Novel Alignment Method of Small Molecules Using the Hopfield Neural Network

Masamoto Arakawa,[†] Kiyoshi Hasegawa,[‡] and Kimito Funatsu*,[†]

Toyohashi University of Technology, Tempaku, Toyohashi 441-8580, Japan, and
Nippon Roche, Kajiwara, Kamakura 247-8530, Japan

Received January 17, 2003

Molecular alignment is an important step in three-dimensional quantitative structure−activity relationship (3D-QSAR) such as comparative molecular field analysis (CoMFA). A reasonable molecular alignment is necessary for building a 3D-QSAR model. In this paper, a novel method for molecular alignment using the Hopfield Neural Network (HNN) is introduced. Four kinds of chemical properties are assigned to each atom of a molecule. Then, those properties between two molecules correspond to each other using HNN. To validate our method, HNN was applied to 12 pairs of enzyme inhibitors cited from the Protein Data Bank (PDB). As a result, our method could successfully reproduce the real molecular alignments obtained from X-ray crystallography.

INTRODUCTION

In Quantitative Structure−Activity Relationship (QSAR), a mathematical model between structural descriptors and biological activities is constructed. Then, biological activities of unknown compounds are predicted from the model. The Hansch−Fujita method[1] is a classical QSAR technique, which constructs the multiple liner regression (MLR) model between physicochemical descriptors of substituents and biological activities. Since the Hansch−Fujita method uses only two-dimensional descriptors, predictive ability of the model is restricted. Molecular properties based on the three-dimensional (3D) structure of compounds may be useful in describing the ligand−receptor interactions. Recently, a variety of ligand-based 3D-QSAR methods such as Comparative Molecular Field Analysis (CoMFA)[2] have been developed and widely used in medicinal chemistry. CoMFA uses steric and electrostatic field variables that are calculated at grid points surrounding a whole molecule. The relationship between these 3D structural descriptors and biological activities are modeled by partial least squares (PLS). The result of CoMFA can be graphically displayed as the 3D contour map of the PLS regression coefficients. Owing to the 3D contour map, the important regions for biological activity can be easily understood.

In CoMFA and other 3D-QSAR methodologies, a proper alignment between molecules is necessary. Since molecular alignment is an important step in 3D-QSAR analysis, a variety of methodologies have been proposed[3−8] and reviewed by Lemmen and Lengauer.[9] A determination of conformation is also the major problem for constructing 3D-QSAR models, and many papers were published. For example, Hopfinger et al. proposed the 4D-QSAR formalism which incorporates conformational and alignment freedom into the development of 3D-QSAR models as the fourth dimension.[10] Doweyko et al. proposed the multiconformer technique as a novel technique.[11] Each molecule in the data set was represented by five separate low-energy conformers, and then all of them were used in the generation of HASL 3D-QSAR models. We also proposed a new method that can solve a conformation/alignment problem using a 4-way PLS formulation.[3,4] Possible 3D conformations of all molecules were generated by conformational analysis, and they were characterized by field variables of CoMFA with some alignment rules. Then, 4-way array for 4-way PLS analysis was created by similarity measure, and conformations largely contributing to inhibitory activity were selected as an active conformer by the regression coefficients. To demonstrate the general utility, the data set of glucose analogue inhibitors of glycogen phosphorylase[3] and benzodiazepine analogue inhibitors of cholecystokinin[4] were used as test samples.

In this paper, a novel method for molecular alignment using the Hopfield Neural Network (HNN) is introduced. Four kinds of chemical properties (hydrophobic group, hydrogen-bonding acceptor, hydrogen-bonding donor, and hydrogen-bonding donor/acceptor) are assigned to each atom of molecule. Then, those properties between two molecules correspond to each other using HNN. To validate our method, HNN was applied to 12 pairs of enzyme inhibitors cited from the Protein Data Bank (PDB). As a result, our method could successfully reproduce the real molecular alignments obtained from X-ray crystallography.

METHOD

The pattern matching problem can be summarized as finding the subset of atoms that have the most similar spatial arrangement as those of a given pattern. Doucet et al. proposed the method to deal with this combinatorial optimization problem with HNN.[12] The proposed methodology in this paper is the extension of this pattern matching method.

Definition of Chemical Property. The definition of chemical properties used in this study is listed below. This definition is cited from ref 5. The user can change this flexible definition depending on the alignment problem to be solved.

