Journal of Computer-Aided Molecular Design, **13:** 499–512, 1999.
KLUWER/ESCOM

Estimation of active conformations of drugs by a new molecular superposing procedure

Kazuhiko Iwase*,** & Shuichi Hirono**
School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

Received 7 July 1998; Accepted 16 December 1998

Key words: functional atom, molecular dynamics, pharmacophore, physicochemical property, stable conformation, superposition

Summary

We have developed a new program, SUPERPOSE, to superpose two molecules based on the physicochemical properties of functional atoms within individual molecules. SUPERPOSE treats a pseudo-molecule consisting of functional atoms instead of a real molecule. Four types of physicochemical properties – hydrophobicity, presence of a hydrogen-bonding donor, presence of a hydrogen-bonding acceptor and presence of a hydrogen-bonding donor/acceptor – were supposed and a score was given to each overlap. When functional atoms with the same physicochemical properties were overlapped, points were added to the score, and when the functional atoms with different physicochemical properties were overlapped, points were subtracted. We applied SUPERPOSE to 12 pairs of 24 enzyme inhibitors and found that the best scored overlay for each inhibitor pair could successfully reproduce the superposition obtained from X-ray crystallography. Next, we applied SUPERPOSE to estimate the active conformations of the thrombin inhibitors MQPA, 4-TAPAP and NAPAP. Superpositions of conformers sampled by the high-temperature molecular dynamics calculation with respect to the three inhibitors were performed, and 13 sets of conformers having the best common overlay to the three inhibitors were selected. One among 13 sets was consistent with the superposition of the active conformations derived from the X-ray crystallography of the thrombin–inhibitor complexes.

Abbreviations: ACE-Ala-Pro-Val-FPA, acetyl-L-alanyl-L-prolyl-L-valyl-difluoro-*N*-phenylethylacetamide; AGF, *O*-[[(1*R*)-[[*N*-(phenylmethoxycarbonyl)-L-alanyl]amino]methyl]hydroxyphosphinyl]-L-3-phenyllactate; AMC, aminomethylcyclohexane; BDK, 2-[5-amino-6-oxo-2-(2-thienyl)-1,6-dihydropyrimidin-1-yl]-*N*-[3,3-difluoro-1-isopropyl-2-oxo-3-[*N*-(2-morpholinoethyl)carbamoyl]propyl]acetamide; BEN, benzamidine; BZS, L-Benzylsuccinate; CBZ-PGL-Leu-Leu, carbobenzoxy-glycyl^{P}-L-leucyl-L-leucine (glycyl^{P} is used to indicate that the trigonal carbon of the peptide linkage is replaced by the tetrahedral phosphorus of a phosphonamidate group.); FVF, *O*-[[(1*R*)-[[*N*-(phenylmethoxycarbonyl)-L-phenylalanyl]amino]isobutyl]hydroxyphosphinyl]-L-3-phenyllactate; MQPA, (2*R*,4*R*)-4-methyl-1-[*N*α-(3-methyl-1,2,3,4-tetrahydro-8-quinolinesulfonyl)-L-arginyl]-2-piperidine carboxylic acid; NAPAP, *N*α-(2-naphthyl-sulfonyl-glycyl)-DL-amidinophenylalanyl-piperidine; OMTKY3, the third domain of turkey ovomucoidinhibitor; PHO-Leu-NH_2, *N*-phosphoryl-L-leucinamide. PRA, 3-Phenylpropylamine; 4-TAPAP, *N*α-(4-toluene-sulfonyl)-DL-m-amidinophenylalanyl-piperidine; TFA-Leu-Ala-ANI, trifluoroacetyl-L-leucyl-L-alanyl-*p*-trifluoromethylphenylanilide; TFA-Lys-Ala-ANI, trifluoroacetyl-L-lysyl-L-alanyl-*p*-trifluoromethylphenylanilide; TFA-Lys-Leu-ISO, trifluoroacetyl-L-lysyl-L-leucyl-*p*-isopropylanilide; TFK, 3-[[(Methylamino)sulfonyl]amino]-2-oxo-6-phenyl-*N*-[3,3,3-trifluoro-1-(1-methylethyl)-2-oxopropyl]-1(2H)-pyridineacetamide.

Introduction

In computer-assisted drug design, it is very important to determine the conformation of the ligand molecules that bind to such proteins as receptors and enzymes, i.e., the active conformation. As experimental methods to determine the active conformation of a ligand molecule, X-ray crystallography or NMR spectroscopy of the protein-containing complex have been used. However, these methods are generally accompanied by a number of difficulties, such as the difficulty of crystallizing the proteins (i.e., the receptors or enzymes), the difficulty of obtaining a sufficient quantity for analysis, and the problem of the molecular weights being

*Present address: Central Research Laboratories, Kyorin Pharmaceutical Co. Ltd., 2399-1 Mitarai, Nogi-matchi, Shimotsuga-gun, Tochigi 329-0114, Japan.
**To whom correspondence should be addressed.

too large to analyze. In computer-assisted drug design, we conveniently use conformations based on the crystal structure of the ligand molecule alone or on the structure obtained from molecular mechanics or the molecular orbital method in a system that does not clearly deal with the solvent molecules.

It has been considered that even in cases where the active conformation differs from the most stable conformation obtained from calculations, the two may still be relatively similar with respect to energy. Moreover, for conformations of the ligand molecules binding to the same receptor or enzyme, the same functional atoms occupy the same three-dimensional space. Hence, it would seem to be possible to sample many stable conformations using high-temperature molecular dynamics (MD) calculations derived from only a few ligand molecules, and to perform the superposition of these stable conformations based on the functional atom in order to estimate the active conformation common to each ligand molecule.

The superposition between two molecules is commonly performed such that the positions between the corresponding atoms coincide with each other. For this reason, there is need of an operation to determine, in advance, the pharmacologically effective atoms of the compounds. However, of many receptors the three-dimensional structures are not known. Hence, the results of interactive superposition tend to be influenced by the intuition and experience of the analyst.

Martin et al. developed a technique for automatically exploring superpositions (DIStance COmparisons; DISCO) by improving the conventional technique of matching the distance between pharmacophore centers. In DISCO, the maximal allowable error was used for the coincidence of atomic positions, and the direction of hydrogen bonding and the orientation of the aromatic ring were also taken into account. The authors applied DISCO to both the dopamine and benzodiazepine agonists [1].

Techniques to evaluate the superposition based on similarities in electron density, electrostatic potential and molecular volume have also been reported by Carbo and co-workers [2,3] and Hodgkin and co-workers [4,5], and the automatic exploration of superposition has been realized using an ASP package [6].

Perkins et al. [7] aroused much interest by developing a technique for automatically exploring the superposition between serine protease inhibitors of various sizes while taking into account the similarities in molecular surface-volume, hydrogen bonding, and electrostatic potential among these inhibitors.

McMartin and Bohacek developed TFIT (Template FITting), which used a molecular superposition force field to flexibly match compounds. The authors classified each atom as hydrogen-bonding donor, hydrogen-bonding acceptor, hydrophobic or charged. Then, the superposition force field applied an attractive force (external force) between atoms of two molecules when these atoms are of similar type. TFIT was successful in flexibly matching thermolysin inhibitors, HIV-1 protease inhibitors, and endothiapepsin inhibitors to template molecules [8]. However, TFIT in general needs a common molecular functional group for a preliminary alignment of molecules. The authors also applied TFIT to a flexible search for the superposition of four ACE inhibitors and found low-energy structures similar to previously proposed solutions [9]. However, common molecular functional groups of ACE inhibitors were connected with zero-order bonds.

