Flexible Alignment of Small Molecules Using the Penalty Method

Whanchul Shin,*,† Seung Ah Hyun,† Chong Hak Chae,‡ and Jae Kyung Chon†,‡

Department of Chemistry and Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea, and Korea Research Institute of Chemical Technology, Yuseong-Gu, Daejeon 305-600, Korea

Received December 11, 2008

An efficient flexible alignment method using the penalty method, called FAP, is described. FAP is a pairwise alignment algorithm that matches a flexible sample to a rigid template. It is a pure atom-based 3D method that utilizes the modified SEAL similarity index combined with an energy penalty term. The penalty term, defined as the third power of the ratio of the local strain energy to its target value, enables effective control of energy increase during alignment. The alignment procedure consists of the seed conformer generation, rigid-body alignment, and flexible optimization steps. Both conformation and alignment spaces are efficiently explored by the sparse, random sampling schemes. FAP has been tested with benchmark sets of seven different classes of ligands taken from the literature. In terms of the ability to produce the bioactive overlays, FAP is comparable to, or in some cases better than, other alignment methods. FAP is accurate, objective, fully automated, and fast enough to be used as a tool for virtual screening.

INTRODUCTION

The alignment of 3D structures of small molecules remains an important step in various drug design strategies such as derivation of the pharmacophore model, receptor mapping, 3D QSAR analysis, and ligand-based virtual screening.[1−4] Numerous methods for molecular superposition have been developed, as extensively reviewed by Lemmen and Lengauer,[2] and new procedures continue to appear.[5−8] They differ in the definition of molecular similarity and in the alignment methods. The similarity index (or alignment score), a means of quantifying the degree of overlap of molecular properties, can be defined in various ways.[9−11] The different alignment methods widely differ in the search and optimization algorithms.[2,12]

Flexible ligands undergo structural rearrangements upon binding to the protein, and their bioactive conformers are likely to differ from low-energy conformers.[13−16] Perola and Charifson have extensively analyzed the energetics of ligands obtained from 150 X-ray structures of protein−ligand complexes.[14] They found that the strain energies calculated for the bound ligands with respect to the closest local minimum ranged from 0.0 to 16.5 kcal/mol with an average value of 1.9 kcal/mol, and the strain energies calculated relative to the global minima ranged from 0.0 to 26.2 kcal/mol with an average value of 4.0 kcal/mol, when the calculations were performed with MMFF force field.[17] They have also found that the thresholds for acceptable strain energies increase with increasing ligand flexibility. These results indicate that flexibility must be taken into account to obtain bioactive overlays.

The alignment methods are usually classified into rigid, semiflexible, or flexible methods depending on how molecular flexibility is treated.[2,12] In the rigid methods, a conformer of the sample molecule is aligned with the template as a rigid body. Obviously, this method is not appropriate for the flexible ligand molecules unless the bioactive conformer is not used. The semiflexible methods utilize a precomputed ensemble of conformers for each compound and rigidly align each conformer with the template in turn. The conformer generators, such as Catalyst, MOE, and Omega, are designed to produce representative conformations that cover conformational space and are capable of generating bioactive conformers.[18−20] The semiflexible methods are usually very fast and have their limitations in terms of sampling conformational space. In contrast, the flexible methods usually handle conformational flexibility on-the-fly. Some of the flexible methods can perform the simultaneous alignment of multiple ligands in addition to the pairwise alignment, and many of them allow both sample and template molecules to be flexible.[8]

Flexible alignment is a difficult multidimensional optimization problem, especially for molecules with many rotatable bonds. Numerous flexible methods that widely differ in similarity criteria, superposition methodology, and optimization algorithm have been developed.[21−33] Among these, TFIT,[23] GASP,[25] Labute's method,[30] and FLAME[33] share a common feature; they simultaneously optimize the alignment score and the internal energy of the sample. These approaches inherently entail a multiobjective optimization problem for which, in principle, no definite solution exists.[34] Indeed, the alignment score must be maximized, while the energy has to remain as low as possible. The relative weights for the alignment score and the energy as well as for many terms in the alignment score itself significantly affect the optimization behavior. Although effective procedures have been devised and validated for these methods, the determination of proper weights remains problematic. In these flexible methods, the accurate control of the increase in energy during alignment is not guaranteed.

* Corresponding author e-mail: nswcshin@plaza.snu.ac.kr.
† Seoul National University.
‡ Korea Research Institute of Chemical Technology.

10.1021/ci8004463 CCC: $40.75
Published on Web 07/31/2009

We have previously reported on a flexible method based on the modified SEAL index.[31] The original SEAL index is defined as the sum of pairwise overlaps of steric volumes and partial atomic charges that are attenuated by a decaying Gaussian function.[35] This is simple and robust but still entails the problem of scaling of the two terms. Our modified index evaluates the steric overlap weighted by electrostatic similarity using transferable parameters.[31] Thus, the scaling problem is apparently eliminated. Our method is unique in that it employs the constrained optimization technique to circumvent the multiobjective nature of the problem.[31,34] The constrained optimizer maximizes the similarity by changing the conformation while allowing the strain energy to increase exactly by the amount specified by the user. Although this approach is in principle well suited to flexible alignments, the demanding computing need prohibits routine use.

A constrained minimization may be converted into an unconstrained one, for example, using classical methods such as the penalty method or the Lagrange multiplier method.[36] For those problems in which the exact fulfillment of constraints is not required, the penalty method is advantageous as it is easy to use. Indeed, flexible alignment is such a problem. In this article, we present a new pairwise alignment method that matches a flexible sample to a rigid template. This algorithm efficiently treats molecular flexibility by using the penalty method. We devised a simple energy penalty function and combined it with a modified SEAL index to obtain an objective function for flexible alignment. This function can be optimized with an efficient, unconstrained quasi-Newton method, and it enables effective control of energy increase during alignment. We have also developed an efficient search protocol in which the global minimum can be found even if the alignment space is very sparsely sampled. Our fully automated method, called FAP (Flexible Alignment using the Penalty method), is much faster than the previous version using the constrained optimizer, and its use in virtual screening becomes feasible. To demonstrate the accuracy and effectiveness of this method, we have tested many alignment problems that were previously studied with other alignment methods.

METHODS

Objective Function for Flexible Alignment. During the alignment of two molecules, one molecule is usually fixed, and the other is allowed to move. In this study, they are termed *template* and *sample*, respectively. The SEAL alignment score in its original form is the sum of steric and electrostatic similarity terms that are attenuated by a decaying Gaussian function[35]

$$S = \sum_{i\in A}^{N_A}\sum_{i\in B}^{N_B}(w_v v_i v_j + w_q q_i q_j)\exp(-\alpha r_{ij}^2) \qquad (1)$$

In this function, q_i and v_i are the partial atomic charge and the steric volume of the atom i, respectively, and r_{ij} is the distance between the atom i in molecule A with N_A atoms and the atom j in molecule B with N_B atoms. w_v and w_q are the relative weighting factors, and α is an adjustable attenuation factor. This index is simple in that it requires only atomic volumes and partial atomic charges as the necessary data. It is, however, still robust. The addition of other terms such as hydrogen bonding capability, hydrophobicity, and/or refractivity does not significantly improve the quality of the aligment.[37]