1. Hydrophobic group (HY): (A) aromatic ring of a phenyl, naphthalene, pyridine, thiophene, etc.; (B) an ali-

* Corresponding author phone: 81 53 244 6879; fax: 81 53 247 9315; e-mail: funatsu@tutkie.tut.ac.jp.
[†] Toyohashi University of Technology.
[‡] Nippon Roche.

10.1021/ci030001l CCC: $25.00
Published on Web 06/18/2003

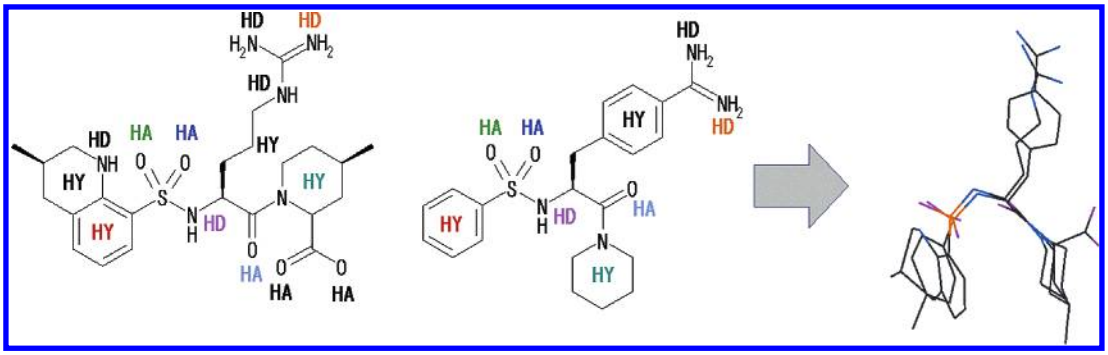


Figure 1. Example of molecular alignment based on chemical properties.

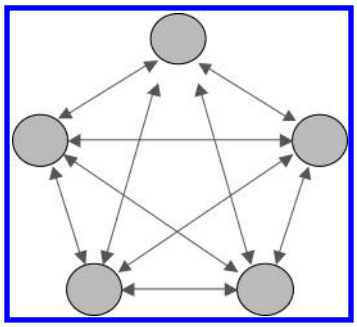
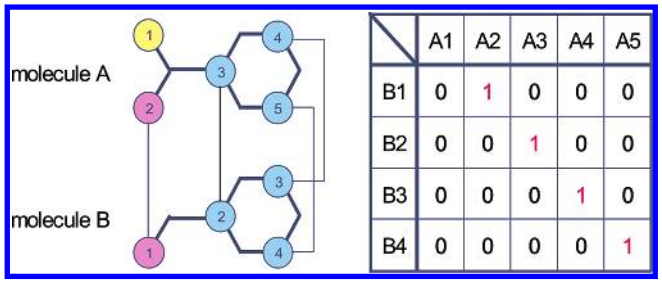


Figure 2. Structure of Hopfield Neural Network.



	A1	A2	A3	A4	A5
B1	0	1	0	0	0
B2	0	0	1	0	0
B3	0	0	0	1	0
B4	0	0	0	0	1

Figure 3. Corresponding of chemical properties.

phatic chain with alkyl chain of three atoms or more; and (C) trifluoromethyl.

2. Hydrogen-bonding donor (HD): the nitrogen atom accompanying a hydrogen atom, such as an amine, amide, amidine or guanidine.

3. Hydrogen-bonding acceptor (HA): the oxygen atom of a carbonyl, sulfone, phosphone, ester, ether, etc.

4. Hydrogen-bonding donor/acceptor (HDA): (A) the oxygen atom of a hydroxyl group and (B) the sulfur atom of a thiol group.

An example of molecular alignment based on chemical property is shown in Figure 1. It is shown that the chemical properties with the same color between two molecules assigned by HNN. To minimize the sum of squares of distance between the 3D coordinates of same chemical properties leads to the real alignment of molecules.

Hopfield Neural Network. The pattern matching problem on chemical properties is a combinatorial optimization problem, and it is said to be an NP-hard problem. If the number of chemical properties is large, calculation cannot be finished within a practical amount of time. To overcome this problem, HNN was used as an optimization tool.

