting program that is suitable for complex flexible molecules (Fig. 1).

The program uses a standard force field for internal energy, combined with a superposition potential function. An extensive conformational search is carried out based on torsion Monte Carlo sampling with co-minimization of internal and superposition energy. By using torsion energy minimization, the method can be made sufficiently rapid for a typical search to be completed in a few minutes on a VAX 8820. The program is easy to use. It is initiated by input of a structure file of the template and

test ligand and generates a set of matched structure and result files without further user intervention. To apply the method, the structure is set up as shown in Fig. 1A and the program carries out a standard series of operations and returns the structures with the best overlaps. The program also produces information about strain energy and a superposition score.

Although the method is easy to use, to be really useful it must also be reliable. Given a test compound which can match a pharmacophore, does the method find the correct bioactive conformation? To answer this question we have applied the method to flexible ligands, where the bioactive conformation is known from the X-ray structure of the ligand–enzyme complex.

This paper reports the results of comparisons of predicted and experimental conformations of inhibitors of three enzymes. For each enzyme, one of the known ligand conformations is taken as a pharmacophore and the remaining ligands are matched.

Methods and data

Outline of the TFIT program

The job is first set up interactively at a graphics terminal as described below in step (i). The remaining processes are carried out by the program.

(i) Align test ligand and template. The test ligand structure is energy minimized using the MacroModel [13] implementation of the AMBER force field [14] to obtain good bond angles and bond lengths. The test ligand is interactively translated and rotated, so that a common feature of the test ligand and the template is superimposed (see Fig. 1A). The atoms which are superimposed are called the root atoms. A single MacroModel structure file, containing both molecules superimposed, is input to the program. TFIT distinguishes between atoms of the template, atoms of the test compound which move and fixed atoms of the test compound by using atom color information, which is a standard feature of MacroModel structure files. This avoids the need to submit lists of atom numbers or several separate structure files. The atoms are colored by the user using MacroModel and written to output as a structure file. The colors shown in Fig. 1 were used for illustration purposes only. In a structure submitted to TFIT a simpler coloring, based on the needs of the program, is used (the template atoms are yellow, root atoms of the ligand are red, and moving atoms are green). When the program starts, it uses atom colors to assign torsional degrees of freedom which allow conformations of the test ligand to be generated without moving the root atoms (Figs. 1B,C). The color scheme is maintained in the output structure file of superimposed molecules. The color coding makes it very easy to use the program and enables anyone viewing the results to understand immediately how the computation was set up.

(ii) Estimate global minimum energy of test ligand. A rapid search for low-energy ligand conformers is carried out, in the absence of a superposition force field, to estimate the global minimum internal energy of the test ligand. This energy is subtracted from the internal energy of fitted conformers to obtain the energy cost or strain energy for formation of that conformation.

(iii) Search for well-fitted ligand conformers. This is a search for ligand conformers with low superposition plus internal energies. The search uses Monte Carlo sampling and minimization, with a number of features added to make the search more rapid and efficient.

(iv) Output structures and energies. A file of superimposed structures and a list of superposition and strain energies is generated. The output is ordered, so that the conformers with lowest total energy come first.

Outline of the superposition force field

The use of a superposition force field is central to the present method. The superposition energy is a readily computed measure of the match between a given conformer of the test ligand and the template. It is defined so that when an atom of the test ligand approaches an atom of the template of the same type, an attractive force is experienced (Fig. 1B). The superposition energy reaches a minimum value when the atoms are superimposed. The distance-dependent superposition potential function is defined so that interactions only take place when the atoms are closer than a specified cutoff distance. This enables test atoms to be specifically forced to superimpose on template atoms that are close and avoids large numbers of conflicting long-range interactions. Atoms are classified as hydrogen bond acceptor, hydrogen bond donor, hydrophobic or charged (atoms with formal positive or negative charges). Thus, for example, a hydrogen bond acceptor atom on the test ligand will experience an attractive force to any nearby hydrogen bond acceptor atoms of the template, but not to atoms of other types.

The superposition potential function defines the range of the attractive force and the energy which can be gained by superposition of similar atoms. The total superposition energy is added to the internal molecular mechanics energy without further scaling and this combined energy is co-minimized and used to rank order the fitted conformations.

Details of the program

The superposition potential function

The superposition potential applies an attractive force between atoms of the test ligand and template when these atoms are of similar type. The energy between a pair of atoms i and j is the product of three terms: a user-defined energy scaling coefficient (E_{sup}) which is negative; a simi-

larity factor (K_{ij}) which is specific for each pair of atom types; and a distance-dependent function. The distance-dependent function has a maximum value of 1 at zero distance and declines smoothly to 0 at a cutoff distance determined by the user (D_{sup}). Different shapes of the potential function curve were tested using the mathematical functions described in step (iii) below.

(i) Description of atom types The similarity factor is based on the values of three chemical properties which are assigned to each atom.

(1) Polarity. The following rules are sequentially applied to each atom: $P_i=1$ for hydrogens attached to a polar atom (hydrogen bonding hydrogen), $P_i=2$ for heteroatoms other than halogens, sulfur or phosphorus (polar atom), $P_i=3$ for carbon, phosphorus or sulfur attached to an oxygen through a double bond (polar carbonyl, phosphonyl, sulfonyl), $P_i=2$ for sulfur or phosphorus if attached to an atom classified as $P_i=2$ (polar atom), $P_i=2$ for carbon attached to an atom classified as $P_i=3$ (polar atom) and $P_i=4$ for carbon, phosphorus, sulfur and halogen atoms that have not already been classified (hydrophobic atom). For molecules containing carbon, hydrogen, nitrogen, sulfur, phosphorus and halogens, every atom receives a non-zero value for P_i , except for hydrogen atoms attached to carbon if these are explicitly present.

(2) Charge. $C_i=0$ for zero formal charge, $C_i=1$ for a cation atom and $C_i=-1$ for an anion atom. When charges are distributed in a conjugated system, all atoms which carry a formal charge in any of the equivalent resonance form assignments get a value of +1 (cation) or -1 (anion). For instance, the two oxygens of a carboxylate are both set to -1, whereas the three nitrogens of a guanidinium ion are all set to +1. Values greater than 1 are not used.

(3) Hydrogen bonding. $H_i=1$ for non-protonated heteroatoms with a lone pair (donor), $H_i=2$ for protonated heteroatoms without a lone pair (acceptor) and $H_i=3$ for protonated heteroatoms with a lone pair (donor or acceptor).

(ii) Potential energy for perfect superposition The minimum superposition potential energy (K_{ij}) for an atom i of the fitted compound and an atom j of the template is defined by the rule:

```
IF  $C_i$  and  $C_j$  are of opposite sign and both are non-zero THEN
   $K_{ij}=0$ 
ELSE
   $K_{ij}=E_{\text{sup}} \{ \text{Sim}(P_i, P_j) + \text{Sim}(C_i, C_j) + \text{Heq}(H_i, H_j) \}$ 
```

where $\text{Sim}(m,n)=0$ if $(m \neq n)$ or $(m \text{ or } n=0)$; $\text{Sim}(m,n)=1$ if $(m=n)$ and $(m > 0)$; $\text{Heq}(m,n)=0$ if $m=0$ or $n=0$; $\text{Heq}(m,n)=1$ if $(m=n)$ and $(m \neq 0)$; and $\text{Heq}(m,n)=1$ if $(m=3 \text{ and } n \neq 0)$ or $(m \neq 0 \text{ and } n=3)$.