Our objective function for flexible alignment, F_{SE}, is defined as

$$F_{\mathrm{SE}} = F_{\mathrm{S}} - F_{\mathrm{E}} = \sum_{i\in A}^{N_A}\sum_{i\in B}^{N_B} f(\Delta q)\cdot f(q_i q_j)\cdot v_i v_j \cdot \exp(-\alpha r_{ij}^2) - \left(\frac{\Delta E_{\mathrm{strain}}}{\Delta E_{\mathrm{target}}}\right)^3 = \sum_{i\in A}^{N_A}\sum_{i\in B}^{N_B}\left(\frac{2-|q_i-q_j|}{2}\right)\cdot\left(\frac{1+\tanh\{\beta(q_i q_j-\gamma)\}}{2}\right)\cdot v_i v_j \cdot \exp(-\alpha r_{ij}^2) - \left(\frac{E_{\mathrm{conf}}-E_{\mathrm{min}}}{\Delta E_{\mathrm{target}}}\right)^3 \qquad (2)$$

where F_{S} is the modified SEAL index, and F_{E} is the newly devised energy penalty function. The alignment score F_{S} is quantified by the sum of steric similarity weighted by electronic similarity. The electronic weighting factor, expressed as a product of the functions of Δq and $q_i q_j$, is designed to facilitate the matching of cognate charges; $f(\Delta q)$ reciprocally depends on the magnitude of the charge difference for an atom pair, and $f(q_i q_j)$ increases in a sigmoid manner as the magnitude of the charge product increases. The default value for α is 0.3, and the default values for β and γ that determine the shape of the hyperbolic tangential curve are 15.0 and 0.04, respectively. These parameters have been shown to be transferable to various samples, and they virtually eliminate the problem of scaling the steric and charge terms. The detailed properties of this index have been discussed in our previous article.[31]

In the penalty function F_E, E_{min} is the initial energy of the seed conformer which corresponds to one of local minima of the sample, and E_{conf} is the energy of the sample during flexible alignment. E_{conf} usually increases during alignment. Thus, $\Delta E_{\mathrm{strain}}$ is the amount of energy increased during alignment and may be called local strain energy. $\Delta E_{\mathrm{target}}$ is the user-specified target value for this strain energy and functions as an inexact constraint. The ratio of $\Delta E_{\mathrm{strain}}$ to $\Delta E_{\mathrm{target}}$ is raised deliberately to the third power. E_{conf} replaces E_{min} if the energy decreases during alignment.

The similarity index F_{S} is a function of conformational parameters as well as translational and rotational parameters, whereas F_{E} is a function of only conformational parameters. Both F_{S} and F_{E} terms are differentiable with respect to these parameters, and F_{SE} can be maximized (or $-F_{\mathrm{SE}}$ minimized) using the efficient Broyden-Fletcher-Goldfarb-Shanno (BFGS) method.[36] If the structure of the sample becomes strained during alignment and thus E_{conf} becomes larger than E_{min}, $\Delta E_{\mathrm{strain}}$ exceeds $\Delta E_{\mathrm{target}}$ and the $\Delta E_{\mathrm{strain}}/\Delta E_{\mathrm{target}}$ ratio becomes larger than 1. Then the penalty function F_{E}, defined as the cube of this ratio, increases very rapidly and becomes detrimental to the maximization of F_{SE}. Thus, during the optimization process, $\Delta E_{\mathrm{strain}}$ cannot become much larger than $\Delta E_{\mathrm{target}}$, and the penalty term effectively prevents the sample from adopting a highly strained conformation. In all reasonably aligned poses obtained in this study, the $\Delta E_{\mathrm{strain}}/\Delta E_{\mathrm{target}}$ ratio is mostly smaller than 2, and thus F_{E} is smaller than 8. The magnitude of F_{S} varies depending on the sizes of both sample and template, but it is always at least 2 orders of

magnitude larger than F_E. It seems that the variation in F_E does not significantly affect the overall optimization behavior.

Search and Optimization. The FAP method does not utilize any prior knowledge on the conformation or predefined relationship of particular groups but requires only the initial 3D atomic coordinates and partial atomic charges as necessary data. 3D structures can be retrieved from a structural database or made using a modeling program. When the molecule is used as a template, its coordinates are usually used without modification. If the molecule is used as a sample to be aligned, its conformation is treated as unknown, that is, bond lengths and bond angles are retained but torsion angles are completely ignored. The overall procedure is simple and straightforward and consists of the three sequential steps of seed conformer generation, rigid-body alignment, and flexible optimization.

Generation of Seed Conformers. To explore the alignment space for a flexible molecule, its conformational space needed to be sampled somehow. In the FAP method, we neither sought the global minimum of the sample molecule nor performed an exhaustive conformational search. Instead, we performed a sparse, random sampling of the conformational space to obtain seed conformers that were used to generate an adequate number of initial configurations in the alignment space. The FAP program automatically detects the 'major' acyclic rotatable bonds, excluding the rotatable bonds in the terminal groups such as a hydroxyl, methyl, carboxyl, methoxy, or *t*-butyl moiety. A number of conformers were generated by randomly assigning a specific torsion angle to each major bond. The sp^3-sp^3 single bond was allowed to adopt 60°, −60°, or 180°, which correspond to *gauche* (+), *gauche* (−), and *trans* conformations, respectively. The sp^3-sp^2 single bond was allowed to adopt 0°, 90°, 180°, or 270° if asymmetric and 0° or 90° if symmetric. When the molecule had n (≥ 4) rotatable bonds, we randomly generated 7n conformers. For example, for a molecule with ten saturated bonds, the systematic combination of the three angle values would yield 3^{10} (= 59049) conformers, but we randomly sampled only 70 conformers. If $n \leq 3$, all possible conformers were generated.

Each conformer was then energy-minimized with respect to exocyclic bond angles and torsion angles. Energy minimization was done with a module taken from an in-house modeling program that operates in the internal coordinate system and uses the BFGS optimizer.[38] In this modeling program, the various potential energy functions and the parameters were taken from Allinger's MM2[39] with a modification; the Coulomb interaction energy term was included instead of the dipole/dipole term. After all initial conformers were energy-minimized, only the conformers with energy within 12 kcal/mol above the lowest energy conformer were retained and checked for duplicate removal. The two conformers were considered redundant when the root-mean-square difference (rmsd) of the torsion angles about the major rotatable bonds was less than 15.0° and the rmsd of internal distances between atoms was less than 1.0 Å, and the one with higher energy was eliminated. Usually, most of the energy-minimized structures survived as unique conformers and were used as the seeds in the configurational search.

Rigid-Body Alignment. To obtain configurations subjected to flexible optimization, initial configurations were generated by aligning seed conformers of the sample molecule rigidly to the template. First, each conformer was positioned in such a way that its center of mass was superposed to the center of mass of the template. If the template was much larger, it was virtually partitioned into segments comparable in size to the sample. The rotational space represented by the quaternion formulation[40] was then searched for each translational configuration (usually 1); a systematic search with a grid of 45°, 60°, or 90° produces 208, 84, or 24 orientations, respectively.[41] A finer grid should be better for the completeness of the search, but the sampling with a coarse 90° grid still gives good results.[31] For efficiency purposes, we tried to reduce further the number of orientations to be optimized. We tested a random search scheme instead of the systematic one. Initial configurations were generated by assigning random numbers to the quaternions. For each configuration, the alignment score F_S in (2) was maximized with respect to three translational and three rotational parameters using the rational function optimization method.[31,35] During this rigid-body optimization, α was set to 0.15 instead of the optimal value of 0.3; this reduces the accuracy but leads to a smoother alignment surface. We found that, in all the studied cases, the completeness of the search could be achieved with only five initial orientations. The correct solutions could be obtained even with three orientations. Thus, the maximum number of initial configurations was five times the number of seed conformers times the number of translational configurations. After every initial configuration was rigidly optimized, the resulting configurations were clustered by considering two configurations identical when the rmsd of the coordinates of identical atoms was less than 1.0 Å. A cluster usually included many configurations that were often derived from several different seed conformers; the configuration with the largest F_S value was selected to represent this cluster. The representative configurations, the number of which was specified by the user, were saved for flexible optimization.