There are two ways to search the possible superposing orientations. One is the flexible superposition that TFIT adopted, and the other is the rigid superposition of various conformations of molecules. The flexible superposition is rapid, but it might miss the best solutions under the influence of the applied external force. On the other hand, the rigid superposition of various conformations of molecules does not miss the best solutions, although the searches take much time.

We aimed at elucidating a pharmacophore and estimating each active conformation from among many stable conformations of each molecule, and developed the program SUPERPOSE to rigidly superpose the physicochemical property spheres of the functional atoms in molecules. SUPERPOSE is suitable for superposing many conformers by a parallel computer because each and every superposition of conformers is independent. Therefore, SUPERPOSE is practical for executing many superpositions. When considering orientations which included the common conformers of a reference molecule, SUPERPOSE can select several conformer sets, which are possible candidates for an active conformer set. In the following section, we describe the detailed features of SUPERPOSE. Some examples are also given.

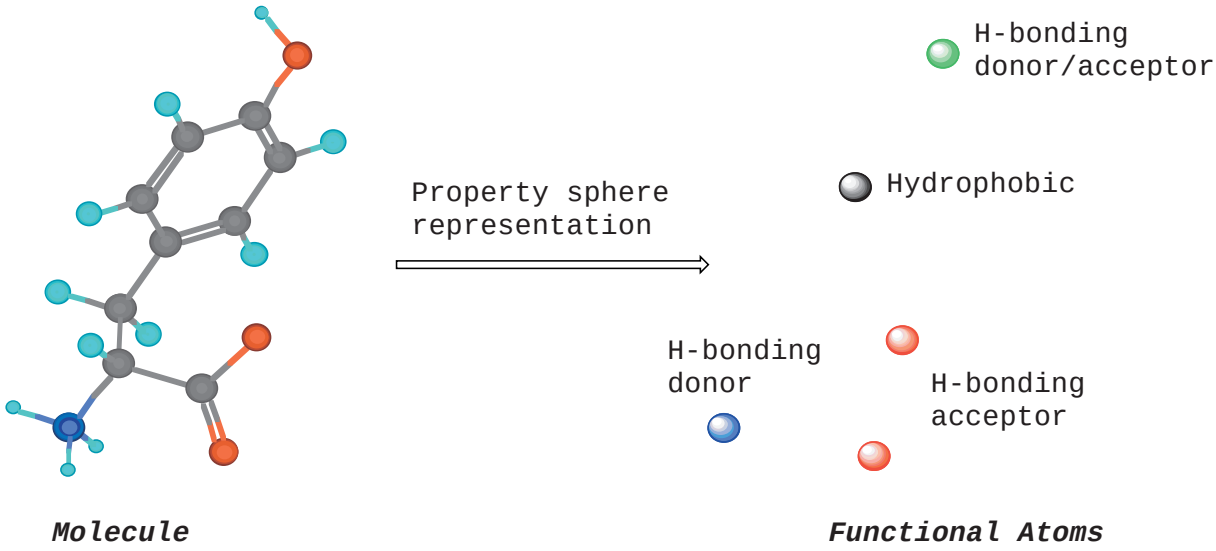

Figure 1. Property sphere representation of a tyrosine molecule.

Methods

Superposition of two conformations

This section describes the program SUPERPOSE, which rigidly superposes the physicochemical property spheres of functional atoms in molecules. For example, a tyrosine molecule is represented by property spheres of five functional atoms, i.e., a hydrophobic atom, a hydrogen-bonding donor atom, two hydrogen-bonding acceptor atoms and a hydrogen-bonding donor/acceptor atom (Figure 1). Therefore, we considered a pseudo-molecule consisting of the five property spheres instead of a tyrosine molecule.

SUPERPOSE needs the three-dimensional coordinates of molecules in the protein data bank (PDB, Brookhaven Protein Data Bank) file format and the list of functional atoms and property radii in plain text format. The superpositions of enzyme inhibitors were performed using the three-dimensional coordinates obtained from the PDB without additional hydrogens. The orientation obtained from X-ray crystallography is derived by removing only the coordinates of the enzyme molecule after a least-squares fitting between the α-carbon atomic coordinates of the enzyme molecules in the enzyme–inhibitor complexes.

Algorithm for SUPERPOSE

(1) The molecule with the large box-volume is fixed and the center of mass for another molecule is translated and rotated.

(2) When functional atoms between two molecules overlap by the operation of translation and rotation in (1), a score is given depending on the overlap between the respective functional atoms.

(3) Every score is summed for all the overlaps of functional atoms between two molecules.

Operations (1)–(3) are repeated to determine the orientation with the highest score.

Using the superposing method, there is no need of: (a) a consideration of individual functional atoms between two molecules; (b) an operation for advance determination of the pharmacophore; or (c) an accurate correspondence between atomic positions of the two superposed molecules.

The process of SUPERPOSE is shown in Figure 2. Each step will now be explained. Four types of physicochemical properties – hydrophobicity, presence of a hydrogen-bonding donor, presence of a hydrogen-bonding acceptor and presence of a hydrogen-bonding donor/acceptor – were supposed and a score was given to every overlap thereof. Also, when functional atoms with the same physicochemical properties were overlapped, points were added to the score, and when functional atoms with different physicochemical properties were overlapped, points were subtracted. Atomic groups having no overlap were not scored (see Table 1). Overlaps of more than one pair with the same combination of physicochemical properties were not allowed with regard to a functional atom.

The scores in Table 1 and the radii of the property spheres were determined to reproduce the true super-

Table 1. Scores between atomic groups

Atomic groups	Hydrophobic	H-donor	H-acceptor	H-donor/ acceptor
Hydrophobic	+3	−2	−2	−2
H-donor	−2	+2	−2	+1
H-acceptor	−2	−2	+2	+1
H-donor/acceptor	−2	+1	+1	+1

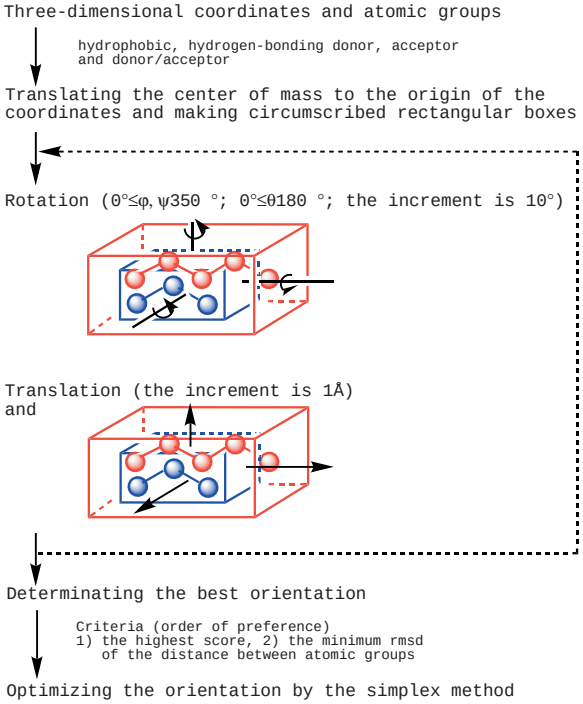

Figure 2. Superposing process of SUPERPOSE.

positions obtained from the X-ray crystallography of the HIV-1 protease inhibitors A-74704 and MVT-101, and the thrombin inhibitors NAPAP and MQPA. Each functional atom was approximated by a sphere and if any two spheres overlapped by even a little, they were treated as a target of a score.