Sim and Heq are functions which define similarity of

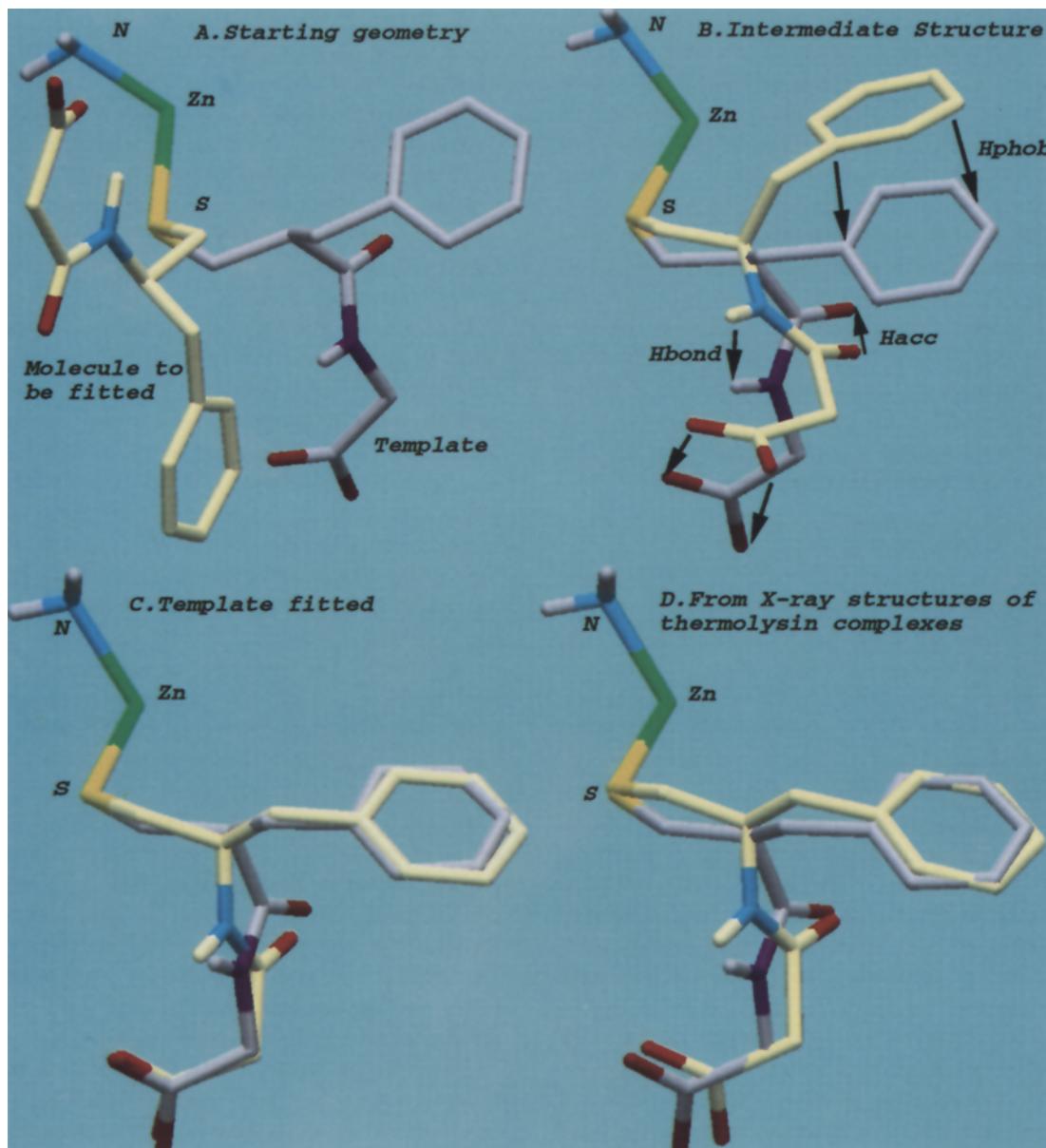


Fig. 1. Template fitting using a superposition force field. (A) Input for template fitting of retro-thiorphan, using thiorphan as a template. The retro-thiorphan–zinc complex (yellow) was energy minimized and aligned with the thiorphan–zinc complex (gray, taken from the X-ray structure of the thermolysin–thiorphan complex) to superimpose the root atoms, nitrogen, zinc and sulfur. Rotatable retro-thiorphan dihedral angles were then randomly perturbed without moving the root atoms. (B) Intermediate in the fitting process. Conformers of retro-thiorphan are generated by rotating atoms about free torsion bonds in a manner which does not move the superimposed root atoms. When atoms of retro-thiorphan are close to atoms of similar type in the template, a force is experienced as shown by the arrows in the diagram. Hphob = hydrophobic, Hacc = hydrogen-bond acceptor, Hbond = hydrogen-bonding hydrogen. (C) The result of co-minimization of internal energy plus superposition energy of the structure in B. Only atoms of retro-thiorphan are allowed to move. The force field combines intermolecular superposition energy with the internal torsional and van der Waals energy of retro-thiorphan. (D) Structures of thiorphan and retro-thiorphan observed in the binding site of thermolysin by X-ray diffraction analysis. The composite structure shown in A is input to TFIT, with the atoms colored to indicate which atoms belong to the template and which atoms move.

atom types. The logic for H_{eq} is necessary to take account of atoms which can be treated as hydrogen bond donors or acceptors, for example a hydroxyl oxygen. E_{sup} is a scaling factor for energy which can be varied by the user and is always negative.

With these definitions, the minimum superposition

potential energy can be as low as $3E_{sup}$ for a combination of charge, polarity and hydrogen bond matching. The hydrophobic energy will never be lower than E_{sup} for a given pair of atoms. Hydrophobic pair matches will, therefore, often have a smaller absolute score than hydrogen bonding matches. It was not found necessary to in-

crease the hydrophobic score to compensate for this because, typically, a hydrophobic atom of the test compound will overlap with several hydrophobic atoms in the template since these atoms tend to be adjacent. As a result, several hydrophobic superposition terms may be summed for each hydrophobic ligand atom. Hydrogen bonding and charge interactions usually involve only single atom matches.

(iii) *Distance-dependent potential energy* The superposition energy at distance D_{ij} is calculated using one of three alternative expressions, depending on which type of potential function has been selected for the application:

```

IF  $D_{ij} > D_{sup}$  THEN
   $E_{ij} = 0$ 
ELSE
   $E_{ij} = K_{ij} (D_{sup} - D_{ij})^2 / D_{sup}^2$  (convex potential)
or
   $E_{ij} = K_{ij} (D_{sup} - D_{ij}) / D_{sup}$  (linear potential)
or
   $E_{ij} = K_{ij} (D_{sup}^2 - D_{ij}^2) / D_{sup}^2$  (concave potential)

```

where D_{sup} is the cutoff distance defined by the user.

The convex potential applies an attractive force which increases as atoms approach. The force is $2K_{ij}/D_{sup}$ when the distance is zero, i.e., the atoms superimpose perfectly. The energy gradient and resulting force decrease at greater distances, becoming zero at the cutoff distance. The linear potential applies a constant attractive force of K_{ij}/D_{sup} at all distances less than the cutoff distance. The concave potential applies zero force at the point of superposition and the force increases to $2K_{ij}/D_{sup}$ at the cutoff distance. At greater distances, the force is zero.

Energy minimization

Energy minimization is the rate-limiting step in the template fitting procedure. The present method uses a torsional minimizer which is approximately two orders of magnitude more rapid compared to Cartesian minimization. To minimize a structure, each torsion angle is sequentially rotated to an energy minimum. After all the rotatable torsion bonds have been energy minimized, the minimization cycle is repeated through all rotatable torsion bonds until no further movement of more than 1° occurs for any of the torsions. The individual torsion bonds are energy minimized by a line search procedure which calculates the energy at 5° rotational steps, taken in a direction of decreasing energy. Once a minimum has been detected (i.e., the energy falls between steps $n-1$ and n and rises between steps n and $n+1$), the minimum position is more accurately estimated using a quadratic fit. Usually three to four steps are required to minimize the energy for each torsional rotation. The calculation typically converges after only three to four complete cycles through the molecule.