HNN was proposed by Hopfield and Tank,[13] and it was applied to an NP-complete problem such as the Traveling Salesman Problem for proving the capability of optimization.[14] The structure of HNN is shown in Figure 2. Neurons arranged in a single layer have values of 1 or 0. Each neuron is connected to each other in symmetrical manner. First, values of all neurons are randomly initialized. Then, each neuron is iteratively updated according to Hebb's rule

$$\sum_j W_{ij}S_j - \theta_i > 0:\ S_i = 1 \qquad (1)$$

$$\leq 0:\ S_i = 0$$

where θ_i is the threshold value and S_i is the state of the neuron i. W_{ij} is the weight value between the neurons i and j. As a result of this iteration, the energy value of the HNN system is minimized. The value of the energy is calculated by the following energy function:

$$E = -\frac{1}{2}\sum_i\sum_{j\neq i} W_{ij}S_iS_j + \sum_i \theta_i S_i \qquad (2)$$

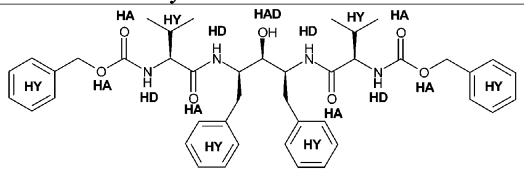
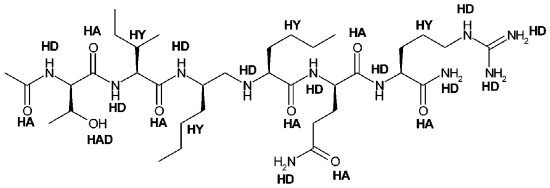
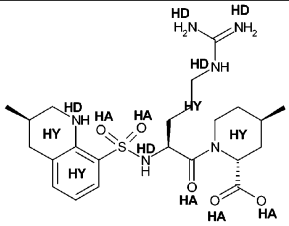
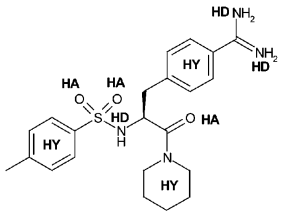
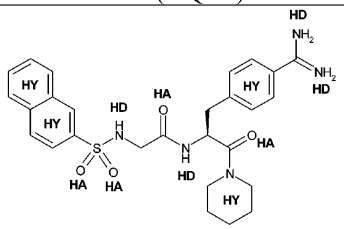
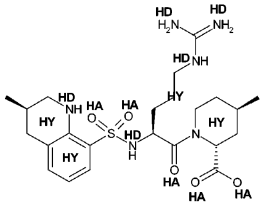
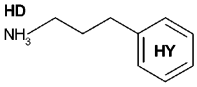
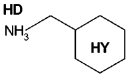
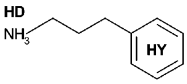
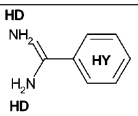
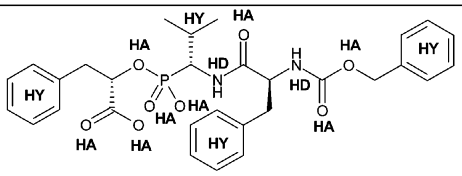
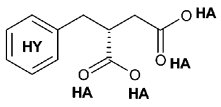
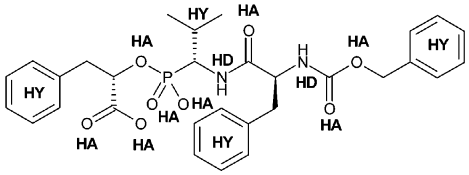
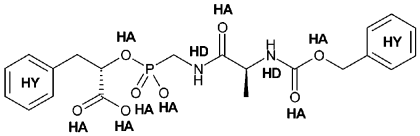
This formula is the general form of the HNN system. More details of the HNN system are described in ref 13. The exact expressions for W_{ij} and θ_i in this alignment problem are shown in ref 15.

Corresponding of Chemical Property. An example of the correspondence matrix of chemical properties is shown in Figure 3. In this example, the numbers of chemical properties assigned on molecule A and molecule B are five and four, respectively. Each chemical property was represented by one of the four kinds of symbols (HY, HD, HA, HDA). Neurons are arranged in 5 * 4 matrix in the HNN system. Neurons are updated iteratively by the Hebbian rule, and the energy value of the HNN system is minimized. When the optimization process is successfully completed, it is expected that the molecular alignment such as shown in Figure 3 would be obtained. The state of the neuron represents a relationship between two properties. For example, if the state of the neuron (A2, B1) is 1, the chemical property 2 of the molecule A is linked to the chemical property 1 of the molecule B. The squared error of distance between these properties is minimized by translation and rotation of the molecule for producing the real molecular alignment.