In the superposition between HIV-1 protease inhibitors, property spheres of radii 0.5, 0.8, 1.0, and 1.3 Å were examined using the scores in Table 1. Here, we put a hydrophobic sphere at the center of each phenyl substituent. As a result, the superposition using property spheres of radius 1.0 Å could reproduce the X-ray orientation. When we used 0.5 and 0.8 Å as the radii of the property spheres, the orientations were different from the X-ray orientation. Therefore, we chose 1.0 Å as the radius of the property spheres.

With regard to score, the value between functional atoms with the same physicochemical property was changed from 2 to 4, except in the case of the hydrogen-bonding donor/acceptor, where it was changed from 1 to 2. Moreover, the value between functional atoms with different physicochemical properties was changed from −2 to −4. When using property spheres of radius 1.0 Å, the superpositions using the scores in Table 1 could correctly reproduce the X-ray orientations of HIV-1 protease inhibitors and thrombin inhibitors.

Each functional atom was defined as follows:

(1) The oxygen atom of a carbonyl, sulfone, phosphone, ester, ether, etc. is a hydrogen-bonding acceptor.

(a) In the resonance systems of CO_2^-, SO_2, PO_2^-, etc., the radius of the hydrogen-bonding acceptor sphere is 0.5 Å.

(b) In (a), it is also possible to place a 1 Å hydrogen-bonding acceptor sphere in the middle of two oxygen atoms.

(2) The oxygen atom of a hydroxyl group and the sulfur atom of a thiol are a hydrogen-bonding donor/acceptor, and the radius is 1 Å.

(3) The nitrogen atom accompanying a hydrogen atom, such as an amine, amide, amidine or guanidine, is a hydrogen-bonding donor.

(a) In the resonance systems of an amidine, etc., the radius of the hydrogen-bonding donor sphere is 0.5 Å.

(b) In (a), it is also possible to place a 1 Å hydrogen-bonding donor sphere in the middle of two nitrogen atoms.

(4) The aromatic ring of a phenyl, naphthalene, pyridine, thiophene, etc. is hydrophobic.

(a) 0.5 Å hydrophobic spheres are placed at the root of the aromatic ring and at every other atom in the clockwise and counterclockwise directions.

(b) In (a), it is also possible to place a 1 Å hydrophobic sphere at the center of the ring.

(5) An aliphatic chain (including branched chain) with an alkyl (including thioether) chain of three atoms or more is hydrophobic.

(a) A 1 Å hydrophobic sphere is placed on a carbon atom at a branched position.

(b) A 1 Å hydrophobic sphere is placed on a carbon atom at the third position counting from the root of the main chain. If the atom at the root is already considered to be a functional atom, counting begins from the next atom.

(c) It is impossible to continuously place the hydrophobic sphere next to the functional atom having polarity.

(d) It is impossible to continuously place the hydrophobic spheres (two atoms should be interposed).

(e) It is possible to place a 1 Å hydrophobic sphere at the center of the piperidine ring and of pyrrolidine ring.

(6) The trifluoromethyl group is hydrophobic and a 1 Å hydrophobic sphere is placed on the central carbon atom.

In order to systematically perform the superposition of two molecules, first the center of mass of each molecule is translated to the origin of coordinates, and then the circumscribed rectangular boxes are calculated. The molecule with a large box-volume is fixed, and then the center of mass for the molecule with a small box-volume is translated and rotated. The range of translation is the maximum distance that the small box can translate inside the large box. The translational increment is 1 Å and the center of mass is translated on the body-centered cubic lattice points made in the circumscribed rectangular box of the large volume. The rotation is performed on each of the lattice points. The ranges of three Eulerian angles are $0 \leq \varphi, \Psi \leq 350°$ and $0 \leq \theta \leq 180°$ and the rotational increment is 10°.

The score is calculated on the orientations of all superpositions, respectively, and the orientation with the highest value is adopted. If the highest value of the score is redundant, the one with the smallest value of the root-mean-square deviation (rmsd) of the distance between every pair of functional atoms is conveniently selected. As a precaution, however, it might be required to check orientations with the same score. For further improving the score of the adopted orientation, three translations and three Eulerian angles are optimized by the simplex method using the rmsd as an objective function to determine the final orientation of superposition. In the calculation of the rmsd, however, overlaps of more than one pair with the same combination of physicochemical properties are allowed with regard to a functional atom.

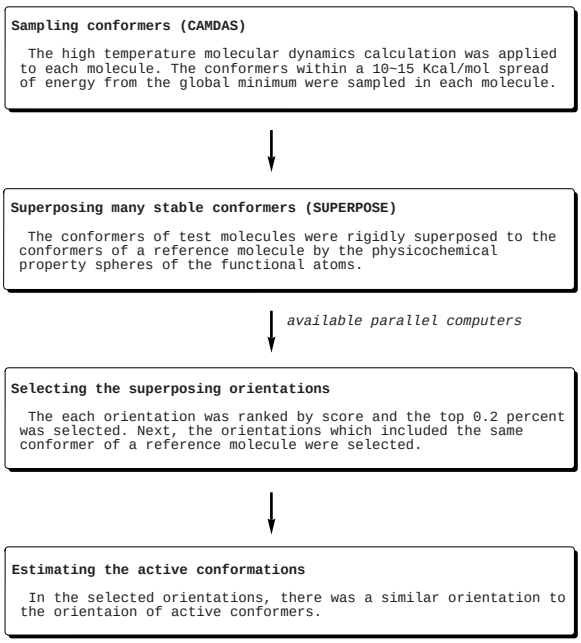

Figure 3. Scheme for estimating the active conformation.

Superposition of many conformations

It has been reported that the conformation of a physiologically active peptide in aqueous solution is nearly identical to the conformation binding to a receptor or enzyme [10]. However, calculations involving water molecules require a considerable amount of time. Therefore, for purposes of simplicity and rapidity in computer-assisted drug design, there is need of a practically applicable molecular mechanics calculation that does not deal with solvent molecules. In vacuum, the intramolecular interaction is estimated in excess because there is no or only weak interaction with the outside. The conformation in vacuum tends to become unnaturally compact in comparison with the conformation in solution [11]. For this reason, we used a potential function that excluded the electrostatic interaction term and hydrogen-bonding term, and performed conformational analysis in vacuum by the high-temperature MD calculation, one of the best ways to generate the energetically possible conformers of a molecule.