The method only requires calculation of energy differences due to the torsional movements (torsional, nonbonded and superposition energy), and it is not necessary to calculate derivatives of energy. The speed of the calculations is further assisted by preparing torsion bond atom assignment lists and energy look-up tables for the rotational energies around each rotatable bond.

At the start of the program, the test structure is analyzed to assign rotatable bonds. A set of moving atoms is listed for each rotation. This set is chosen so that only the nonrooted part of the structure is moved. Torsional energies are computed for each rotatable bond at 1° resolution at the start of the program and stored in look-up tables. These tables are used to compute energies due to torsional changes. During a minimization, the change in nonbonded energy for a torsional change is only calculated for the atoms which are moved.

Energies were calculated using the AMBER force field [14]. Van der Waals energies were set to zero above the van der Waals cutoff distance. At shorter distances, the energy was taken as the energy computed with the Lennard-Jones equation for the observed distance minus the energy computed for the cutoff distance.

Estimating the global minimum of the test compound

A method was found which requires only a few search steps and frequently obtains the global minimum energy or a value close to it for small molecules. Each torsion is successively explored. The torsion angle being explored is set to values which are at local energy minima. For each angle, the entire structure is torsionally minimized, and the lowest energy conformer is kept and used for exploration of the next torsion. Once all torsions have been explored, the cycle is repeated until no further changes occur in the structure, i.e., no new low-energy conformers are found.

Searching for the best matches

In addition to using a rapid minimizer, the conformational search was further accelerated by adding a number of features to improve the efficiency of a Monte Carlo sampling plus minimization algorithm [10]. The basic algorithm selects a structure from a store of low-energy conformers and makes random changes to a limited number of torsion angles. The new structure is minimized and stored if the energy is less than a threshold value above the lowest energy structure in the store. This method returns a set of low-energy structures. The use of a restricted energy window and limited perturbations tends to be efficient, because the best structures found are taken as starting points for further exploration. For template fitting, difficulties were experienced in setting the energy threshold and the number of bonds to be perturbed at each step. Sometimes, near the start of the search, a low-energy solution would be found which was conformation-

ally very different from the best solution. The search would then become trapped, i.e., it would only find conformers close to this solution. To prevent this from happening, it was necessary to use higher energy windows and larger numbers of bond perturbations. At this point, however, the search does not take full advantage of existing low-energy structures since the highly perturbed conformations are largely independent of the geometry before perturbation. This problem was solved by introducing a modification which mimics simulated annealing algorithms. The energy window and the number of bonds perturbed are set to be high at the start of the run and are progressively lowered. It was found that this greatly improves the efficiency of the procedure by reducing the necessary number of search steps. A further improvement was achieved by filtering the perturbed structures before minimization. The filter was based on superposition energy. Only structures with at least 60 percent of the best superposition energy found to date were accepted for further minimization. The internal energy was also tested for use as a filter and found to be inefficient. The starting energy is a poor predictor of minimized energy. This, in fact, is one of the main reasons why introduction of minimization into Monte Carlo conformation searching represented a major advance in conformational searching methods [10].

With these modifications to the searching procedure, it was found to be possible to complete the search with fewer energy minimizations. The program was set up to automatically control the length of the search. With the default setting, the maximum number of minimizations is calculated by multiplying the number of rotatable bonds by 100 and is not allowed to exceed 1500. The default setting was used throughout this study. For large molecules, a thorough search may require a larger number of steps. This number can be set by the user.

Sources of structures

The following enzyme inhibitor complexes were obtained from the Brookhaven Protein Data Bank (PDB) [15]: thermolysin complexed with ZFLA (Cbz-Phe^p-L-Leu-L-Ala, 5TMN) [19], ZGLL (Cbz-Gly^p-L-Leu-L-Leu, 5TMN) [18], phosphoramidon (1TLP) [19], CLT (*N*-(1-carboxy-3-phenylpropyl)-L-Leu-L-Trp [20], HOX-BAG (HONH-(benzylmalonyl)-L-Ala-Gly-*p*-nitroanalide, 5TLN) [20], and P-Leu-NH₂ (*N*-phosphoryl-L-leucinamide, 2TMN) [16]; HIV-1 protease complexed with acetyl pepstatin (5HVP) [22], L-700,417 (P4PHV) [23], and A-74,704 (9HVP) [24]; and endothiapepsin complexed with L-364,099 (2ER0) [25], H-256 (2ER6) [25], H-261 (2ER7) [26], and L-363,564 (2ER9) [25]. The letters after the compound name or code name are PDB identifiers. Where necessary, HIV and endothiapepsin complexes were aligned to a common frame of reference by rotation and trans-

lation, in order to superimpose the catalytic aspartic acids of the active site.

Additional inhibitors complexed to thermolysin, for which only the inhibitor coordinates have been published, were thiophan [12], retro-thiophan [12], and BAG ((2-benzyl-3-mercaptopropanoyl)-L-alanylglycinamide) [18]. For the analysis, polar hydrogens were added using MacroModel [13]. For the inhibitors for which the complete thermolysin/inhibitor complex coordinates were not available, minimization was performed using the active site of 5TMN and the MacroModel [13] implementation of the AMBER force field [14]. During the energy minimization, Glu¹⁴³, Leu²⁰², Leu¹³³, Val¹³⁹ and the inhibitors were allowed to move without constraints; the other atoms of the active site were constrained with a 10 kJ/Å² force constant. The high-resolution structures ZFLA, ZGLL, CLT and P-Leu-NH₂, and all HIV-1 protease and endothiapepsin complexes were used without energy minimization.

The inhibitors selected for fitting were first energy minimized in the absence of an enzyme binding site using the MacroModel [13] implementation of the AMBER force field [14]. The inhibitors were then aligned with a template before using TFIT. Thermolysin inhibitors bind to zinc through a variety of zinc-binding functional groups, e.g. sulfur or carboxylate. To position these groups into a common reference frame, each molecule was translated and rotated to superimpose its functional group on a standard reference geometry for that functional group, complexed to zinc. The zinc complex reference geometry for each type of functional group was taken from one of the thermolysin/inhibitor complexes. Reference geometries were taken from thiophan for inhibitors bound through a sulfur, from phosphoramidon for inhibitors bound through a phosphorus acid, from CLT for inhibitors bound through a carboxylic acid, and from HOX-BAG for inhibitors bound through a hydroxamic acid. To obtain a starting alignment, the functional group of the test compound was superimposed on that of the reference compound. All the thermolysin/inhibitor X-ray coordinates are in the same enzyme reference coordinate frame, so that the alignment of the ligated functional groups described above automatically positions them correctly with respect to each other.

Results

The inhibitors of thermolysin and HIV-1 protease, shown in Fig. 2, were used to evaluate the method presented here. These inhibitors were selected because their bioactive conformations are known from the X-ray crystal structure of the appropriate enzyme inhibitor complex. The bioactive conformations of one or more of the inhibitors were used to form a template and the other inhibitors were flexibly fitted to this template.