Flexible Optimization. For each optimized configuration with its own E_{min} value, the objective function F_{SE} was maximized with respect to rotational, translational, and conformational parameters of the sample molecule using the BFGS optimizer. We fixed the bond lengths and allow bond angles and torsion angles to be varied similarly to the energy minimization process. Optimization was performed in two steps varying α in F_S and ΔE_{target} in F_E to facilitate the finding of global minimum. Smaller α and larger ΔE_{target} values increase the convergence radius by making the alignment space smoother and the conformational change easier, even though the accuracy of the alignment is reduced. In the first step, α and ΔE_{target} were set to 0.20 and 6.0 kcal/mol, respectively. In the second step, only the highest scoring configurations (with F_{SE} larger than 85% of its maximum value in this study) were further optimized, setting α and ΔE_{target} to their optimal values of 0.30 and 3.0 kcal/mol, respectively. This two-step optimization scheme somewhat mimics a simulated annealing. Finally, disregarding the F_E term, the normalized alignment score F was calculated from the F_S value according to

$$F = \frac{F_S}{\sqrt{F_S^{AA} F_S^{BB}}} \tag{3}$$

where the two F_S terms in the denominator represent the self-similarity scores of molecules A and B, respectively.

Data Sets. We tested the performance of FAP with the benchmark data set described by Marialke et al.[42] It includes various ligands, 20 in total, in complex with seven proteins (dihydrofolate reductase, elastase, thrombin, estrogen receptor, carboxypeptidase A, thermolysin, and rhinovirus), which were originally used in studies of FlexS and fFlash.[26,32] The atomic coordinates of these compounds were retrieved from the PDB (http://www.pdb.org)[43] and were used without modification except for methotrexate (MTX), a dihydrofolate reductase inhibitor. The MTX structure from 4dfr contained several bonds with bad geometry and had to be corrected for proper energy evaluation. The idealized structure made using Chem3D[44] was superposed to the crystal structure with an rmsd of 0.23 Å by manually adjusting torsion angles to put every functional group as closely as possible. The modified MTX coordinates were used for all analyses involving the 4dfr structure. The absent hydrogen atoms were attached, and the partial atomic charges were calculated using the OpenBabel program (version 2.1.1, http://www.openbabel.org).[45]

For each protein, the bioactive conformations of ligands were indirectly aligned by aligning their protein backbone. In each group, every ligand served as template for both self- and cross-alignments. Pairwise rmsd values were used to assess the predictive quality of alignment and were obtained by comparison of the FAP aligned structure of the sample molecule with its X-ray aligned structure. All calculations were performed on a Pentium 4 computer (1.6 GHz process, 512 MB RAM, OS; Linux 2.6.11 Kernel) using the FAP program written in ANSI C. All figures were drawn using PyMOL (http://www.pymol.org).[46]

RESULTS AND DISCUSSION

The results for FAP flexible alignments of seven groups of ligands with the rmsds to the corresponding X-ray aligned structures are given in Table 1 and compared to the results for GMA,[42] FlexS,[26] and fFLASH[32] reported by Marialke et al.[42] Some results are also compared to those of Labute's method and FLAME.[30,33]

Details of the Methodology. The major goal in developing FAP was to obtain the highest possible accuracy with maximum efficiency. The overall procedure of the FAP method, including adjustable parameters, was established from studies on self-alignments of 20 compounds in the data set and on alignment of MTX on dihydrofolate (DHF). Self-alignment may be a trivial problem for some flexible methods, such as GMA[42] that is based on molecular graphs, FlexS[26] that is based on an incremental construction procedure, or GASP[25] that utilizes a genetic algorithm representing the conformations with the torsion angles encoded in the chromosomes. For the methods using pure atom-based 3D search procedures, however, self-alignment is as difficult as cross-alignment especially for flexible molecules with many rotatable bonds due to the large number of accessible conformations.

The results of self-alignments are shown in Table 1(a). The data set included 20 ligands for 7 proteins (the structure diagrams in Supporting Information Figure S1). The number of heavy atoms ranged from 13 to 41, and the number of major rotatable bonds ranged from 3 to 13, as are typical for drugs. For FAP alignments, the normalized F scores ranged from 0.981 to 0.997, and the rmsds to original conformations ranged from 0.12 to 0.50 Å. They should be 1.0 and 0.0 Å, respectively, for perfect overlays. As shown in Table 1(a), the rmsds ranged from 0.10 to 0.88 Å for GMA, 0.37 to 1.31 Å for FlexS, and 0.88 to 2.32 Å for fFLASH.[42] The average rmsds were 0.29, 0.38, 0.68, and 1.41 Å for FAP, GMA, FlexS, and fFLASH, respectively. These indicate that FAP, a pure atom-based 3D method, is comparable to, or in some cases even better than, GMA and FlexS in accuracy. The fact that the rmsds of FAP generally spread more narrowly indicates that it consistently works well regardless of the kind of molecules.

The FAP results were obtained by minimizing the initial configurations in two steps: first setting α in F_S (the similarity score) as 0.20 and ΔE_{target} in F_E (the energy penalty) to 6.0 kcal/mol and then setting them to 0.30 and 3.0 kcal/mol, respectively. Among the many protocols tested, this type of adjustment of the two parameters was found to be the most effective for the finding of global minima. The energy penalty term F_E was defined so as to satisfy the following requirements: (1) it should be differentiable, (2) it should increase rapidly when the energy increases over the allowed value and decrease, even to a negative value, when the lowering of the energy increases the similarity score, (3) it should not, however, become excessively large; otherwise, the system becomes stiff and the problem becomes ill-conditioned,[36] and (4) its variation should not significantly affect the optimization behavior of the whole objective function. Despite being a simple function, the third power of the $\Delta E_{strain}/\Delta E_{target}$ ratio was found to meet these criteria very effectively, making it possible to avoid unrealistic conformations during alignment.

For each self-alignment, the local strain energy ΔE_{strain}, i.e., the energy difference between the final aligned form and the initial form of the sample, is listed in Table 1(a). The ΔE_{strain} values ranged from 1.6 to 7.0 kcal/mol, indicating that none of the aligned configurations assumed the original local minimum conformation. Correspondingly, the energy penalty term F_E varied from 0.15 to 12.7. For 7 of 20 cases, ΔE_{strain} was smaller than ΔE_{target} of 3.0 kcal/mol, and thus F_E was smaller than 1.0. After final optimization, the internal energy usually decreased as compared to that obtained for the configuration after the first optimization. The energy increase was not controlled as precisely as in our previous study with the constrained optimizer.[31] Fortunately, flexible alignment is not a problem that requires an exact fulfillment of constraints. As Perola demonstrated, the internal energy of ligands in their bioactive conformations can be significantly high as long as they do not exceed ~20 kcal/mol above the global minima.[14] Thus, FAP may be a good method because it compromises accuracy and efficiency that may well be the conflicting issues in flexible alignments. The ΔE_{strain} values for cross-alignments were similar to those for self-alignments and are not given in Table 1(b).

To help elucidate the FAP process, detailed intermediate results are described here only for the cross-alignments of MTX (4dfr) and DHF (1dhf). The dihydrofolate reductase inhibitors are frequently used as test samples for alignment programs.[5,6,8,22,26,30,31,35] An intuitive topological overlay is the 'hetero' mode with the apparently similar heterocycles on top of each other but is incorrect. The correct overlay is the experimentally observed 'X-ray' mode that differs from

Table 1. Results of FAP Alignments and Comparison to Other Alignment Methods with Respect to the RMSD of the Aligned Pose to the Known Crystal Structure[a]