We used an automated conformational analysis program, CAMDAS (Conformational Analyzer with

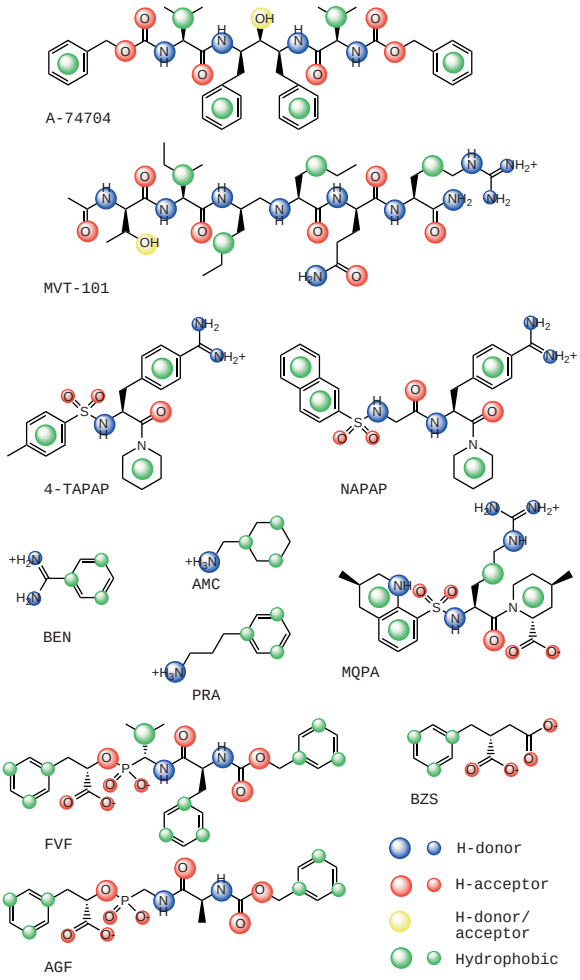

Figure 4. Chemical structures and functional atoms of the inhibitors. The large circles represent 1 Å and the small circles represent 0.5 Å with respect to the property radii of the functional atoms.

Molecular Dynamics And Sampling) [12]. CAMDAS performs MD calculations for a molecule and samples conformers from the trajectory of the MD. The program then minimizes conformers, evaluates the similarities between each of the sampled conformers in terms of rmsd of the atomic positions, clusters similar conformers, and finally prints out the clustered conformers. CAMDAS is intended to provide a convenient framework for the method and has the ability to find the representative conformers automatically from an arbitrarily given structure of the molecule. The accuracy of CAMDAS was examined using *N*-acetylalanine-*N*′-methylamide, and the obtained result was consistent with that of the systematic search method.

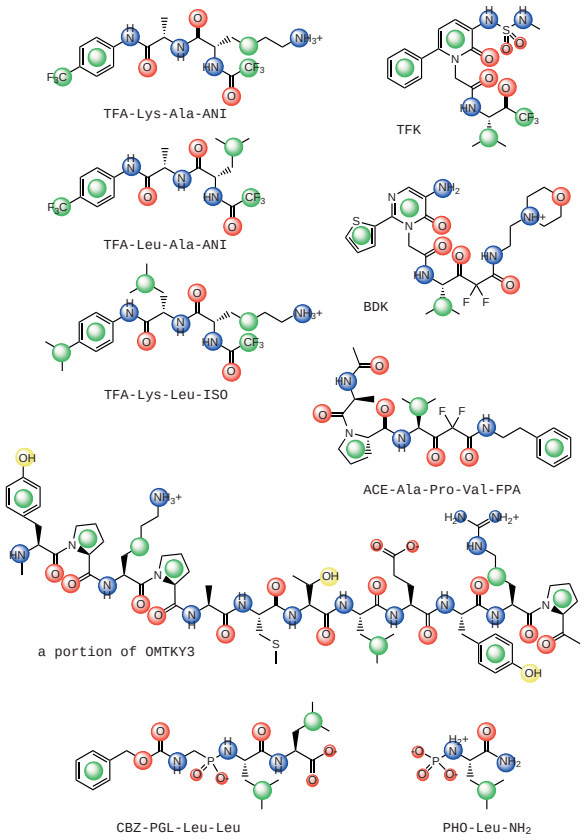

Figure 4. (continued).

For the conformations of compounds binding to the same receptor or enzyme, the same functional atoms occupy the same three-dimensional space. Hence, a lenient superposition of functional atoms between two molecules seems to be effective for the extraction of an active conformation. SUPERPOSE was applied to estimate the active conformation among many stable conformations. We carried out the scheme (Figure 3) for estimating the active conformations.

Results and discussion

Superposition of two conformations

We applied SUPERPOSE to 12 pairs of 24 enzyme inhibitors for comparison with the superpositions obtained from the X-ray crystallography of an enzyme–inhibitor complex. The PDB entry codes of the enzyme–inhibitor complexes and the names of the inhibitors are shown in Table 2. Their chemical structures and the functional atoms are shown in Figure 4.

Table 2. PDB entry codes of the enzyme–inhibitor complexes and the names of inhibitors

Enzyme name	PDB entry codes and inhibitor names[a]
1. HIV-1 protease	9HVP(A-74704), 4HVP(MVT-101)
2. Thrombin	1ETR(MQPA), 1ETT(4-TAPAP)
3. Thrombin	1ETS(NAPAP), 1ETR(MQPA)
4. Trypsin	1TNK(PRA), 1TNG(AMC)
5. Trypsin	1TNK(PRA), 2TBS(BEN)
6. Carboxypeptidase-A	7CPA(FVF), 1CBX(BZS)
7. Carboxypeptidase-A	7CPA(FVF), 8CPA(AGF)
8. Elastase	1EAU(BDK), 1EAS(TFK)
9. Elastase	2EST(TFA-Lys-Ala-ANI), 7EST(TFA-Leu-Ala-ANI)
10. Elastase	1EAS(TFK), 1ELB(TFA-Lys-Leu-ISO)
11. Elastase	1PPF(part of OMTKY3), 4EST(ACE-Ala-Pro-Val-FPA)
12. Thermolysin	5TMN(CBZ-PGL-Leu-Leu), 2TMN(PHO-Leu-NH_2)

[a]Abbreviations of the inhibitor names are given in parentheses.

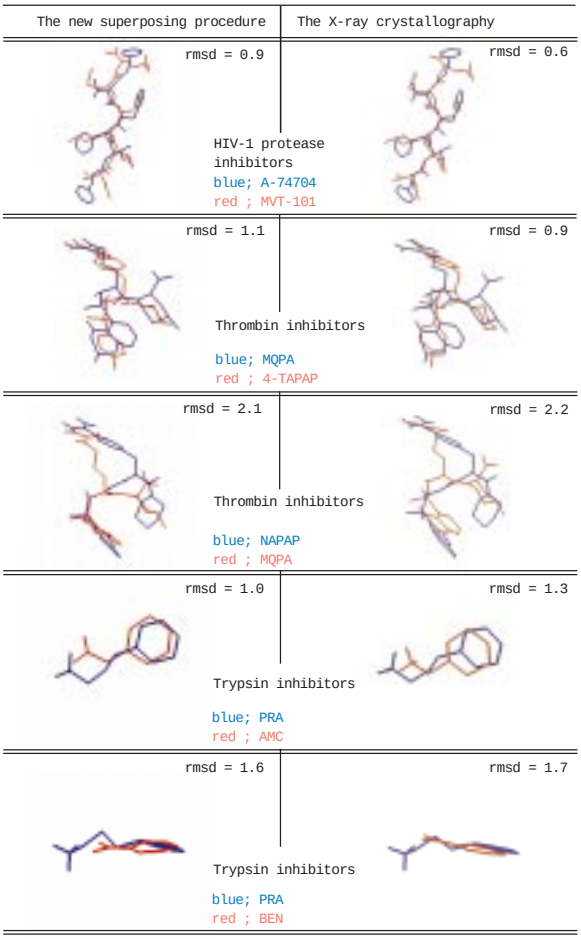

Figure 5. Comparison of the best orientations with the crystal orientations. The rmsd values of several corresponding atoms between two inhibitors are shown.