Application to thermolysin inhibitors

Alignment of the molecules

The method in its present form only searches geometries of the test molecule generated by variation of torsion angles. As a result, a common feature must be present in the test substance and the template which can be used for a preliminary alignment of the molecules. The zinc-binding functional group of thermolysin inhibitors is a natural feature to use as a root (Fig. 1A), since these groups will all occupy similar positions in the enzyme-ligand complex. The functional groups which bind to zinc are of different chemical types. However, it is possible to obtain a reference orientation for each type of functional group from an appropriate enzyme structure (see the Methods section). The coordinates of the functional group along with zinc and one of the imidazole nitrogens of the enzyme are used. The molecules to be tested are complexed to the zinc-nitrogen fragment. Each molecule now has a common feature which can be used for superimposition (see Fig. 1A).

Results of template fitting

Eight thermolysin inhibitors, bound to zinc, were

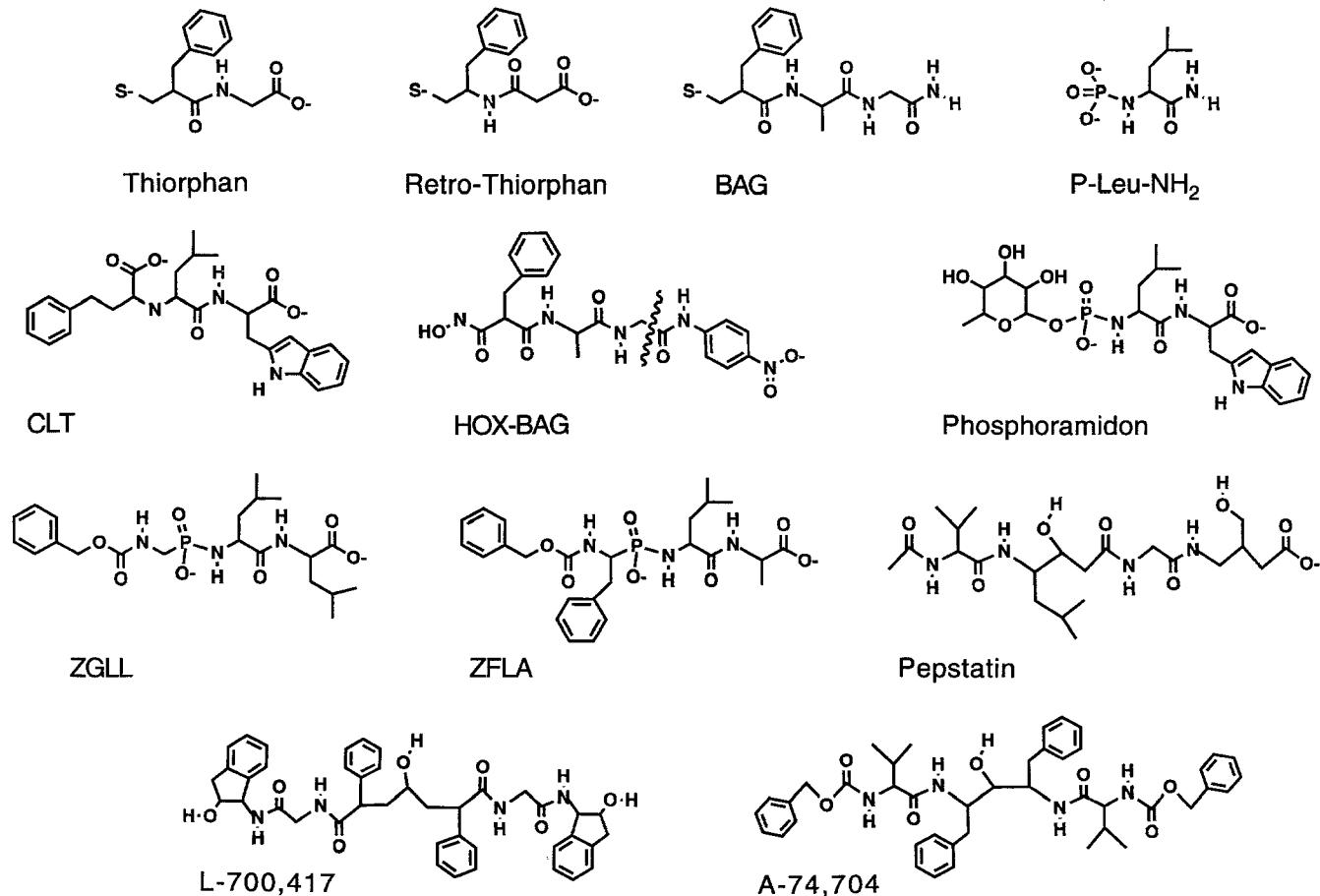


Fig. 2. Inhibitors used to evaluate TFIT. See the Methods section for references. The portion of HOX-BAG lying to the right of the wavy line was not defined in the X-ray analysis and was not included in the structure used for template fitting.

aligned using the X-ray structure of ZFLA as a template. The inhibitors were then fitted to the template using TFIT, and the match between the fitted structure and the bioactive conformation (X-ray structure of the complexed inhibitor) was examined.

Parameters for the superposition force field were determined by preliminary experimentation. We expected that a suitable energy for perfectly superimposed matching atoms would be on the order of 5 kJ, corresponding very approximately to energies which might be achieved in a binding site through hydrophobic or hydrogen bonding interactions. The cutoff distance was expected to be between 0.5 and 5 Å, to allow for a certain amount of movement which might occur in a binding site. Our experiments showed that, if the energy was too high or the cutoff distance too short, the atoms would be attracted too strongly to the template, resulting in high internal energies. In addition, a short cutoff made the search much more lengthy, since the conformation, generated by a Monte Carlo perturbation prior to minimization, had to be very close to a matching conformation before the attractive force would minimize the structure to a well-matched solution.

The preliminary experiments showed that values of

-5.0 kJ for the superposition energy (E_{sup}) and 3.0 Å for the cutoff distance (D_{cut}), combined with the molecular mechanics parameters given in the footnote to Table 2, gave reasonable results, as shown in Table 1 and Fig. 3. The effect of varying these values is discussed in the next section. The result for thiorphan (Fig. 3A) shows that a leucyl side chain of ZFLA acts as a very effective guide for the phenylalanine of thiorphan. For retro-thiorphan (Fig. 3B), the program correctly aligned the phenyl group and the retro-amide bond. To align the latter bond, a different geometry is required for the C^α and C^β of the phenylalanine, compared to the normal amide. This geometry was predicted correctly. It was clear, however, that in some compounds not all of the flexible side chains were correctly positioned by fitting. For instance, although much of the fitted CLT molecule (Fig. 3C) lies close to the crystal structure, the tryptophan has not been aligned correctly. This is to be expected, since in the bioactive conformation the indole ring does not lie close to any part of the template. This illustrates an important limitation of the use of templates for compound design, namely that only the parts of the new compound which are represented by template atoms can be modeled with confidence using the template.

Table 1 shows fitting statistics calculated for those 'matchable' atoms which can be expected to be correctly fitted, i.e., where the atom in the bioactive conformation lies close to a template atom. When this criterion was applied, satisfactory results were found for all compounds, with phosphoramidon as the single exception. Phosphoramidon contains a sugar residue. Atoms of this residue are found in a binding pocket occupied by a phenyl group of ZFLA, and the atoms of the sugar do not correspond well to atoms of the phenyl with regard to atom type or specificity.

It is possible to use more than one molecule to form a template. The combination of the X-ray structures of ZFLA and ZGLL provides a template which gives more guidance for the tryptophan of CLT (Fig. 4A). The positioning of the tryptophan of CLT is improved, but the indole ring is still not oriented correctly. Combining the X-ray structures of ZFLA and phosphoramidon gives an even better template, which now results in a fitted conformation for CLT that is close to the X-ray geometry (Fig. 4B).

These results show that the method only works reliably if the template has atoms which can provide a guide to the alignment of ligand atoms of similar type.