(a) Self-Alignments

sample[b]	template[c]	F score/ΔE[d]	FAP	GMA	FlexS	fFlash
			Dihydrofolate Reductase			
1dhf (7)	1dhf (32)	0.996/2.1 (44.2)	0.12	0.45	0.48	1.54
4dfr (7)	4dfr (33)	0.990/2.9 (48.8)	0.31	0.68	0.46	1.84
			Elastase			
1ela (11)	1ela (32)	0.985/5.3 (85.3)	0.50	0.35	0.62	0.89
1ele (8)	1ele (29)	0.992/2.8 (25.6)	0.40	0.35	0.37	1.20
			Thrombin			
1dwc (11)	1dwc (35)	0.983/3.5 (99.3)	0.35	0.30	0.60	1.28
1dwd (11)	1dwd (37)	0.984/4.8 (99.1)	0.37	0.48	0.45	1.86
			Estrogen Receptor			
1err (7)	1err (34)	0.996/3.2 (31.8)	0.17			
3ert (9)	3ert (29)	0.981/3.5 (46.0)	0.34			
			Carboxypeptidase A			
6cpa (10)	6cpa (33)	0.990/5.1 (68.8)	0.30	0.32	1.31	1.37
7cpa (13)	7cpa (41)	0.992/7.0 (166.5)	0.26	0.17	1.25	2.32
1cbx (3)	1cbx (15)	0.996/2.7 (1.8)	0.16	0.10	0.44	0.88
3cpa (4)	3cpa (17)	0.997/2.6 (4.7)	0.15			
			Thermolysin			
1tlp (10)	1tlp (37)	0.992/4.3 (92.8)	0.28	0.87	0.86	1.48
1tmn (11)	1tmn (36)	0.991/4.7 (100.5)	0.31	0.88	0.72	1.85
2tmn (5)	2tmn (13)	0.994/1.6 (7.2)	0.17	0.37	0.37	0.97
3tmn (6)	3tmn (22)	0.996/2.7 (19.4)	0.18	0.37	1.15	1.00
			Rhinovirus			
2r04 (10)	2r04 (25)	0.984/5.7 (41.3)	0.40	0.32	0.67	1.37
2r06 (8)	2r06 (23)	0.994/3.8 (30.3)	0.23	0.11	0.44	1.28
2rr1 (10)	2rr1 (26)	0.989/4.9 (50.7)	0.34	0.14	0.73	1.21
2rs3 (11)	2rs3 (27)	0.981/5.0 (54.6)	0.46	0.14	0.69	1.64
		average rmsd (Å)	0.287	0.376	0.683	1.411

(b) Cross Alignments[e]

sample template	F score[f]	FAP[g]	GMA[h]	FlexS	fFlash	Labute	Flame
			Dihydrofolate Reductase				
4dfr ≈ 1dhf	0.914/0.943	1.62/2.23	2.45 (0.89)	1.36	2.30		
1dhf ≈ 4dfr	0.927/0.937	1.45/2.20	2.35 (0.92)	1.47	1.73		
1dhf ≈ 1dls	0.900/0.962	1.16/1.87				1.40	
			Elastase				
1ele ≈ 1ela	0.910	0.68	0.79 (0.84)	0.65	1.22		
1ela ≈ 1ele	0.918	1.47	0.86 (0.73)	0.96	1.08		
			Thrombin				
1dwc ≈ 1dwd	0.858	1.07	1.69 (0.74)	1.64	2.28		
1dwd ≈ 1dwc	0.878	1.01	1.68 (0.71)	1.44	1.93		
1dwd ≈ 1dwe	0.851	0.90	2.45				1.42
			Estrogen Receptor				
3ert ≈ 1err	0.872	5.00	4.80				1.68
1err ≈ 3ert	0.889	2.26				1.60	
			Carboxypeptidase A				
7cpa ≈ 6cpa	0.908	1.36	2.13 (0.81)	0.93	×		
6cpa ≈ 7cpa	0.914	0.91	0.68 (1.00)	0.69	1.39		
1cbx ⟨⟨ 6cpa	0.681	0.91	0.66 (0.73)	0.78	1.03		
1cbx ⟨⟨ 7cpa	0.612	1.07	8.50 (0.87)	0.88	1.13		
1cbx ≈ 3cpa	0.934	0.70				1.30	
3cpa ≈ 1cbx	0.910	0.87	1.10	1.42			1.28
6cpa ⟩⟩ 1cbx	0.724	×	5.30 (0.31)	12	×		
7cpa ⟩⟩ 1cbx	0.657	×	8.49 (0.30)	12	×		
			Thermolysin				
1tmn ≈ 1tlp	0.912	1.46	1.68 (0.82)	1.24	3.43		
1tlp ≈ 11tmn	0.898	1.20	2.30 (0.74)	1.34	2.94		
3tmn ⟨ 1tmn	0.803	1.15	1.09 (0.89)	0.78	1.24		
3tmn ⟨ 1tlp	0.763	1.16	1.05 (0.89)	0.95	1.25		
1tlp ⟩ 3tmn	0.785	4.90	4.02 (0.56)	12	3.86		
1tmn ⟩ 3tmn	0.813	4.70	4.18 (0.64)	12	4.10		
2tmn ⟨⟨⟨ 1tlp	0.662	3.53	1.08 (0.88)	0.79	1.10		
2tmn ⟨⟨⟨ 1tmn	0.652	5.10	1.49 (0.65)	0.63	1.24		
2tmn ⟨⟨ 3tmn	0.845	5.03	3.19 (0.59)	3.01	1.22		
1tlp ⟩⟩⟩ 2tmn	0.705	5.91	2.75 (0.35)	12	×		
1tmn ⟩⟩⟩ 2tmn	0.677	8.67	6.84 (0.28)	12	×		
3tmn ⟩⟩ 2tmn	0.831	5.94	5.76 (0.37)	12	×		
			Rhinovirus				
2r06 ≈ 2r04	0.950	1.01	1.33 (0.96)	1.30	1.80		
2r04 ≈ 2r06	0.973	1.45	1.96 (0.88)	0.78	1.49		
2rs3 ≈ 2rr1	0.972	0.75	0.58 (0.96)	0.68	1.66		
2rr1 ≈ 2rs3	0.971	0.59	0.43 (1.00)	0.90	1.29		
2r04 ≈ 2rr1	0.877/0.970	1.16/12.68	12.82 (1.00)	10	R		
2rr1 ≈ 2r06	0.880/0.964	1.77/12.27	12.82 (0.85)	10	R		
2r04 ≈ 2rs3	0.866/0.958	1.14/12.72	12.96 (1.00)	10	R		
2rs3 ≈ 2r04	0.887/0.957	1.67/13.50	13.91 (0.93)	10	R		
2rr1 ≈ 2r04	0.884/0.976	1.33/12.99	13.32 (0.96)	10	R		
2rs3 ≈ 2r06	0.870/0.942	2.32/12.69	13.08 (0.81)	10	R		
2r06 ≈ 2rr1	0.850/0.934	1.53/11.26	11.55 (0.96)	10	R		
2r06 ≈ 2rs3	0.821/0.909	1.64/11.33	11.59 (0.96)	10	R		

[a] Rmsds (Å) are given in the column of each alignment method; those for the other methods were taken from ref 42. [b] Number of major torsion angles in parentheses. [c] Number of heavy atoms in parentheses. [d] ΔE in kcal/mol and CPU time in seconds in parentheses. [e] Results are shown only for the pairs studied by other methods; R for reverse alignment; × for meaningless comparison; The symbols denote the ratio (r) of the number of atoms in the two molecules: ≈ if $r < 1.5$; ⟨ or ⟩ if $1.5 \le r < 2.5$; ⟨⟨ or ⟩⟩ if $2.5 \le r < 4.0$; ⟨⟨⟨ or ⟩⟩⟩ if $r \ge 4.0$. [f] Scores and [g] rmsds for the X-ray/the highest scoring poses when two numbers are given; otherwise for the highest scoring poses. [h] The numbers in parentheses denote the proportion of atoms in the sample molecule that are matched to an atom in the template on the molecular graphs.