To achieve real alignment of molecules, the appropriate energy function is necessary. The energy function used in this study is described below.

The energy function consists of five terms.

$$\text{Energy} = \sum_i^5 T_i \qquad (3)$$

Enzyme name	PDB entry codes and inhibitor names
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)

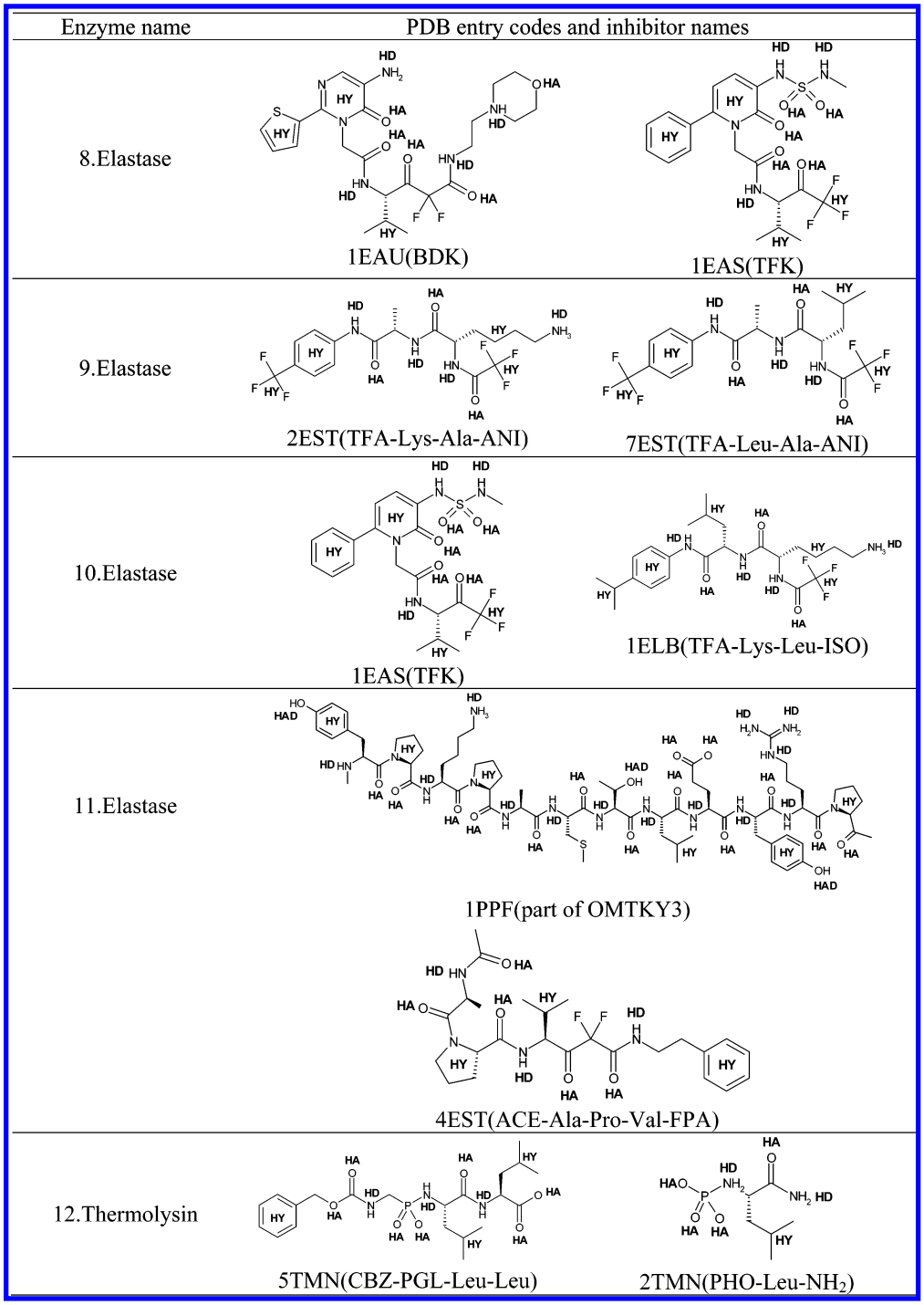


Figure 4. Chemical structures and chemical properties of 12 pairs of inhibitors.