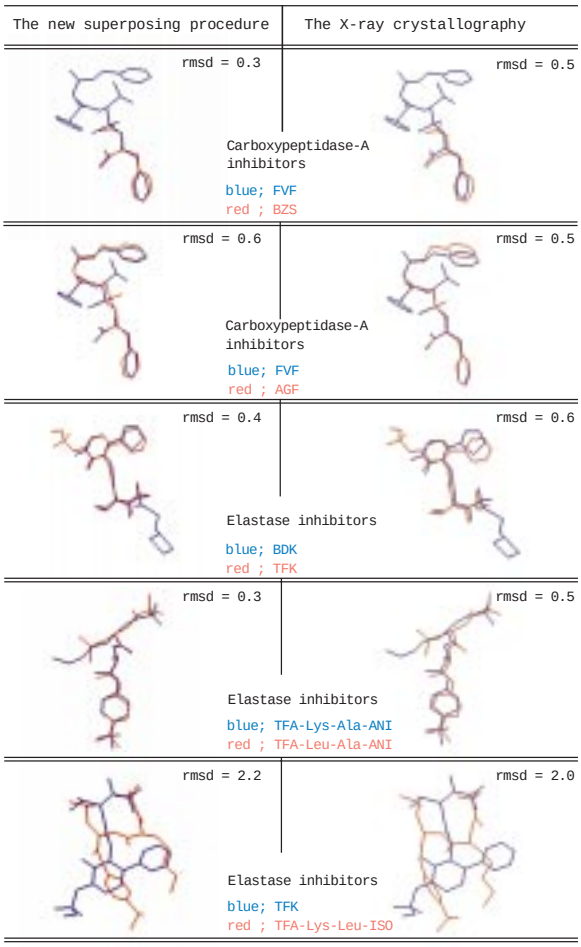

Figure 5. (continued).

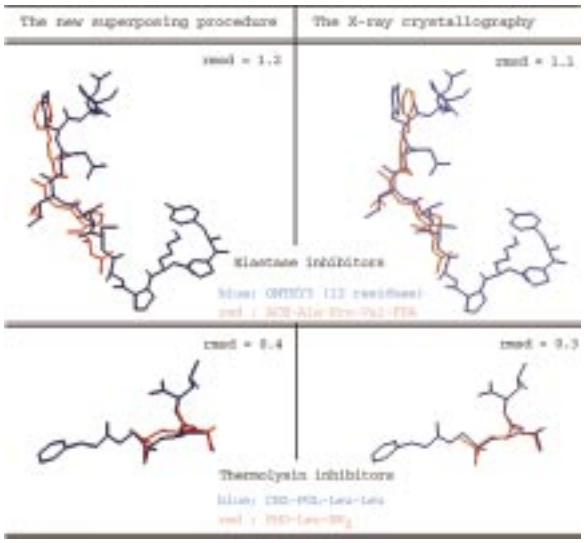

Figure 5. (continued).

The results of the superpositions between two active conformations are shown in Figure 5. The left side of each diagram shows the orientation obtained by SUPERPOSE and the right side shows that obtained by X-ray crystallography. All the orientations on the 12 pairs of 24 enzyme inhibitors reproduced the orientations from X-ray crystallography with much the same rmsd values. In the superposition between the HIV-1 protease inhibitors A-74704 and MVT-101, we put a hydrophobic sphere at the center of each phenyl substituent. As a result, the orientation with the highest score (26) was the same as that from X-ray crystallography. The next highest score was 25, and this orientation was very similar to the highest score orientation. The orientation with the highest score was optimized by the simplex method, but neither the score nor the rmsd were improved.

In the case of the thrombin inhibitors, two superpositions, one between NAPAP and MQPA and one between MQPA and 4-TAPAP, seemed a bit unusual for their lack of coincidence with the main-chain atoms. We used a hydrophobic sphere to represent the phenyl moiety. In the superposition between MQPA and 4-TAPAP, the highest score of the top two orientations was 19, and both had an rmsd of 0.9 Å. These top two orientations were basically the same, except for a difference of the Eulerian angles, φ,Ψ, which were 10°, 340° in one case and 20°, 330° in the other. The former orientation was optimized by the simplex method, but neither the score nor the rmsd were improved. In the superposition between NAPAP and MQPA, the top 10 scoring orientations had a score of 19 and an rmsd of 1.0 Å. These 10 orientations were basically identical, differing only with respect to the Eulerian angles, φ,Ψ,θ. After one among the 10 orientations was optimized by the simplex method, the score was 19 and the rmsd was 0.9 Å.

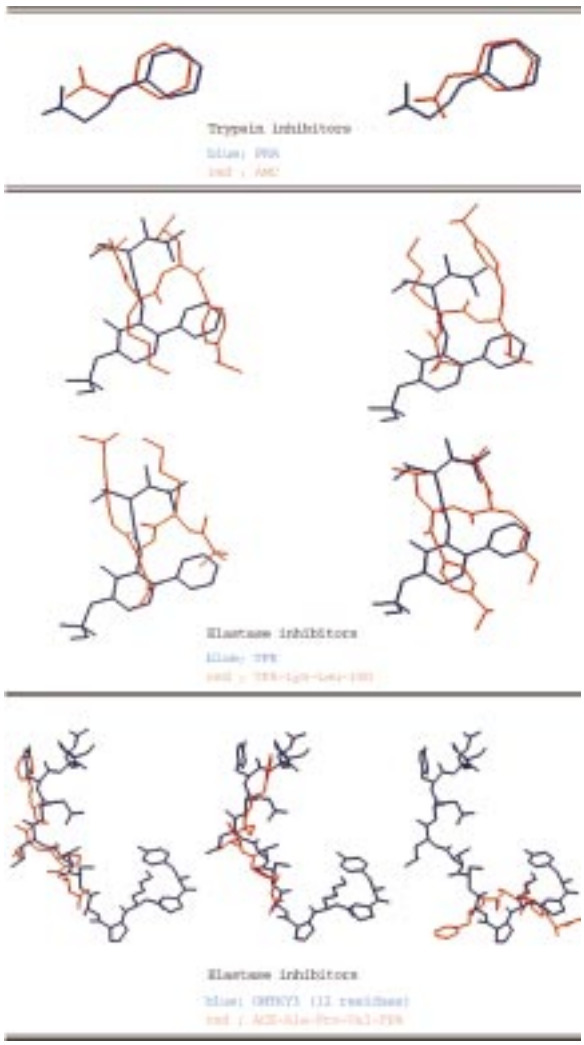

Figure 6. The several kinds of orientations obtained from SUPERPOSE.

In the case of the trypsin inhibitors, superpositions between PRA and AMC and between PRA and BEN were carried out. These compounds possess only two physicochemical properties – hydrophobicity and the presence of a hydrogen-bonding donor – so we used three hydrophobic spheres instead of a single one to represent the phenyl or cyclohexyl substituents. In the superposition between PRA and AMC, the top seven orientations all had a score of 11 and an rmsd

of 1.0 Å. These orientations were classified into two groups with respect to a turnover of the phenyl ring of AMC, as shown in Figure 6. In the superposition between PRA and BEN, the top seven orientations all had a score of 11 and an rmsd of 1.0 Å. However, these seven orientations were all identical due to the symmetrical structure of BEN. In both cases, orientations reproduced the ring orientations of the X-ray crystallography of the trypsin inhibitors.