Effects of force field parameters

Table 2 shows how the effectiveness of the method depends on the correct choice of force field parameters. The most sensitive parameter is the shape of the forcing potential. Changing to a linear potential greatly decreased the fitting efficiency. Earlier experiments using a concave

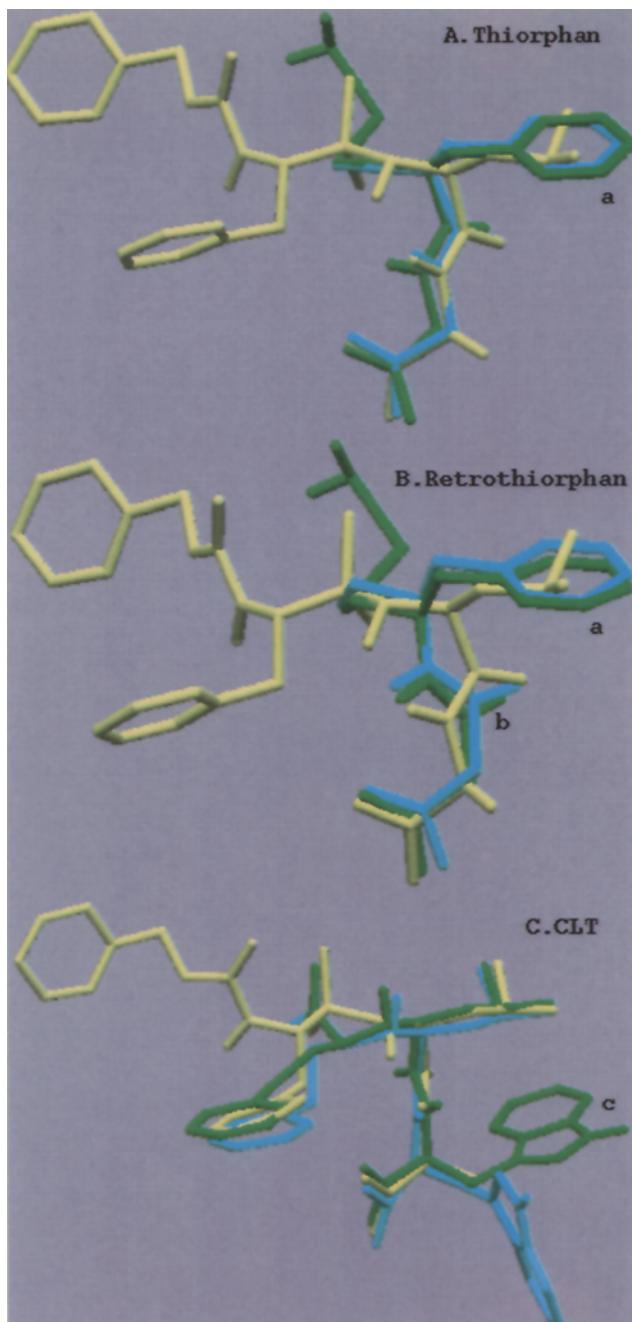


Fig. 3. Comparison of fitted versus experimental conformers of thermolysin inhibitors using, as a template, coordinates from an X-ray analysis of the thermolysin-ZFLA complex (see Fig. 1 for references and details). Templates are yellow, fitted structures are green and structures determined by X-ray analysis of the thermolysin complex of the test compound are blue. (A) thiorphane; (B) retrothiorphane; (C) CLT. Before fitting, the inhibitors were aligned to ZFLA using a reference geometry for the functional group bound to zinc (see the Methods section). The phenyl groups (a) are well oriented by the leucine side chain of ZFLA. The retro-amide bond (b) is also well fitted. The indole side chain of CLT is not oriented correctly (c), because no template atom is in the appropriate region to guide the orientation.

(quadratic potential) had given even less satisfactory results. Smith [8] suggests that the shape of the potential

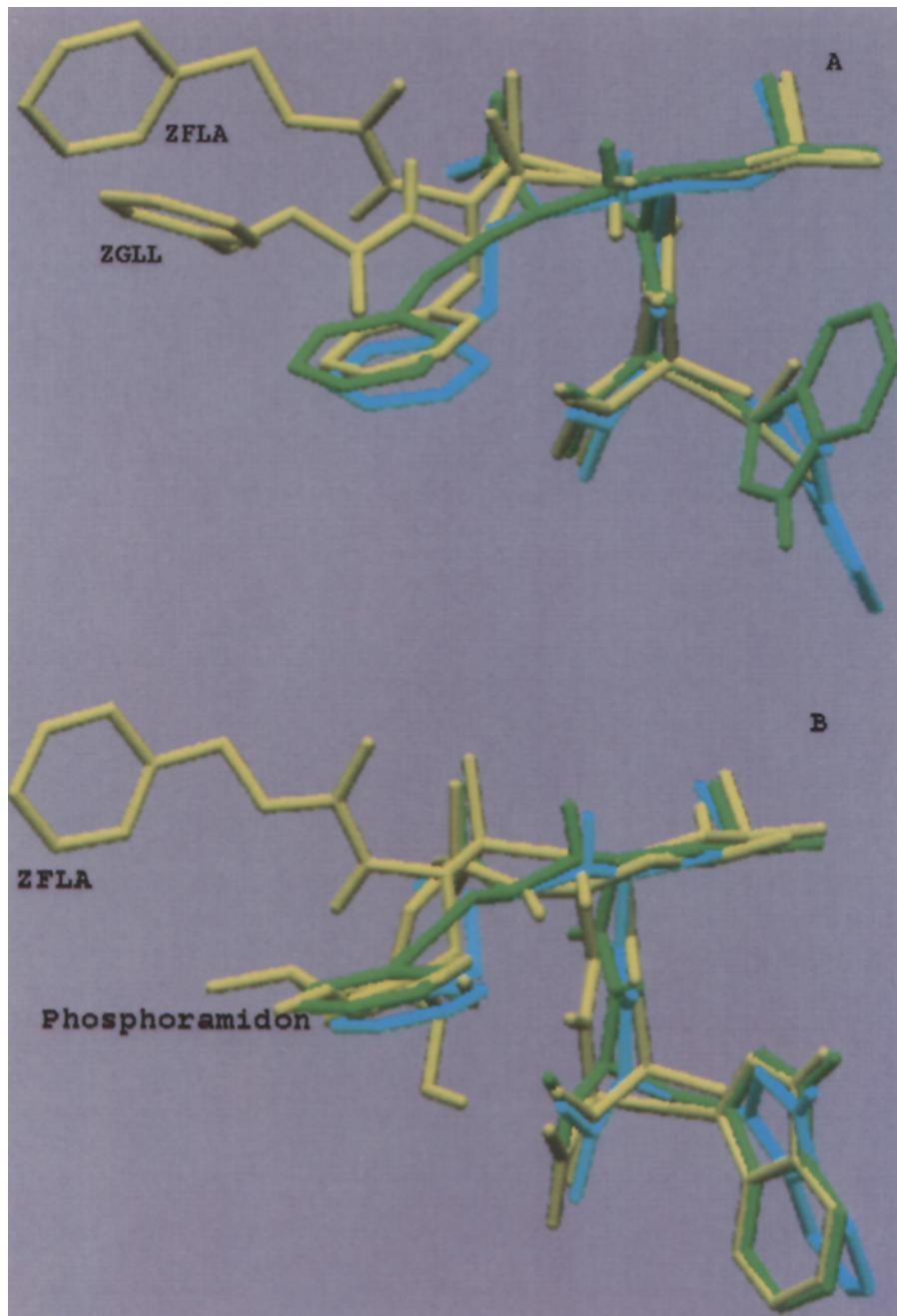


Fig. 4. Effect of template on fitting. Templates (taken from the X-ray structures of thermolysin complexes) are yellow, fitted structures are green and bioactive conformers of the test compound, determined by X-ray analysis of the thermolysin complex, are blue. (A) CLT fitted to a template based on both ZFLA and ZGLL. The valine side chain of ZGLL provides a better orientation of the indole ring of CLT than ZFLA alone (Fig. 2C). (B) CLT fitted to a template based on both ZFLA and phosphoramidon. The indole ring of phosphoramidon provides a better guide for the indole ring of CLT and the rings of the template and test compound are perfectly superimposed and cannot be distinguished in the figure. This orientation is close, but still does not exactly correspond to the X-ray structure.