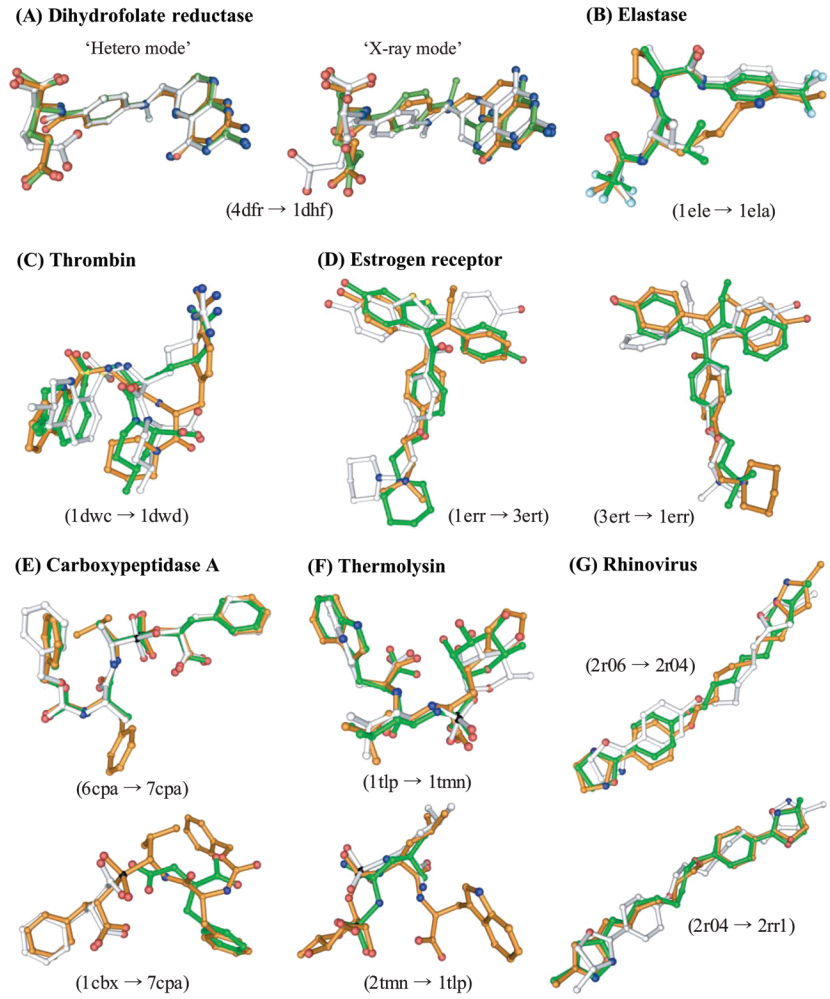

Figure 1. Alignments of molecules within each data set. Each figure contains the template molecule (orange) taken from the X-ray structure, the sample (green) aligned by FAP with the highest F score, and the sample (white) aligned in the crystal. In the 'hetero' mode shown in part (A), there is no corresponding crystal structure, and the sample molecule in white is the FAP aligned structure with the second highest F score. The O, N, F, S, and P heteroatoms are colored in red, blue, cyan, yellow, and black, respectively.

the hetero mode by a ring flip of 180°.[47] The alignment methods that utilize the pharmacophoric features readily give the X-ray mode that is consistent with hydrogen bonding patterns of the two heterocycles.[6,22] However, as noted by Wolber et al., the alignments of the two inhibitors are problematic for pure atom-based methods.[6] These methods, including our previous version with the similarity index F_S and a constrained optimizer, tend to give the hetero mode as the top-scoring solution.[26,31,35] Although the Labute's method is also an atom-based method, it gave the correct overlay but only when the relative weights between several terms defining its similarity measure were properly adjusted.[30] In the present study, we did not attempt to adjust the similarity index to obtain better results for this particular case.

In the 4dfr →1dhf alignment (a notation of sample → template denoted with PDB codes will be used hereafter), MTX used as a sample contains seven 'major' rotatable bonds (see Figure S1). Initial conformers were generated by randomly assigning 0° or 90° to τ_1 and τ_4 and 60°, 180°, or 300° to the others. A complete search, even with such coarse grids, would give a total of 972 ($2^2 \times 3^5$) conformers. Forty-nine (7n) conformers, i.e., 5.0% of the total, were generated and used as seeds. After energy minimization and clustering, 47 local minima were identified with internal energies ranging from −2.21 to 9.43 kcal/mol. A total of 235 initial configurations, 5 for each conformer, were then generated. These were rigidly optimized, and 104 remained after clustering. Among these, 36 configurations with F_S values larger than 85% of the maximum value were subjected to the first round of flexible optimization, with $\alpha = 0.2$ and $\Delta E_{target} = 6.0$. Finally, 20 configurations with F_S values larger than 85% of the maximum were further optimized, with $\alpha = 0.3$ and $\Delta E_{target} = 3.0$. It took 40.2 s for the whole process on a Pentium 4 computer (1.6 GHz), 6.8 s for conformer generation, 10.5 s for rigid-body optimization, and 22.9 s for flexible optimization.

As expected, the top-scoring solution ($F = 0.943$) was the hetero mode (Figure 1A). For this solution, the initial energy (E_{min}) of the seed conformer was 4.29 kcal/mol, $\Delta E_{strain} = 2.59$ kcal/mol, and $F_E = 0.65$ (Table 1(b)). For the second solution ($F = 0.918$), which also corresponds to the hetero mode, they were 1.68 kcal/mol, 5.05 kcal/mol, and 4.78, respectively. The two differ only in the conformation of the glutamate moiety, as shown in Figure 1A, and originated from different seed conformers. The third solution ($F = 0.914$) was the correct X-ray mode with an rmsd of 1.62 Å to the X-ray aligned structure (Figure 1A and Table 1(b)). Its energy was 4.70 kcal/mol and both ΔE_{strain} and F_E were 0.0, whereas the energy of the seed conformer was 4.92

kcal/mol. It is an interesting unusual case in that the flexible alignment induced one local minimum conformation to be changed into another local minimum with lower energy instead of a higher energy conformation. Such lowering of the energy during flexible alignment had not been observed in other cases. In these solutions, the initial conformation of the sample was significantly altered after flexible optimization. The torsion angles flanked by two rings varied up to 78°, whereas those in the glutamate moiety varied up to 53°. Two distinct seed conformers did not evolve to the same solution. Because the search was performed in a completely random manner, the results from repeated runs of the same job were not exactly the same. However, the essential features of the high-scored poses never changed; the *F* scores varied only slightly (<∼0.1%).

In the 1dhf → 4dfr alignment performed with the same protocol, 46 local minima of 49 initial seeds were obtained with internal energies ranging from −3.26 to 6.96 kcal/mol, and a total of 230 initial configurations were generated. Among these, 120 remained after rigid-body optimization and clustering, and 49 and 25 configurations were subjected to the first and second rounds of flexible optimization, respectively. It took 40.3 s for the whole process. The top-scoring solution ($F = 0.937$) was the hetero mode. For this solution, the initial energy (E_{min}) of the seed conformer was −0.24 kcal/mol, $\Delta E_{strain} = 4.21$ kcal/mol, and $F_E = 2.77$. For the second solution ($F = 0.927$), which corresponds to the X-ray mode, they were 1.31 kcal/mol, 3.45 kcal/mol, and 1.52, respectively. The rmsd to the X-ray aligned structure was 1.45 Å.

In both alignments, the other solutions with lower *F* scores were mostly a subset of either the hetero or the X-ray mode that differed in the conformation of the bonds linking the two rings or the glutamate moiety. The fact that these solutions were obtained as distinct local minima substantiates that the search protocol developed in this study seems to ensure the completeness of the FAP method.

The CPU time required for each self-alignment is listed in Table 1(a). Obviously it took a longer time to compute the alignment as the flexibility and size of molecule increased. The number of rotatable bonds apparently exerted a greater effect on the computing time than the number of atoms. For a given sample, the total computing time depends approximately linearly on the number of seeds used in the search of conformational and configurational spaces as well as on the number of initial configurations subjected to flexible optimization. Thus, computing time can be saved by reducing these numbers. Test runs with several samples indicated that the minimum number of initial conformers and configurations required for obtaining correct solutions are 5n and 3 × 5n, respectively, where n is the number of major rotatable bonds. For the 4dfr →1dhf alignment, for example, both hetero and X-ray modes were obtained when 102 initial configurations were rigidly optimized, and only the 5 highest-scored ones were flexibly optimized. However, the second solution described above disappeared. It took only 12.8 s (4.7, 4.2, and 3.9 s for each of the three steps, respectively), which is less than one-third of the computing time necessary with more rigorous conditions. Thus this set of smaller numbers may be suitable for virtual screening, in which completeness may well be risked in favor of efficiency.