In the carboxypeptidase-A inhibitors, superpositions were carried out between FVF and BZS and between FVF and AGF. In the superposition between FVF and BZS, each phenyl substituent of FVF was positioned around two acidic moieties, PO_2^- and CO_2^-. Therefore, we used three hydrophobic spheres to represent each phenyl substituent. The highest score of the top four orientations was 17, and the rmsd was 0.5 Å in both cases.

These four orientations were basically the same, except for a difference of the Eulerian angles, φ,Ψ. After one among the four orientations was optimized by the simplex method, the score was 17 and the rmsd was 0.3 Å. In the superposition between FVF and AGF, the highest score of the top two orientations was 34, but the rmsd was 0.8 in one case and 1.0 Å in the other. The two orientations differed with respect to three translations. We selected the orientation with the smallest rmsd. After the orientation was optimized by the simplex method, the score was 36 and the rmsd was 0.4 Å. We next executed four superpositions between the elastase inhibitors, i.e., a superposition between BDK and TFK, between TFA-Lys-Ala-ANI and TFA-Leu-Ala-ANI, between TFK and TFA-Lys-Leu-ISO, and between ACE-Ala-Pro-Val-FPA and a portion of OMTKY3. We put a hydrophobic sphere at the center of each of the aromatic rings (i.e., phenyl, pyridine, pyrimidine, and thiophene) in each compound. As a result, every superposition reproduced in good detail the orientations from X-ray crystallography. In the superposition between BDK and TFK, the highest score of the top two orientations was 19 and the rmsd was 0.5 Å in both cases.

These two orientations were basically the same, except for a difference of the Eulerian angle, θ, which was 90° in one case and 100° in the other. After the former orientation was optimized by the simplex method, the score was 19 and the rmsd was 0.3 Å. In the superposition between TFA-Lys-Ala-ANI and TFA-Leu-Ala-ANI, the highest score of the top two orientations was 24, but the rmsd was 0.4 in one case and 0.7 Å in the other. The two orientations differed with respect to the Eulerian angles, φ,Ψ, which were 0°, 0° in one case and 0°, 10° in the other. The orientation with the smallest rmsd was optimized by the simplex method, but neither the score nor the rmsd were improved. In the superposition between TFK and TFA-Lys-Leu-ISO, the top eight orientations all had a score of 16 and an rmsd of 1.5 Å. These eight orientations were classified into four groups, as shown in Figure 6. The number of orientations in each group was 3, 2, 2, and 1. The three orientations in a group were the same as that from X-ray crystallography.

In these cases, we treated the trifluoromethyl substituent with a hydrophobic group, and the hydrophobic sphere was placed on the central carbon atom, but there seemed to be room to reassign the physicochemical property with respect to the trifluoromethyl substituent, which possessed a strong ability to withdraw electrons.

Perkins et al. developed a superposing technique that accounted for similarities in molecular surface-volume, hydrogen bonding and electrostatic potential, and applied this technique to a superposition between OMTKY3 and ACE-Ala-Pro-Val-FPA. They reported a successful reproduction of the X-ray orientation between OMTKY3 and ACE-Ala-Pro-Val-FPA, but this reproduction required a significant amount of time [7]. We also attempted a superposition between OMTKY3 and ACE-Ala-Pro-Val-FPA, but for reasons of practicality treated 12 rather than 56 residues of OMTKY3. We put a hydrophobic sphere at the center of their phenyl and pyrrolidine substituents in each compound, and were also successful in estimating the X-ray orientation between ACE-Ala-Pro-Val-FPA and a portion of OMTKY3. The top three orientations all had a score of 16 and an rmsd of 1.3 Å. Two orientations were much the same as that from X-ray crystallography and the remaining orientation was different, as shown in Figure 6.

In the case of the thermolysin inhibitors, superposition between CBZ-PGL-Leu-Leu and PHO-Leu-NH_2 was carried out. We put a hydrophobic sphere at the center of a phenyl substituent in CBZ-PGL-Leu-Leu. The highest score of the top two orientations was 13, but the rmsd was 0.5 in one case and 0.6 Å in the other. These orientations were identical, differing only with respect to the Eulerian angles, φ,Ψ, which were 140°, 70° in one case and 150°, 60° in the other. The orientation with the smallest rmsd was optimized by the simplex method, but neither the score nor the rmsd were improved.

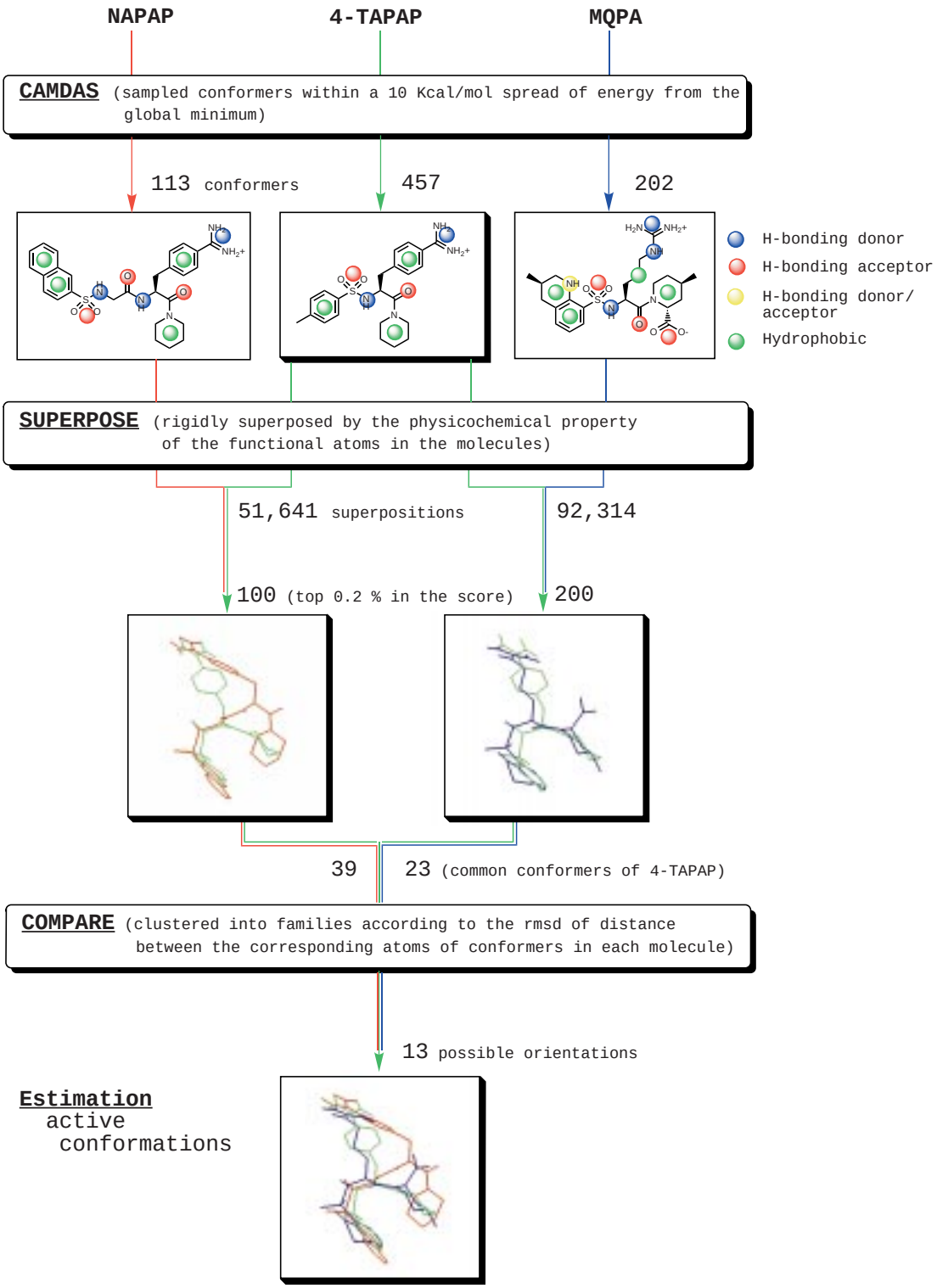

Figure 7. Process of estimating the active conformations of the thrombin inhibitors MQPA (blue), 4-TAPAP (green) and NAPAP (red).