Table 1. Scores between Chemical Properties

	HY	HD	HA	HDA
HY	−3	Inf	Inf	Inf
HD	Inf	−2	Inf	−1
HA	Inf	Inf	−2	−1
HDA	Inf	−1	−1	−1

The first two terms are

$$T_1 + T_2 = \sum_{iI}^{mn}\sum_{jJ}^{mn} S_{iI}S_{jJ}[A\delta_{IJ}(1-\delta_{ij}) + B\delta_{ij}(1-\delta_{IJ})] \quad (4)$$

where m and n are the numbers of chemical properties assigned to molecules A and B. S_{iI} is the state of the neuron placed (i, I). δ_{ij} is the Kroneker delta function, that returns 1 if i is equal to j and returns 0 otherwise. If there is more than one 1 in each row or column, the penalty value is added to the energy value. As a result of minimization of these terms, each chemical property of the molecule corresponds to at most one chemical property of the other molecule.

The third one

$$T_3 = C[\min(m,n) - \sum_{iI}^{mn} S_{iI}] \quad (5)$$

maximizes the number of correspondence of chemical properties.

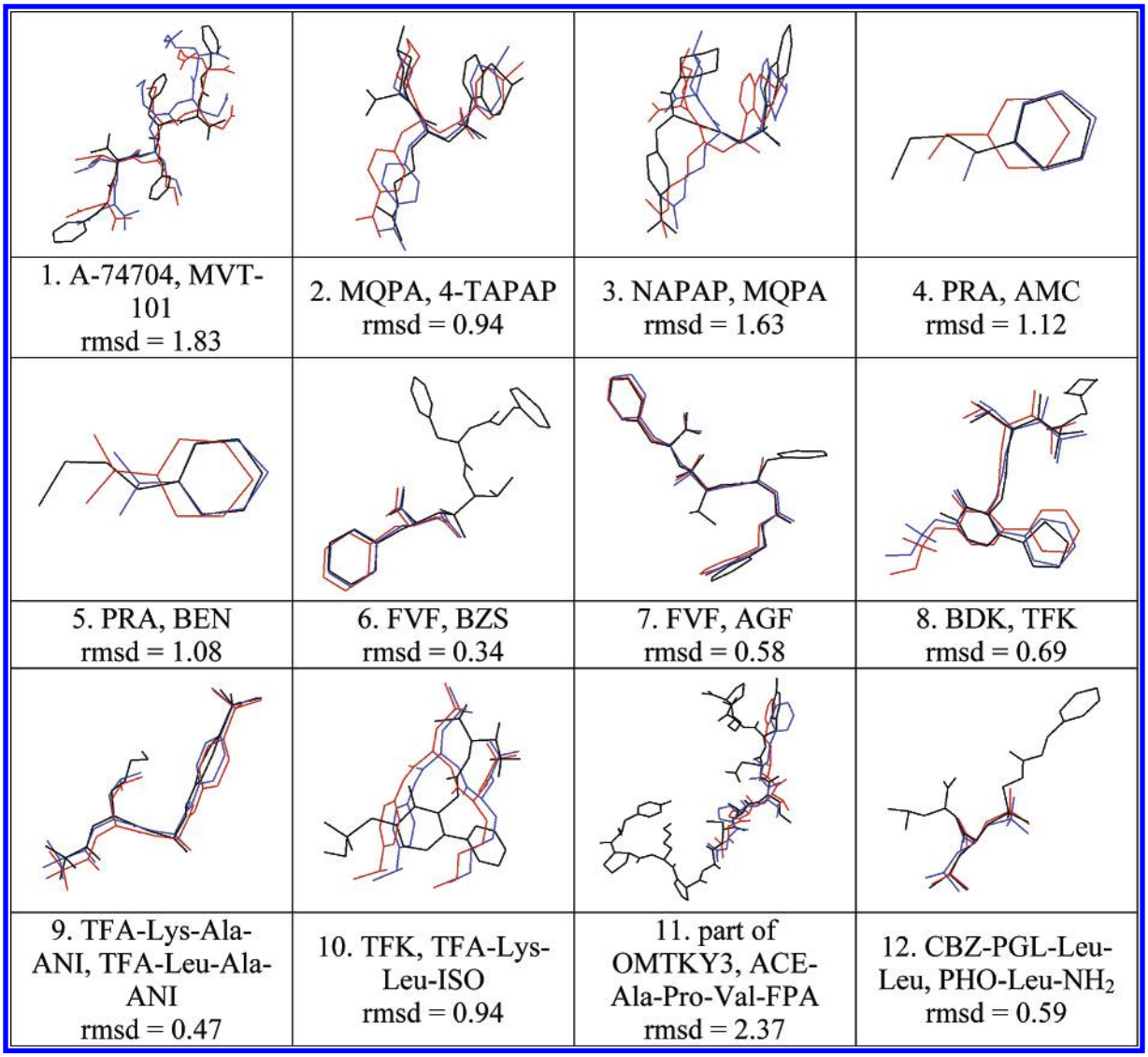


Figure 5. Comparison of HNN results (blue) with X-ray crystal structures (red).

The fourth one

$$T_4 = D[\sum_{iI}^{mn}\sum_{jJ}^{mn} S_{iI}S_{jJ}(1-\delta_{ij})(1-\delta_{IJ})|d_A(ij) - d_B(IJ)|] \quad (6)$$

evaluates the difference in distance between corresponding chemical properties.

The fifth one

$$T_5 = E[\sum_{iI}^{mn} S_{iI}P(i,I)] \quad (7)$$

evaluates the consistency of each kind of chemical property. The definition of the score matrix P is shown in Table 1. If the kind of chemical property is different, the infinite value is added to energy term. These values are cited from ref 5. The definition of energy function described above is based on the previous definition by Doucet et al.,[12] and the chemical meaning of each term is explained in the reference paper.

A, B, C, D, and E are constants used to adjust the weight of each term. These coefficients are determined depending on the precision of molecular alignments or the number of chemical properties. The ratio of coefficients used in this study is decided as follows by empirical trials:

$$A{:}B{:}C{:}D{:}E = 4{:}4{:}8{:}1{:}1 \quad (8)$$