is critical and recommends the use of a convex potential. This finding is of interest, because a quadratic potential corresponds to the standard method of performing superpositions using a least-squares fit. The convex potential probably works better because it allows a weak force to operate when similar atoms are at some distance from each other, while providing a powerful force for atoms which are close to each other. The weak force at longer

distances enables atoms which can be potentially superimposed to be recognized, thus increasing the chance that a test conformer will be minimized to yield a good superposition. The concave potential, on the other hand, apparently provides too much attractive force for atoms which are not close and at the same time does not adequately force superposition of close atoms.

The cutoff distance for the forcing potential is also

critical. When the range is too small, the search for matches becomes less efficient because fewer of the randomly perturbed conformers place atoms within the range of the forcing potential field of the template and, as a result, superposition will not be forced during minimization. If the range is too large, the forcing potential becomes distributed over too many atoms and will lack spatial resolution.

The superposition energy constant (E_{sup}) is also important. A too low energy allows internal forces to dominate and may not force the conformers to optimize the overlap. A too high energy may result in too much straining of the conformer. Going from 5 to 10 kJ reduces the rms deviation for matched atoms slightly, but causes an increase in strain energy. The default value of -5 kJ/mol gives the greatest number of successfully matched molecules.

In addition to affecting the geometries of superimposed structures, the superposition energy constant plays a second, very important role. The program returns a number of fitted solutions, ordered according to the combined superposition and internal energies. If the superposition energy constant is set high, then badly strained conformers may appear at the top of the list if the atoms match well. Setting the constant to a low value will give results that are ordered mostly by internal or strain energy. In practice, we use -5 kJ/mol but we routinely examine more than the first solution. The results are most convincing if the top structure is well separated in total energy from the next solution, which has a significantly different geometry. When this is not true, the fitting exercise will usually still yield valuable insights.

The parameters used for computing internal energy were also tested. We found no advantage in including van der Waals attractive energy; this usually results in a

higher apparent strain value, because a global minimum will often be found where the molecule folds up to maximize the van der Waals energy. This energy is the result of using an uncorrected vacuum force field, which is not appropriate for solution conformers. Electrostatic energy was also found to make little difference in the present tests. For structures where internal hydrogen bonding may occur in the fitted molecule, it may be necessary to include internal electrostatic energy or an equivalent hydrogen bonding energy term.

Although the superposition force field parameters are critical to the performance of the method, we have found the default values surprisingly robust and it has not been necessary to change them in numerous applications.

Application to HIV inhibitors and endothiapepsin inhibitors

The X-ray structure of the pepstatin/HIV-1 protease complex was used as a template [22]. Pepstatin has two docking modes and these were combined to provide a template. The combination provides a much more complete definition of the binding pockets than that obtained with a single docking mode (Fig. 5). Two test structures were fitted to this template. The test structures and the template all have a backbone hydroxyl group, providing a tetrahedral geometry. The hydroxyl and the attached carbon were used to align the test structures to the template. The results (Fig. 5) show that the fitting procedure was effective for the backbone and many of the side chains of the structures. However, for A-74,704 (Fig. 5A) the terminal phenyl groups were not aligned correctly. The fitting procedure actually found an alignment between the phenyl and valine side chains which looks highly plausible. However, the X-ray structure of A-74,704 has the phenyl groups in a place which the template does

TABLE 1
MATCH BETWEEN THE STRUCTURES OF INHIBITORS FITTED TO A ZFLA TEMPLATE^a AND THEIR BIOACTIVE CONFORMATIONS DETERMINED BY X-RAY ANALYSIS OF THE INHIBITORS COMPLEXED WITH THERMOLYSIN

Compound	Number of atoms			Rms of matched atoms (Å)	Time ^b (min:s)
	Total	Matchable ^c	Matched ^d		
Thiophan	18	17	17	0.33	6:36
Retro-thiophan	18	17	17	0.52	6:54
BAG	26	20	20	0.52	14:37
P-Leu-NH ₂	16	16	16	0.44	3:38
CLT	38	30	28	0.56	33:06
HOX-BAG	29	20	16	0.82	7:58
Phosphoramidon	43	31 ^e	21	0.73	25:15
ZGLL	35	24	22	0.52	20:24

^a The template is based on the X-ray structure of the ZFLA/thermolysin complex [18].

^b CPU time on a VAX 8820.

^c Atoms of the test compound are considered to be matchable if an atom in the X-ray structure of the test compound is less than 1.5 Å from an atom of the template.

^d Atoms are considered to be matched if the fitted atom is less than 2.0 Å from the atom in the X-ray structure of the test compound.

^e Nine of these atoms are in a sugar residue which is not structurally similar to the corresponding part of ZFLA.

TABLE 2
EFFECT OF FORCE FIELD PARAMETERS ON PREDICTION OF BIOACTIVE CONFORMATIONS BY FITTING TO A TEM-PLATE (ZFLA)^a

Parameter ^b	Strain energy ^c range (kJ/mol)	No. of matched ^d compounds	Matched atoms ^e (rms (Å))
Default ^f	16.5–2.5	8	0.56 (0.15)
Linear ^g	14.4–2.3	5	0.54 (0.17)
D _{sup} =4.0	17.6–2.4	6	0.54 (0.13)
D _{sup} =2.0	14.0–2.4	6	0.54 (0.15)
E _{sup} =–10.0	22.6–3.1	7	0.53 (0.14)
E _{sup} =–2.0	7.5–1.9	7	0.65 (0.13)
V _{cut} ^h =1.5	29.4–2.0	6	0.54 (0.19)
Charge ⁱ d=4.0	14.7–2.4	8	0.60 (0.16)

^a The eight test compounds shown in Fig. 1 were used. After Monte Carlo searching, the lowest energy structure of each inhibitor was compared to the experimentally determined structure. Values for the test compounds are combined.

^b Single changes are made to the default force field, as specified.

^c Internal energy of the fitted molecule minus estimated global minimum of the free molecule. Values for the compounds with the lowest and highest strain energies are listed.

^d A compound matches if the number of matched atoms is at least as great as the number given for that compound in Table 1.

^e See Table 1 for definition of a matched atom. The mean and standard deviation of the rms values for matched atoms of fitted molecules compared to X-ray structures are reported.

^f Convex superposition potential; superposition energy: E_{sup}=–5.0 kJ/mol; superposition range: D_{sup}=3.0 Å; van der Waals cutoff ratio: V_{cut}=1.0; no electrostatic energy.

^g A linear superposition potential function was used (see Experimental section).

^h V_{cut} is the ratio of the sum of the van der Waals distances to the cutoff distance. V_{cut}=1.0 means that only repulsive van der Waals energy will be used. V_{cut}=1.5 sets the cutoff to approximately 6 Å for aliphatic carbon atoms.

ⁱ A distance-dependent dielectric constant of 4.0 r was used.

not indicate to be hydrophobic and which could not have been anticipated given only the template structure.