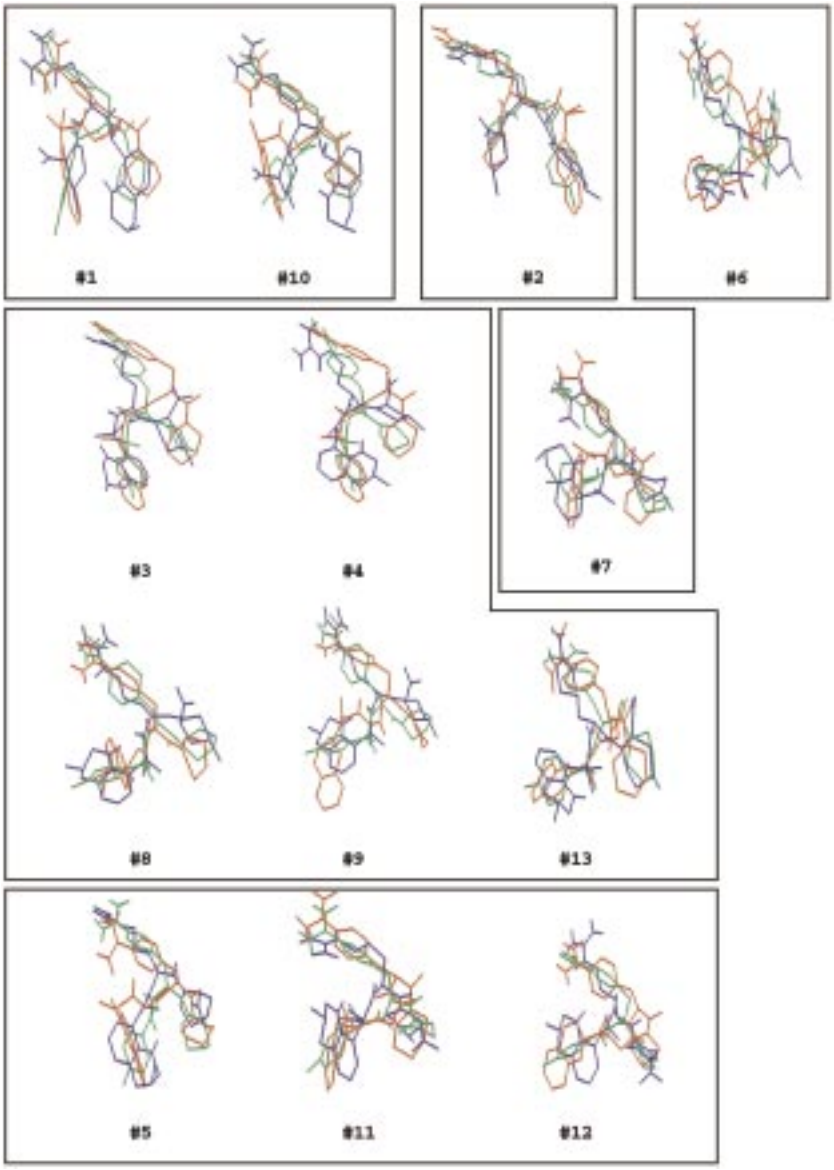

Figure 8. The 13 pairs of conformers common to MQPA (blue), 4-TAPAP (green) and NAPAP (red).

All 12 orientations with the highest score by SUPERPOSE were consistent with the orientations obtained from X-ray crystallography, indicating the efficacy of SUPERPOSE.

Superposition of many conformations

The process of estimating the active conformations of three thrombin inhibitors is shown in Figure 7. The high-temperature MD calculations for the thrombin inhibitors MQPA, 4-TAPAP and NAPAP were executed, and the sampled conformers were energetically minimized and clustered into families using the CAMDAS program. CAMDAS was executed under conditions of a 900 K NVT ensemble over 3 ns using the MM2 force field without electrostatic and hydrogen-bonding interactions. A total of 60 000 conformers, which were sampled every 50 fs during the MD calculations, were finally clustered into families according to the dihedral-angle values. When the difference between dihedral angles for each rotatable bond between conformers was within $\pm 30°$, the conformers were included in the same group. The conformers within a 10 kcal/mol spread of energy from the global min-

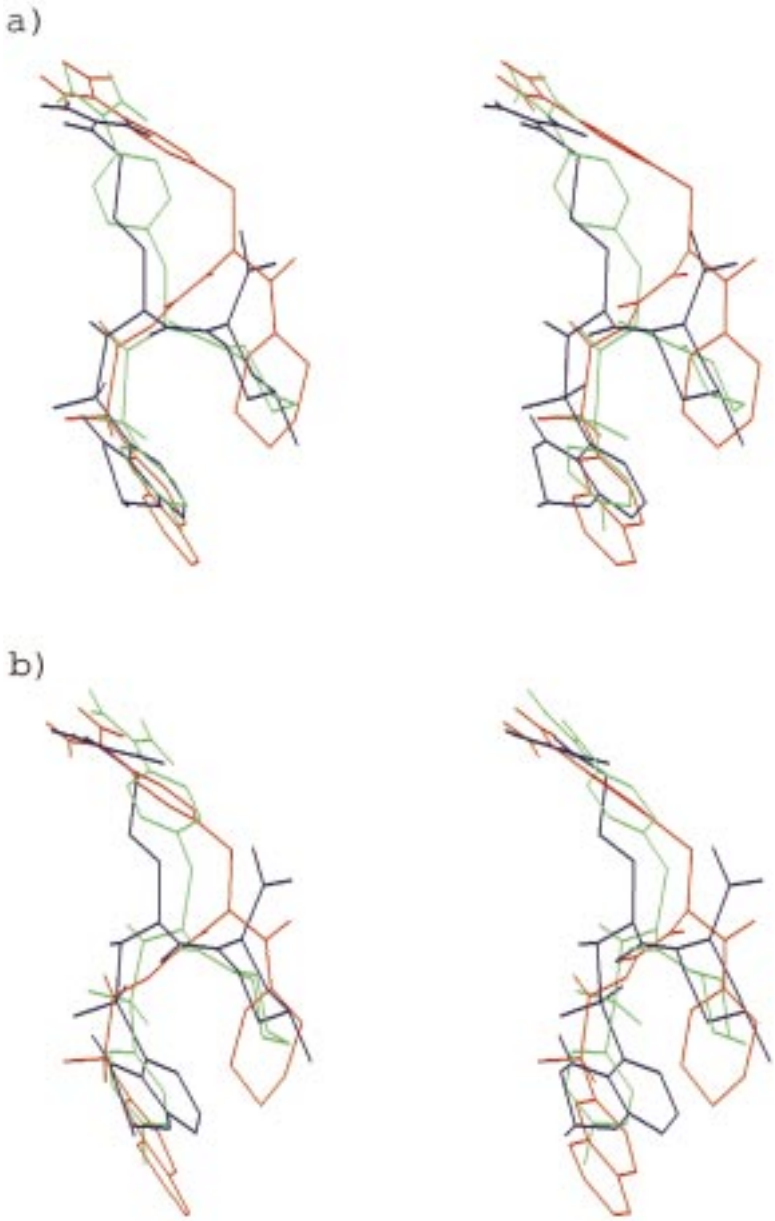

Figure 9. Stereo diagrams of (a) the best orientation obtained by superpositions of the thrombin inhibitors MQPA (blue), 4-TAPAP (green) and NAPAP (red), and (b) their crystal orientation.