Prediction of Crystal Binding Modes. The results for other cross-alignments are given in Table 1(b) and discussed below for the ligands of each protein. All results were obtained using 7n initial conformers and 5 × 7n configurational seeds. Only the top-scoring solutions were considered, except for rhinovirus ligands.

Elastase. The inhibitors 1ela (trifluoroacetyl-L-lysyl-L-prolyl-*p*-isopropylanilide) and 1ele (trifluoro-acetyl-L-valyl-L-alanyl-*p*-trifluoromethylanilide) have similar structures, and the smaller 1ele can be considered a substructure of 1ela.[48,49] These molecules aligned well on each other. The 1ele → 1ela alignment was relatively trivial, showing excellent predictability. The rmsd to the X-ray aligned structure was 0.68 Å. For the 1ela → 1ele alignment (Figure 1B), the rmsd was 1.47 Å and is thus larger than the values obtained with other methods. The largest difference between the predicted and the X-ray aligned structures was in the conformation of the long lysyl side chain in 1ela. The corresponding Nξ and Cε atoms were 5.7 and 4.1 Å apart, respectively. Excluding these two atoms, the rmsd was 0.81 Å and is comparable to the values obtained with GMA, FlexS, and fFLASH. As Marialke et al. have previously reported, the lysyl side chain that binds in the S2 subsite of elastase cannot be aligned because the corresponding valyl side chain in 1ela is significantly shorter and does not provide any information with respect to the conformation of the side chain.[42] They suggested that their solution as well as the other reported placements in Table 1(b) were either fortuitous or simply due to sampling around solutions with an optimal score and picking out the one with the best rmsd.

Thrombin. The ligands 1dwd (N^{α}-(2-naphthyl-sulfonyl-glycyl)-D-*para*-amidinophenylalanyl-piperidine) and 1dwc ((2R, 4R)-4-methyl-l-[N^{α}-[(3-methyl-1,2,3,4-tetrahydro-8-quinolinyl)-sulfonyl]-L-arginyl)]-2-piperidine-carboxylic acid) show relatively low similarity at the graph level; 71 and 74% of the atoms can be matched. Nevertheless, both 1dwd → 1dwc (Figure 1C) and 1dwc → 1dwd alignments were excellent. In the protein complexes, the piperidine and the larger hydrophobic ring bind in the P- and D-pockets of thrombin, respectively, and pack tightly against each other, the SO_2 groups are exposed to solvent, and the amidino moieties bind deep inside the binding pocket.[50] These groups were aligned very well despite the shorter "backbone" of 1dwc. The rmsds to the X-ray aligned structures were ∼1 Å for both alignments, and are thus significantly smaller than those obtained with other methods. In fact, the FAP aligned structure and the X-ray aligned structure themselves were more similar than implied by rmsds. In both alignments, they were shifted by about 0.8 Å from one another owing to the alignment of protein structures. The accuracy of the FAP method is well illustrated by this example.

Estrogen Receptor. Upon binding of antagonists 1err (raloxifene) and 3ert (4-hydroxytamoxifen), the estrogen receptor does not change its conformation.[51,52] This case is interesting in that the result can be quite different depending on which antagonist is used as template. Both 3ert → 1err and 1err → 3ert alignments seemed reasonable, but when compared to the X-ray structure the former was wrong with an rmsd of 5.0 Å, whereas the latter was correct. In the aligned structure of 3ert, the phenol ring was matched to the phenol moiety of 1err, but it was directed toward the benzo[*b*]thiophen-6-ol moiety in the X-ray aligned structure.

The X-ray structure could be obtained if the single bond linking the stilbene moiety to the other phenyl group was rotated by approximately 180°. GMA gave a similar result with an rmsd of 4.8 Å for the 3ert → 1err alignment. Among FAP solutions, the fourth-ranked solution with an *F* score of 0.819 was the closest to the X-ray aligned structure with an rmsd of 1.51 Å. In contrast, Labute's method and FLAME gave correct results with rmsds of 1.60 and 1.68 Å, respectively. For FLAME, however, this result was obtained only after a heavy weight was assigned to the hydroxyl O atom of the fused heterocycle.[33] Without parameter tuning, FLAME also gave the inverted orientation with an rmsd of 5.10 Å. The 1err → 3ert alignment, for which the other methods did not report any result, was good with an *F* value of 0.889 and an rmsd of 2.26 Å (Figure 2D). The rmsd became 1.37 Å when the misplaced pyridine moiety was excluded.

Carboxypeptidase A. The studied ligands included two large phosphonate inhibitors, 7cpa and 6cpa, and two small ones, 1cbx and 3cpa. 6cpa (N-carbobenzoxy-Ala-Ala^{P}-(O)-Phe) is a substructure of 7cpa (N-carbobenzoxy-Phe-Val^{P}-(O)-Phe). Both compounds have four distinct groups, as shown by their chemical formulas that bind the protein in the S3, S2, S1, and S1′ subsites.[53] Both 1cbx (L-benzylsuccinate) and 3cpa (glycyl-L-tyrosine) bind in the S1′ subsite. The 6cpa → 7cpa alignment was nearly perfect (Figure 1E). The rmsd of 0.91 Å was due mainly to an overall shift with respect to the X-ray aligned structure rather than to a difference in the conformations. The 7cpa → 6cpa alignment displayed a slightly worse predictability with an rmsd of 1.67 Å. The bulkier side chain of phenylalanine in 7cpa was positioned correctly even though the corresponding methyl group in 6cpa did not provide any information. In the predicted structure, the peptide moiety linking the phenylalanine and valine residues was completely flipped compared to the one in the crystal structure, and this was the major difference between the two.

Both 1cbx and 3cpa aligned well on each other with excellent predictability; *F* and rmsd were 0.934 and 0.70 Å for the 1cbx → 3cpa and 0.910 and 0.87 Å for the 3cpa → 1cbx, respectively. The small 1cbx aligned relatively well on the larger 6cpa and 7cpa, with rmsds of ∼1 Å, but the two alignments behaved differently. For the 1cbx → 6cpa alignment, a single distinct solution that corresponded to the binding mode in the S1′ subsite was obtained. In the case of 1cbx → 7cpa, the top-scoring solution was the correct binding mode in the S1′ site. However, there were two more competing solutions with slightly smaller *F* scores, and they corresponded to the binding modes in the S2 and S3 sites, respectively. The GMA method failed in this alignment because 7cpa has three groups of equal similarity to 1cbx at the graph level and the mapped group that was not the correct solution was arbitrarily chosen.[42] Similarly to other methods, alignments of large ligands on 1cbx failed. Both 6cpa and 7cpa resulted in compactly folded conformations, and it was meaningless to compare them to the X-ray structures.