The results for L-700,417 (Fig. 5B) show that the ability of phenyl groups to bridge two side chains of pepstatin was predicted correctly. The bulky hydroxyl-indanyl side chains of the fitted compound are located in the correct region, but there is not enough information in the template to position them accurately.

The method was also applied to endothiapepsin inhibitors. The results confirm the findings for thermolysin and HIV inhibitors and are not presented in detail. The inhibitors are large and only fragments were selected for template fitting.

Iso-valeryl-His-Pro-Phe-His-5-cylohexyl-4-amino-3-hydroxy-pentanoamide (2ER0) [25] was fitted to the X-ray structure of tBoc-His-Pro-Phe-His-statine (2ER7) [26]. Alignments were made using the backbone aliphatic hydroxyl group. The fitted inhibitor showed an excellent alignment for all parts of the structure.

Pro-Phe-His-Leu-CH(OH)-CH₂-C(=O)-Leu-NH₂ (2ER9) [25] was fitted to Pro-Thr-Glu-Phe-Phe-Ala-Glu (2ER6) [25], with the Phe-Phe amide bond reduced to CH₂-NH. Alignments were made using the tetrahedral carbon atom, which mimics the transition state, and atoms immediately adjacent to it. The fitted structure was in good agreement with the bioactive structure, except for the alignment of histidine. The template contained glutamic acid and phenylalanyl side chains in this region, and it is not surprising that the fitting procedure could not align the histidine correctly.

Discussion

Application to drug design

The usefulness of a template fitting program ultimately depends on the purpose for which it is used. When applied to drug design, the program will be used to evaluate molecules intended to mimic the three-dimensional arrangement of key functional groups of the template. The question here is whether a low-energy conformer of the test compound can be found which has a reasonable overlap of equivalent atom types.

In our experience, TFIT is very valuable to answer this question. The results presented here confirm this by showing that, if a well-matched superimposition with low strain energy exists, the program nearly always finds it. This means that if, for a given test molecule, the program only finds matches with a high strain energy or with a low superposition score, it is unlikely that the molecule is a good mimic from a structural point of view. This does not mean that such a compound will necessarily have low potency; the template will rarely, if ever, contain enough information to make such a prediction. It does mean that, if the goal is to produce a structural mimic for a three-dimensional template, the program can be very helpful in the design process by enabling ideas to be rapidly and objectively tested, modified and retested.

The reliability and speed of the method are the result of selecting suitable superposition force field parameters (Table 2), and the development of highly efficient search and minimization algorithms.

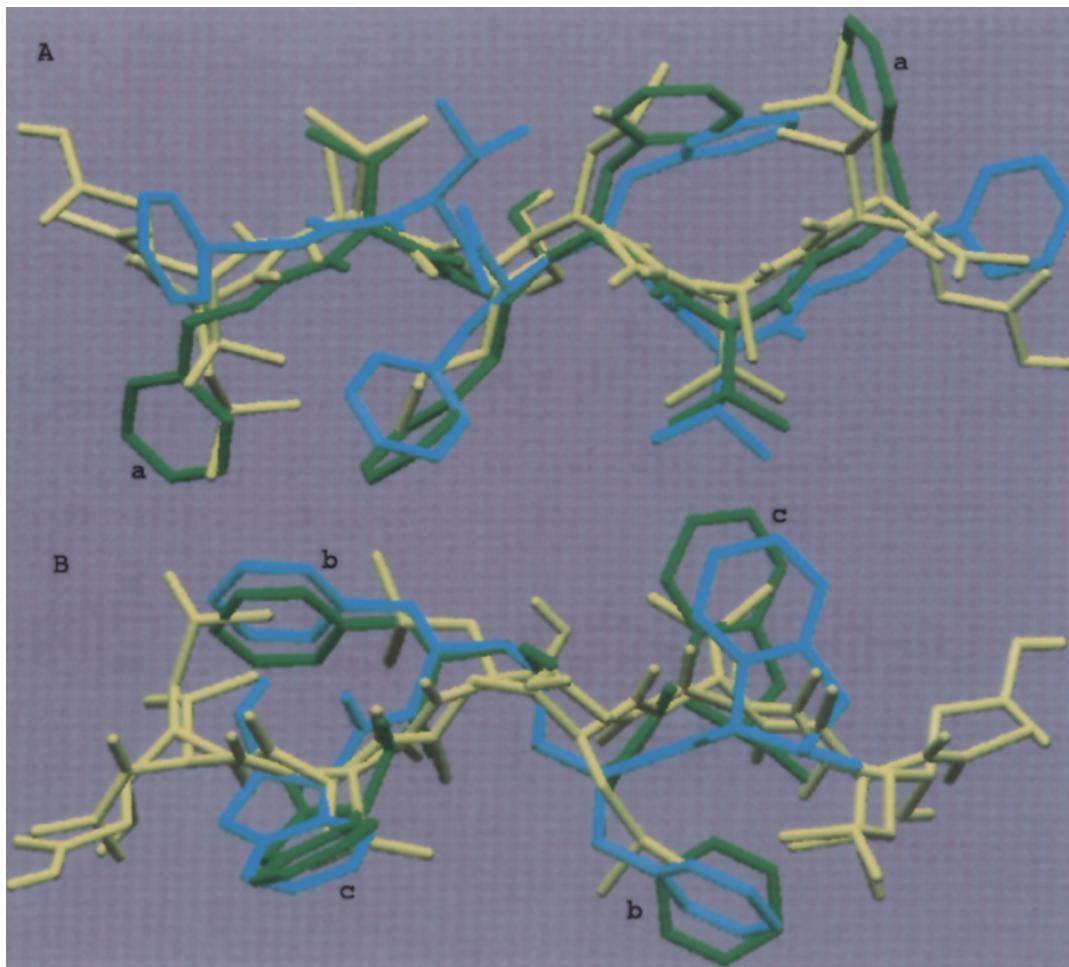


Fig. 5. Template fitting of HIV protease inhibitors to the structure of acetyl pepstatin found in the HIV-1 protease complex [22]. The two observed pseudo-symmetric docking modes of acetyl pepstatin [22] were superimposed to form a template shown in A and B (yellow). Templates are yellow, fitted structures are green and structures determined by X-ray analysis of the HIV protease complex are blue. (A) A-74,704. The fitting program has aligned the side chains with hydrophobic residues of the template. Comparison of the fitted to the experimental results [24] shows that the phenyl groups (a) of the fitted structure, which match a hydrophobic region of the template well, are actually found in a different orientation in the X-ray structure of A-74,704. The pepstatin template failed to induce the correct orientation, because it lacks hydrophobic atoms in this region. (B) L-700,417. Phenyl groups (b) which bridge hydrophobic side chains of acetyl pepstatin are correctly positioned. The bulky hydroxyl-indanyl groups (c) are also in the correct regions, although the positioning is only moderately good. These groups do not receive very effective guidance from the template.

Application to 3D QSAR

Template fitting procedures can also be used to prepare alignments of molecules for structure-activity relationship analysis using methods such as CoMFA [27] or HASL [28]. For this application, one is usually trying to obtain a set of superimposed conformers for a considerable number of molecules, covering a wide range of potencies. If the molecules are fairly rigid and differ only little, it may not be necessary to use the type of fitting procedure described in this paper.

Where the molecules are flexible, TFIT may be a useful way of aligning them. However, the present study has made it clear that the reliability of this procedure will break down if parts of the molecule are not well represented by the template. This is particularly likely to occur with the less potent molecules which are usually import-

ant for establishing a QSAR model. On the other hand, it is not certain that misalignment of certain functional groups, which may not be important for binding, will be detrimental to the subsequent QSAR analysis. Therefore, while it is hoped that procedures such as TFIT will increase the scope of 3D QSAR analysis, the value of TFIT in this respect must still be established.