Simulated Annealing. Simulated annealing methods are based on an analogy with thermodynamics and the way that liquids freeze and crystallize. At high temperature the molecules in liquid move freely. If the liquid is cooled slowly they gradually lose mobility and often form a pure crystal that is completely ordered. It corresponds to the global minimum energy state for the system. However, if the system is cooled rapidly it falls into a polycrystalline or amorphous state. It corresponds to a local minimum with higher energy than the global minimum energy state. To avoid convergence to these local minimum, neurons are updated stochastically

$$p = 1/(1 + \exp(\Delta E/T)) \quad (9)$$

where ΔE is the difference of energy and T is the pseudo-temperature. The change is accepted if p is greater than a random number taken in the interval [0 1]. In this study, the temperature was set to 2.0 at the beginning of annealing and decreased lineally.

RESULTS

To validate the proposed alignment method, 12 pairs of enzyme inhibitors cited from the Protein Data Bank (PDB) were aligned. The chemical structures and chemical properties are shown in Figure 4. In this study, bounded conformations obtained from X-ray structures of the protein−ligand complexes have been used. To each pair of inhibitors the simulations of HNN were carried out 100 times with different initial values of the HNN neurons. The result that gave the minimum energy was adopted as the final molecular alignment. These calculations were performed using the program developed in our laboratory with C++ language. This program was running on 1 GHz PC with a Pentium III processor under Windows 2000.

In Figure 5 the result of molecular alignment using HNN is shown. The structure colored black is the first inhibitor in each pair and the structure colored blue is the second inhibitor that is aligned by HNN. The structure colored red is the second inhibitor in the X-ray crystal structure. The values of rmsd represent the root-mean-squared deviation of distances between all heavy atoms of two inhibitors derived from HNN and X-ray. HNN could successfully reproduce the real molecular alignments obtained from X-ray crystallography. The values of rmsd ranged from 0.34 to 2.37.

In the fourth (PRA, AMC) and fifth (PRA, BEN) pairs, the results were slightly worse. The reason for these bad results is derived from the number of chemical properties. The number of chemical properties assigned on BEN is three, and the number on PRA and AMC is only two. It is suggested that the presented method may not work well in such special situations.