Thermolysin. The studied ligands included two large inhibitors of 1tlp and 1tmn, a medium-sized one, 3tmn, and a very small one, 2tmn.[54−56] 1tlp (*N*-(α-L-rhamnopyranosyloxyhydroxyphosphinyl)-L-leucyl-L-trptophan) and 1tmn (*N*-(1-carboxy-3-phenylpropyl)-L-leucyl-L-tryptophan) aligned very well on each other (Figure 1F). The rmsds to the X-ray aligned structures were 1.46 and 1.20 Å, respectively, and are comparable to those obtained with FlexS and better than those obtained with GMA and fFLASH. 3tmn (L-valyl-L-tryptophan) aligned well on 1tlp and 1tmn. However, the smaller 2tmn (*N*-phosphoryl-L-leucinamide) failed to align on the three larger ligands. This contrasts with the successful alignments of carboxypeptidase A ligands and is the only case for which the other methods outperformed the FAP method. The failure of the 2tmn → 1tlp alignment is particularly noteworthy because 2tmn is a substructure of 1tlp (Figure 1F). In the aligned X-ray structures, 2tmn aligned well on the *N*-phosphinylleucyl moiety of 1tlp. This failure indicates that the algorithm in the translational search during the generation of initial configurations needs to be improved. As in the case of carboxypeptidase A, the alignments of large ligands on smaller ones failed. As noted by Marialke et al., this is clearly a general method-independent problem, because not enough information is given by the alignment to place the complete ligand.[42]

Rhinovirus. Rhinovirus shows alternative binding modes with the antiviral agents called WIN compounds.[57] They are composed of an extended phenyl-O-$(CH_2)_n$ spacer with heterocycles at both terminal ends, and they show a certain degree of pseudosymmetry. The studied ligands included 2r04, 2r06, 2rr1, and 2rs3; n of the spacer was 5 for 2r06 and 7 for the others. One pair (2r04 and 2r06) showed an inverse binding mode compared to the other pair (2rr1 and 2rs3). Ligands within a pair aligned very well on each other (Figure 1G); the rmsds to the X-ray aligned structures ranged from 0.59 to 1.45 Å. Without exception, the alignment of two ligands from each pair yielded a high-scoring solution with an orientation inverse to that of the X-ray aligned structure. The average *F* and rmsd were 0.95 and 12.4 Å, respectively. In each case, the alignment also yielded the crystal binding mode but only as a low-scoring solution. The average *F* and rmsd were 0.87 and 1.6 Å, respectively.

Comparison of Flexible Alignment Methods. Among many flexible alignment methods, Labute's method[30] and FLAME[33] are the most relevant for FAP in that they optimize the objective function consisting of the similarity score and the term for internal energy based on a complete, all-atom force field. However, the objective functions and the search methods differ and are compared below from methodological viewpoints. The details of similarity scores are not discussed here.

In Labute's method, conformational and alignment spaces are simultaneously searched using a random incremental pulse search procedure.[30] Trial configurations are minimized using a truncated Newton optimizer. The collection of configurations, in which the average potential energy is lower than the minimum observed energy plus some predefined threshold, is then examined. In FLAME, initial alignments are generated by first finding the maximum common pharmacophores between the two molecules using a genetic algorithm.[33] The encoding scheme consists of torsion angles of noncyclic rotatable bonds, and the fitness score is the number of common pharmacophores. Trial configurations are then minimized using the BFGS algorithm, and alignments with high scores are visually inspected. The objective function for flexible optimization is $(-kT \log F + U)$ for Labute's method and $(-TF + U_{sample} + U_{template})$ for FLAME, where *F* is the alignment score of each method, *U* is the

internal energy, and T is a factor that controls the relative weighting of alignment and energy. T significantly affects the optimization behavior, and the determination of a proper value is problematic. In Labute's method, $T = 30{,}000$, which produced approximately 1% frequency of occurrence of internal energies greater than 20 kcal/mol, was selected as a default value.[30] In FLAME, the authors set $T = 1.0$ as a default value, asserting that the rule of thumb is to maintain a similar order of magnitude between internal energy and alignment score.[33]

In comparison, our search procedure is a simple combination of sparse, random sampling of the conformational space followed by energy minimization, and random sampling of the configurational space followed by rigid-body optimization. Each initial configuration is minimized in two steps using the BFGS algorithm. The objective function is $-F_S + F_E$, where F_S is the alignment score and F_E is the energy penalty term. It does not include a scaling factor. FAP is unique in that the energy penalty term is used instead of the internal energy itself.

For Labute's method and FLAME, only several pairwise alignments were reported,[30,33] and some of the results are listed in Table 1(b). The FAP results were better than those obtained with Labute's method for 1dhf → 1dls and 1cbx → 3cpa alignments but not for 1err → 3ert. They were also better than those obtained with FLAME for 1dwd → 1dwe and 3cpa → 1cbx alignments, but FLAME outperformed for 3ert → 1err. It should be noted, however, that the FLAME results were obtained only after adjusting the probe weights of specific atoms for each alignment. The limited data sets of the other methods hampered a rigorous comparison of the performances. Overall, FAP performed well except for estrogen receptor ligands.

CONCLUSIONS

In summary, we have developed a FAP methodology that is capable of aligning small molecules in a truly flexible manner. It was our aim to develop a fully automated method that does not require a priori knowledge on the conformation or predefined relationship of particular functional groups but is still accurate and efficient. FAP performs alignments in an objective manner. It uses only 3D structural and partial atomic charge information. The optimized default values for adjustable parameters were previously found to be transferable.[31] In terms of the ability to produce the bioactive overlays, FAP is comparable to, or in some cases even better than, the existing methods such as GMA, FlexS, fFlash, Labute's, and FLAME methods. FAP is relatively fast, albeit being a pure atom-based 3D method, largely due to its efficient search and optimization schemes. The sparse, random sampling scheme developed in this study proved to be a very efficient way to search the highly complex conformational and alignment spaces. The two-step optimization scheme that mimics a simulated annealing allows the objective function to maintain a large radius of convergence without losing accuracy. These search protocols seem to work well regardless of the type of molecule. In developing a flexible alignment method, accuracy and efficiency are the issues difficult to achieve simultaneously. FAP may be assessed as an alignment method with excellent accuracy and reasonable efficiency.

A conformer DB is not necessary for FAP alignments but can be used. Then the computing time will be reduced by ~30% because the conformer generation step can be skipped. Although FAP is not a very fast method, it can still be used as a virtual screening tool. A state of the art CPU for a desktop PC (for example, the Intel Core 2 E8500, 3.16 GHz) is about ten times faster than the CPU (Pentium 4, 1.6 GHz) used in this study (http://www.cpubenchmark.net/ common_cpus.html). It took 12.8 s when the 4dfr →1dhf alignment was done using minimum number of initial seeds. Extrapolating this case, therefore, at least 70,000 molecules comparable to MTX can be screened in a day with such a fast CPU. MTX has seven 'major' rotatable bonds, and most compound DBs contain many molecules with less rotatable bonds. Thus a single CPU can treat more than 100,000 molecules a day, and a DB containing millions of compounds can be screened within days with several PCs. In its present form, FAP works very well not only for congeneric molecules but also for ostensibly dissimilar ones provided that the overlaid molecules do not significantly differ in size and may thus be used as a powerful tool in particular for scaffold hopping.

ACKNOWLEDGMENT

We are grateful to the anonymous referees for their valuable comments and suggestions. This research was supported by the grant (CBM32-B2000-01-03-00) from the Center for Biological Modulators of the 21C Frontier R&D Program, Ministry of Education, Science and Technology, Korea.

Supporting Information Available: Chemical structure diagrams for 20 molecules in the data set. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993.
(2) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215–232.
(3) Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M. Novel Technologies for Virtual Screening. *Drug Discovery Today* **2004**, *9*, 27–34.
(4) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225–233.
(5) Richmond, N. J.; Willett, P.; Clark, R. D. Alignment of Three-Dimensional Molecules Using an Image Recognition Algorithm. *J. Mol. Graphics Modell.* **2004**, *23*, 199–209.
(6) Wolber, G.; Dornhofer, A. A.; Langer, T. Efficient Overlay of Small Organic Molecules Using 3D Pharmacophores. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 773–788.
(7) Ballester, P. J.; Richards, W. G. Ultrafast Shape Recognition to Search Compound Databases for Similar Molecular Shapes. *J. Comput. Chem.* **2007**, *28*, 1711–1723.
(8) Todorov, N. P.; Alberts, I. L.; de Esch, I. J. P.; Dean, P. M. QUASI: A Novel Method for Simultaneous Superposition of Multiple Flexible Ligands and Virtual Screening Using Partial Similarity. *J. Chem. Inf. Model.* **2007**, *47*, 1007–1020.
(9) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; John Wiley: New York, 1990.
(10) Bender, A.; Glen, R. C. Molecular Similarity: A Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
(11) Sheridan, R. P.; Kearsley, S. K. Why Do We Need So Many Chemical Similarity Search Methods. *Drug Discovery Today* **2002**, *7*, 903–911.