Limitations of the TFIT program

The current version of the program is restricted to situations where the test ligand and the template contain a common molecular functional group, and it is reasonable to assume that this feature should be superimposed. This reduces the number of situations where the program can be applied. In addition, the fact that the root atoms do not move may result in a biased fit, since the 'best' fit

might involve a compromise where some of the root atoms do not exactly overlay the corresponding template atoms. This limitation can be readily removed by adding translational and rotational degrees of freedom to the test molecules. Originally it was intended that the extra degrees of freedom would be added immediately, but it was found that a surprisingly large number of applications could be carried out with the restricted version. Where it is possible to make an initial alignment, the search is likely to be more rapid and less ambiguous.

The method is currently also restricted to exploration of conformers based on variation of freely rotatable torsion angles. It does not search different conformers of cyclic parts in molecules, nor does it perform Cartesian minimization. There is a trade-off between the speed achieved by torsional minimization and errors which may be introduced by not allowing full Cartesian optimization. For many molecules, our experience has been that the errors are less than a few kilojoules. Such errors are not significant for determining the most likely superimposed conformation when one solution is well differentiated from the rest. However, care should be taken when some of the conformers have internal interactions which can be appreciably relaxed by full minimization, such as bad 1,4 contacts. This limitation can also be surmounted. An extended version of the program has recently been developed which allows molecules to rotate and translate freely and uses Cartesian minimization to handle conformations of cyclic systems and to fully minimize structures after torsional minimization. The program does not require a preliminary root overlap, and searches conformations of cyclic as well as noncyclic systems [29]. We have found that the new program extends the range of problems to which a superposition force field approach can be applied, but does not modify any of the conclusions of the present study.

Absolute limitations of template fitting

Intrinsic limitations of the template fitting process arise from the relative lack of information contained in a template model compared to a high-resolution enzyme binding site model. One way of considering these limitations is to regard the superposition force field as a very rough first pass approximation to the force field which might be found in a binding site. The present method makes many simplifying assumptions, for example assigning equal weight to atoms of a similar type. This is necessary when, as is often the case, the data on analogues is limited. In the present study, the values of the superposition force field parameters were established by trial and error using thermolysin inhibitors. These values seem to have rather general application, and we have not found it necessary to modify them after a large number of applications, including many using the extended program with Cartesian minimization. However, an ideal program would

make optimal use of structure-activity information of analogues, perhaps by generating a hypothetical lattice model. Models of this sort (CoMFA [27], HASL [28]) can represent the relative importance of specific functional groups and provide indications of excluded volumes. In this way, a superposition force field might be produced which more closely resembles the force field that actually exists in a binding site. The fitting algorithms described in the present paper could be readily applied to such a force field.

Conclusions

The template fitting program (TFIT) has proved to be a reliable method for evaluating molecules intended to mimic a 3D pharmacophore. The program is easy to run and provides useful information about the degree to which conformers match the template and about the internal energy cost of matching the template. The results are obtained rapidly, and it is possible to perform several iterations of design and testing in a single interactive session at a graphics terminal.

Acknowledgements

The authors wish to thank Dr. Wayne C. Guida and Dr. Frank H. Clarke for helpful discussions and advice.

References

- Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds.) Computer-Assisted Drug Design, ACS Symposium Series, Vol. 112, American Chemical Society, Washington, DC, 1979, pp. 205–226.
- Motoc, I., Dammkoehler, R.A., Mayer, D. and Labanowski, J., Quant. Struct.–Act. Relatsh., 5 (1986) 99.
- Motoc, I., Dammkoehler, R.A. and Marshall, G.R., In Trinajstic, N. (Ed.) Mathematical and Computational Concepts in Chemistry, Ellis Horwood Ltd., Chichester, 1986, pp. 222–251.
- Mayer, D., Naylor, C.B., Motoc, I. and Marshall, G.R., J. Comput.-Aided Mol. Design, 1 (1987) 3.
- Kolossvary, I. and Guida, W.C., J. Chem. Inf. Comput. Sci., 32 (1992) 191.
- Sheridan, R.P. and Venkhtaravaghan, R., Acc. Chem. Res., 20 (1987) 322.
- Dammkoehler, R.A., Karasek, S.F., Shands, E.F.B. and Marshall, G.R., J. Comput.-Aided Mol. Design, 3 (1989) 2.
- Smith, G.M., SEA: Steric and Electrostatic Alignment Molecular Superposition Program, QCPE Program No. 567, Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN.
- Duchamp, D.G., In Christoffersen, R.E. (Ed.) Algorithms for Chemical Computation, American Chemical Society, Washington, DC, 1987, pp. 98–121.
- Li, Z. and Scheraga, H.A., Proc. Natl. Acad. Sci. USA, 84 (1987) 6611.
- Chang, G., Guida, W.C. and Still, W.C., J. Am. Chem. Soc., 111 (1989) 214.

- 12 Roderick, S.L., Fournie-Zaluski, M.C., Roques, B.P. and Matthews, B.W., *Biochemistry*, 28 (1989) 1493.
- 13 Mohamadi, F., Richards, N.G., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrikson, T. and Still, C., *J. Comput. Chem.*, 11 (1990) 440. MacroModel, Version 2.5, was obtained from Professor W. Clark Still, Columbia University, New York, NY.
- 14 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., *J. Am. Chem. Soc.*, 106 (1984) 765.
- 15 a. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, T., Shimanouchi, T. and Tasumi, M., *J. Mol. Biol.*, 112 (1987) 535.
 b. Abol, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J., In Allen, F.H., Bergerhoff, G. and Sievers, R. (Eds.) *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, Data Commission of the International Union of Crystallography, Cambridge, 1987, pp. 107–132.
- 16 Holden, H.M., Tronrud, D.E., Monzingo, A.F., Weaver, L.H. and Matthews, B.W., *Biochemistry*, 26 (1987) 8542.
- 17 Monzingo, A.F. and Matthews, B.W., *Biochemistry*, 23 (1984) 5724.
- 18 Tronrud, D.E., Holden, H.M. and Matthews, B.W., *Science*, 235 (1987) 571.
- 19 Monzingo, A.F. and Matthews, B.W., *Biochemistry*, 21 (1982) 3390.
- 20 Holmes, M.A. and Matthews, B.W., *Biochemistry*, 20 (1981) 6912.
- 21 Tronrud, D.E., Monzingo, A.F. and Matthews, B.W., *Eur. J. Biochem.*, 157 (1986) 261.
- 22 Fitzgerald, P.M.D., McKeever, B.M., Van Middlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.-T., Herner, W.K., Dixon, R.A.F. and Darke, P.L., *J. Biol. Chem.*, 265 (1990) 14209.
- 23 Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., *J. Am. Chem. Soc.*, 113 (1991) 9382.
- 24 Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, N.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, W., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M., *Science*, 249 (1990) 527.
- 25 Cooper, J.B., Foundling, S.I., Blundell, J.L., Boger, J., Jupp, K. and Kay, J., *Biochemistry*, 28 (1989) 8596.
- 26 Veerapandian, B., Cooper, J.B., Sali, A. and Blundell, T.L., *J. Mol. Biol.*, 216 (1990) 1017.
- 27 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
- 28 Doweyko, A.M., *J. Med. Chem.*, 31 (1988) 1396.
- 29 QXP (Quick eXplore) is a program which contains routines for torsion space and Cartesian space searching and minimization. It has user-friendly (single command line) applications for conformational searching, template fitting and docking to binding sites. A manuscript is in preparation and, once published, the program will be submitted for public release through the Quantum Chemical Program Exchange.