DISCUSSION

In this paper, the novel method for molecular alignment using the Hopfield Neural Network is introduced. Chemical properties are assigned to each molecule. Chemical properties between two molecules correspond to each other by HNN. To demonstrate the utility of this method, HNN was applied to 12 pairs of enzyme inhibitors. As a result, our method could successfully reproduce the real molecular alignments obtained from X-ray crystallography. Because the purpose of this paper is whether our method can reproduce real molecular alignment, bounded conformations obtained from X-ray structures have been used. This simplifies the conformation/alignment problem and focuses on the performance of alignment. However, in the case that there is no 3D information of protein−ligand complex, conformational analysis is necessary to estimate bioactive conformation of ligand molecules. In the subsequent paper,[16] our method was applied to the real 3D-QSAR data. Two data sets (Human epidermal growth factor receptor-2 inhibitors and cyclooxygenase-2 inhibitors) were investigated to validate our method, and predictive 3D-QSAR models were successfully obtained.

In many cases, the molecular alignment of the molecules is a very time-consuming task. In our method, owing to the small number of chemical properties and high performance of HNN, the accurate alignment can be obtained quickly. Indeed, in this study the calculation time for 100 simulations of Simulated Annealing was a few seconds or less with a Pentium III processor of 1 GHz. Therefore, it is shown that HNN is a very effective method that solves this kind of combinatorial optimization problem.

In this study, only four properties (hydrophobic group, hydrogen bonding donor, hydrogen bonding acceptor, hydrogen bonding donor/acceptor) were used as a chemical property. Of course, other kinds of chemical properties can be used in this framework. Moreover, it is possible to customize HNN easily by changing the form of energy function.

REFERENCES AND NOTES

(1) Hansch, C.; Fujita, T. ρ-σ-π-Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
(2) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA) 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(3) Arakawa, M.; Hasegawa, K.; Funatsu, K. Selection of Bioactive Conformations and Alignment Rules by 4way PLS Analysis. *J. Comput. Aided Chem.* **2000**, *1*, 1−7.
(4) Hasegawa, K.; Arakawa, M.; Funatsu, K. Simultaneous determination of bioactive conformations and alignment rules by multi-way PLS modeling. *Comput. Chem.* In press.
(5) Iwase, K.; Hirono, S. Estimation of active conformations of drugs by a new molecular superposing procedure. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 499−512.
(6) Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schidt, J. M. Flexible Alignment of Small Molecules. *J. Med. Chem.* **2001**, *44*, 1483−1490.
(7) Feher, M.; Schmidt, J. M. Multiple Flexible Alignment with SEAL: A Study of Molecules Acting on the Colchicine Binding Site. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 495−502.
(8) Mills, J. E. J.; de Esch, I. J. P.; Perkins, T. D. J.; Dean, P. M. SLATE: A method for the superposition of flexible ligands. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 81−96.
(9) Lemmen, C.; Lengauer, T.; Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.
(10) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C.; Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509−10524.
(11) Guccione, S.; Doweyko, A. M.; Chen, H.; Barretta, G. U.; Balzano, F.; 3D-QSAR using 'Multiconformer' alignment: The use of HASL in the analysis of 5-HT_{1A} thienopyrimidinone ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 647−657.
(12) Doucet, J. P.; Panaye, A.; 3D Structural information: from property prediction to substructure recognition with neural networks. *SAR QSAR Environ. Res.* **1998**, *8*, 249−272.
(13) Hopfield, J. J.; Tank, D. W. Neural Computation of Decisions in Optimization Problems. *Biol. Cybern.* **1985**, *52*, 141−152.
(14) Martin-Valdivia, M.; Ruiz-Sepulveda, A.; Triguero-Ruiz, F. Improve local minima of Hopfiled networks with augmented Lagrange multipliers for large scale TSPs. *Neural Networks* **2000**, *13*, 283−285.
(15) The values of W_{ij} and θ_i are deduced from eq 2 and 3 by comparing the coefficient of each term. From complex mathematical manipulation, the coefficient W_{ij} of $-1/2\sum_i\sum_{j\neq i}S_iS_j$ and the coefficient θ_i of $\sum_i S_i$ can be expressed by simple equation with five empirical parameters (A, B, C, D, and E). The fully mathematical derivation is described in ref 12. $W_{ij} = -2A\,\delta_{IJ}(1-\delta_{ij}) - 2B\,\delta_{ij}(1-\delta_{IJ}) - 2D(1-\delta_{ij})(1-\delta_{IJ})|d_A(ij) - d_B(IJ)|$; $\theta_i = -C + EP(i, I)$.
(16) Arakawa, M.; Hasegawa, K.; Funatsu, K. Application of novel molecular alignment method using Hopfield Neural Network to 3D-QSAR. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1396−1402.

CI030001L