(12) Melani, F.; Gratteri, P.; Adamo, M.; Bonaccini, C. Field Interaction and Geometrical Overlap: A New Simplex and Experimental Design Based Computational Procedure for Superposing Small Ligand Molecules. *J. Med. Chem.* **2003**, *46*, 1359–1371.
(13) Boström, J.; Norrby, P. O.; Liljefors, T. Conformational Energy Penalties of Protein-Bound Ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383–396.
(14) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* **2004**, *47*, 2499–2510.
(15) Boström, J.; Hogner, A.; Schmitt, S. Do Structurally Similar Ligands Bind in a Similar Fashion. *J. Med. Chem.* **2006**, *49*, 6716–6725.
(16) Hao, M. H.; Haq, O.; Muegge, I. Torsion Angle Preference and Energetics of Small-Molecule Ligands Bound to Proteins. *J. Chem. Inf. Model.* **2007**, *47*, 2242–2252.
(17) Halgren, T. A. The Merck Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
(18) Boström, J.; Greenwood, J. R.; Gottfries, J. Assessing the Performance of OMEGA with Respect to Retrieving Bioactive Conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449–462.
(19) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative Performance Assessment of the Conformational Model Generators Omega and Catalyst: A Large-Scale Survey on the Retrieval of Protein-Bound Ligand Conformations. *J. Chem. Inf. Model.* **2006**, *46*, 1848–1861.
(20) Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. Conformational Sampling of Bioactive Molecules: A Comparative Study. *J. Chem. Inf. Model.* **2007**, *47*, 1067–1086.
(21) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. The Ensemble Approach to Distance Geometry: Application to the Nicotinic Pharmacophore. *J. Med. Chem.* **1986**, *29*, 899–906.
(22) Kato, Y.; Inoue, A.; Yamada, M.; Tomioka, N.; Itai, A. Automatic Superposition of Drug Molecules Based on Their Common Receptor Site. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 475–486.
(23) McMartin, C.; Bohacek, R. S. Flexible Matching of Test Ligands to a 3D Pharmacophore Using a Molecular Superposition Force Field: Comparison of Predicted and Experimental Conformations of Inhibitors of Three Enzymes. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 237–250.
(24) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. Sampling Conformational Hyperspace: Techniques for Improving Completeness. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 491–499.
(25) Jones, G.; Willett, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–549.
(26) Lemmen, C.; Lengauer, T.; Klebe, G. FlexS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
(27) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220–232.
(28) Mills, J. E.; de Esch, I. J.; Perkins, T. D.; Dean, P. M. SLATE: A Method for the Superposition of Flexible Ligands. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 81–96.
(29) Pitman, M. C.; Huber, W. K.; Horn, H.; Krämer, A.; Rice, J. E.; Swope, W. C. FLASHFLOOD: A 3D Field-Based Similarity Search and Alignment Method for Flexible Molecules. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 587–612.
(30) Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schmidt, J. M. Flexible Alignment of Small Molecules. *J. Med. Chem.* **2001**, *44*, 1483–1490.
(31) Chae, C. H.; Oh, D. G.; Shin, W. Flexible Molecular Superposition: Development of a Combined Similarity Index and Application of the Constrained Optimization Technique. *J. Comput. Chem.* **2001**, *22*, 888–900.
(32) Krämer, A.; Horn, H. W.; Rice, J. E. Fast 3D Molecular Superposition and Similarity Search in Databases of Flexible Molecules. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 13–38.
(33) Cho, S. J.; Sun, Y. FLAME: A Program to Flexibly Align Molecules. *J. Chem. Inf. Model.* **2006**, *46*, 298–306.
(34) *Multiobjective Decision Making: Theory and Methodology*; Chankong, V., Haimes, Y. Y., Eds.; North-Holland: New York, 1983.
(35) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615–633.
(36) Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley: New York, 1987.
(37) Klebe, G.; Mietzner, T.; Weber, F. Methodological Developments and Strategies for a Fast Flexible Superposition of Drug-Size Molecules. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35–49.
(38) Shin, W.; Oh, D. G.; Chae, C. H.; Yoon, T. S. Conformational Analyses of Thiamin-Related Compounds: A Stereochemical Model for Thiamin Catalysis. *J. Am. Chem. Soc.* **1993**, *115*, 12238–12250.
(39) Allinger, N. L.; Zhou, X.; Bergsma, J. Molecular Mechanics Parameters. *J. Mol. Struct.-Theochem.* **1994**, *312*, 69–83.
(40) Karney, C. F. F. Quaternions in Molecular Modeling. *J. Mol. Graphics Modell.* **2007**, *25*, 595–604.
(41) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. MIMIC: A Molecular-Field Matching Program. Exploiting Applicability of Molecular Similarity Approaches. *J. Comput. Chem.* **1997**, *18*, 934–954.
(42) Marialke, J.; Körner, R.; Tietze, S.; Apostolakis, J. Graph-Based Molecular Alignment (GMA). *J. Chem. Inf. Model.* **2007**, *47*, 591–601.
(43) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(44) *Chem3D, Version 9.0*; CambridgeSoft Co.: Cambridge, MA, U.S.A., 2004.
(45) Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L. The Blue Obelisk-Interoperability in Chemical Informatics. *J. Chem. Inf. Model.* **2006**, *46*, 991–998.
(46) *The PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, CA, U.S.A., 2008.
(47) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. Crystal Structures of Escherichia coli and *Lactobacillus casei* Dihydrofolate Reductase Refined at 1.7 Å Resolution. I. General Features and Binding of Methotrexate. *J. Biol. Chem.* **1982**, *257*, 13650–13662.
(48) Mattos, C.; Rasmussen, B.; Ding, X.; Petsko, G. A.; Ringe, D. Analogous Inhibitors of Elastase Do Not Always Bind Analogously. *Nat. Struct. Biol.* **1994**, *1*, 55–58.
(49) Mattos, C.; Giammona, D. A.; Petsko, G. A.; Ringe, D. Structural Analysis of the Active Site of Porcine Pancreatic Elastase Based on the X-ray Crystal Structures of Complexes with Trifluoroacetyl-dipeptide-anilide Inhibitors. *Biochemistry* **1995**, *34*, 3193–3203.
(50) Banner, D. W.; Hadvary, P. Crystallographic Analysis at 3.0-Å Resolution of the Binding to Human Thrombin of Four Active Site-Directed Inhibitors. *J. Biol. Chem.* **1991**, *266*, 20085–20093.
(51) Brzozowski, A. M.; Pike, A. C.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular Basis of Agonism and Antagonism in the Oestrogen Receptor. *Nature* **1997**, *389*, 753–758.
(52) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell* **1998**, *95*, 927–937.
(53) Kim, H.; Lipscomb, W. N. Comparison of the Structures of Three Carboxypeptidase A- Phosphonate Complexes Determined by X-ray Crystallography. *Biochemistry* **1991**, *30*, 8171–8180.
(54) Tronrud, D. E.; Monzingo, A. F.; Matthews, B. W. Crystallographic Structural Analysis of Phosphoramidates as Inhibitors and Transition-State Analogs of Thermolysin. *Eur. J. Biochem.* **1986**, *157*, 261–268.
(55) Monzingo, A. F.; Matthews, B. W. Binding of N-Carboxymethyl Dipeptide Inhibitors to Thermolysin Determined by X-ray Crystallography: A Novel Class of Transition-State Analogues for Zinc Peptidases. *Biochemistry* **1984**, *23*, 5724–5729.
(56) Holden, H. M.; Matthews, B. W. The Binding of L-Valyl-L-tryptophan to Crystalline Thermolysin Illustrates the Mode of Interaction of a Product of Peptide Hydrolysis. *J. Biol. Chem.* **1988**, *263*, 3256–3260.
(57) Badger, J.; Minor, I.; Oliveira, M. A.; Smith, T. J.; Rossmann, M. G. Structural Analysis of Antiviral Agents That Interact with the Capsid of Human Rhinoviruses. *Proteins* **1989**, *6*, 1–19.

CI8004463