imum were used as superposing conformers. As a result, 457 conformers in 4-TAPAP, 113 conformers in NAPAP and 202 conformers in MQPA were selected (see Table 3).

We considered that 4-TAPAP was similar to both NAPAP and MQPA in view of the chemical structure among the three inhibitors. We then applied 51 641 (457 multiplied by 113) superpositions between 4-TAPAP and NAPAP and 92 314 (457 multiplied by 202) superpositions between 4-TAPAP and MQPA. As already mentioned, it was slightly difficult to perform superpositions between thrombin inhibitors due to the lack of coincidence with the main-chain atoms. We represented a phenyl moiety as a hydrophobic sphere, and used a 1 Å sphere, as shown in Figure 7. We tentatively treated the nitrogen atom of the tetrahydroquinoline in MQPA as a hydrogen-bonding donor/acceptor instead of a hydrogen-bonding donor,

Table 3. CAMDAS conditions

Dynamics: −ES (no dipole and no H-bond)	
NVT ensemble	
Temperature	900 K
Simulation time	3 ns
Time step	1 fs
Shake	
vdW cut	20 Å
Clustering	Torsion angle (within ±30°)
Minimizer	Steepest descent ≤ 100 steps
	Conjugate gradient ≤ 5000 steps
Convergence	≤ 0.08 kcal/(mol Å)
Sampling	≤ 10 kcal/mol
Candidate conformers	4-TAPAP: 457
	NAPAP: 113
	MQPA: 202

but there was no difference between the orientations of 4-TAPAP and MQPA in a preliminary superposition. After all the superpositions were finished, we ranked the orientations by score and rmsd, and selected the top 0.2%. The number of orientations selected was 100 in 4-TAPAP and NAPAP, and 200 in 4-TAPAP and MQPA. In 4-TAPAP and NAPAP, the score was between 8 and 22 in all 51 641 orientations, and between 20 and 22 in the top 100 orientations. In 4-TAPAP and MQPA, the score was between 9 and 24 in all 92 314 orientations, and between 22 and 24 in the top 200 orientations.

The number of orientations was reduced to 39 between 4-TAPAP and NAPAP, and to 23 between 4-TAPAP and MQPA, when considering orientations which included the common conformers of 4-TAPAP. Subsequently, the conformers of each inhibitor were clustered into families according to the values of the rmsd of the distance between the corresponding atoms. When the rmsd values of the distance between conformers was found to be lower than 1.4 Å using the in-house program COMPARE, the conformers were classified as belonging to the same group. The representative conformer of each group was the lowest energy conformer. Finally, the conformer sets of three inhibitors were reduced to 13. These results are shown in Figure 8. The resulting conformer sets contained the superposition of the active conformations derived from the X-ray crystallography of the thrombin–inhibitor complexes, as shown in Figure 9.

Conclusions

We have developed SUPERPOSE based on the physicochemical property spheres of functional atoms in individual molecules. Our method does not require beforehand correspondence between functional atoms of each compound, and does not attempt a perfect coincidence of atomic positions during the superposing process.

First, in order to check the effectiveness of SUPERPOSE, we performed superpositions on 12 pairs of 24 enzyme inhibitors, with the result that the best scored overlay for each inhibitor pair successfully reproduced the true superposition obtained from X-ray crystallography.

Next, we performed superpositions for estimating the active conformation among many stable conformations. The high temperature MD calculations for the thrombin inhibitors MQPA, 4-TAPAP and NAPAP were executed, and 457 conformers in 4-TAPAP, 113 conformers in NAPAP, and 202 conformers in MQPA were collected. After 51 641 superpositions between 4-TAPAP and NAPAP and 92 314 superpositions between 4-TAPAP and MQPA were performed, we selected the 100 orientations between 4-TAPAP and NAPAP, and the 200 orientations between 4-TAPAP and MQPA. Moreover, the number of orientations was reduced to 39 between 4-TAPAP and NAPAP, and to 23 between 4-TAPAP and MQPA, when considering orientations which included the common conformers of 4-TAPAP. Finally, the conformers of each inhibitor were clustered into families according to the values of the rmsd of the distance between the corresponding atoms, and the conformer sets of three inhibitors were reduced to 13. The resulting conformer sets contained the superposition of the active conformations derived from the X-ray crystallography of the thrombin–inhibitor complexes.

We have found that when SUPERPOSE was applied to the five inhibitors of another enzyme, the resulting conformer sets were reduced to only a few. We are convinced that we would find an active conformer set using SUPERPOSE. Moreover SUPERPOSE is suitable for superposing many stable conformers by a parallel computer because each and every superposition of conformers is independent. Therefore, SUPERPOSE is practical for executing many superpositions. When we used 256 CPUs in the Hitachi SR2201 parallel computer, 92 314 superpositions between 4-TAPAP and MQPA took 20 h. When 1024 CPUs were used, the superpositions were complete in about 5 h. The

spread of parallel computers makes SUPERPOSE a useful tool for estimating the active conformation in the near future.

Acknowledgements

We thank Dr. Hideki Tsujisita for his help with the implementation of the CAMDAS program.

References

1. Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, A, J. Comput.-Aided Mol. Design, 7 (1993) 83.
2. Carbo, R., Leyda, L. and Arnau, M., Int. J. Quant. Chem., 17 (1980) 1185.
3. Carbo, R. and Domingo, L., Int. J. Quant. Chem., 32 (1987) 517.
4. Richards, W.G. and Hodgkin, E.E., Chem. Br., 24 (1988) 1141.
5. Hodgkin, E.E. and Richards, W.G., Int. J. Quant. Chem., 14 (1987) 105.
6. Burt, C., Richards, W.G. and Huxley, P., J. Comput. Chem., 11 (1990) 1139.
7. Perkins, T.D.J., Mills, J.E.J. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 479.
8. McMartin, C. and Bohacek, R.S., J. Comput.-Aided Mol. Design, 9 (1995) 237.
9. McMartin, C. and Bohacek, R.S., J. Comput.-Aided Mol. Design, 11 (1997) 333.
10. Warshel, A., Åqvist, J. and Creighton, S., Proc. Natl. Acad. Sci. USA, 86 (1989) 5820.
11. Levitt, M. and Sharon, R., Proc. Natl. Acad. Sci. USA, 85 (1988) 7557.
12. Tsujishita, H. and Hirono, S., J. Comput.-Aided Mol. Design, 11 (1997) 305.