VISCANA: Visualized Cluster Analysis of Protein−Ligand Interaction Based on the ab Initio Fragment Molecular Orbital Method for Virtual Ligand Screening

Shinji Amari,*[,†,‡] Masahiro Aizawa,[†,‡] Junwei Zhang,[†,‡] Kaori Fukuzawa,[§] Yuji Mochizuki,[†,‡,||] Yoshio Iwasawa,[‡,||] Kotoko Nakata,[‡,||] Hiroshi Chuman,[⊥] and Tatsuya Nakano[†,‡,#]

Collaborative Research Center of Frontier Simulation Software for Industrial Science, Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan, Revolutionary Simulation Software (RSS21) Project, Collaborative Research Center of Frontier Simulation Software for Industrial Science, Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan, Mizuho Information & Research Institute, Inc., 2-3 Kanda-Nishikicho, Chiyoda-ku, Tokyo 101-8443, Japan, AdvanceSoft Corporation, Center for Collaborative Research, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan, Institute of Health Biosciences, University of Tokushima, 1-78-1 Shoumachi, Tokushima 770-8505, Japan, and Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, Kamiyoga 1-18-1, Setagaya-ku, Tokyo 158-8501, Japan

Received June 24, 2005

We have developed a visualized cluster analysis of protein−ligand interaction (VISCANA) that analyzes the pattern of the interaction of the receptor and ligand on the basis of quantum theory for virtual ligand screening. Kitaura et al. (*Chem. Phys. Lett.* **1999**, *312*, 319−324.) have proposed an ab initio fragment molecular orbital (FMO) method by which large molecules such as proteins can be easily treated with chemical accuracy. In the FMO method, a total energy of the molecule is evaluated by summation of fragment energies and interfragment interaction energies (IFIEs). In this paper, we have proposed a cluster analysis using the dissimilarity that is defined as the squared Euclidean distance between IFIEs of two ligands. Although the result of an ordered table by clustering is still a massive collection of numbers, we combine a clustering method with a graphical representation of the IFIEs by representing each data point with colors that quantitatively and qualitatively reflect the IFIEs. We applied VISCANA to a docking study of pharmacophores of the human estrogen receptor α ligand-binding domain (57 amino acid residues). By using VISCANA, we could classify even structurally different ligands into functionally similar clusters according to the interaction pattern of a ligand and amino acid residues of the receptor protein. In addition, VISCANA could estimate the correct docking conformation by analyzing patterns of the receptor−ligand interactions of some conformations through the docking calculation.

INTRODUCTION

Virtual ligand screening (VLS) methods are becoming increasingly important in drug discovery.[1−3] The VLS method involves two critical issues: docking programs and scoring functions.[4,5] Most current screening methods employ flexible docking tools, such as fast shape matching (DOCK[6]), incremental construction (FlexX[7]), and evolutionary algorithms (AutoDock,[8] GOLD,[9] and GEMDOCK[10]). The performance of scoring functions that calculate the binding (free) energy is often inconsistent across different target proteins and combinations of the docking program and the scoring function. It has been proposed that the combination of multiple scoring functions (consensus scoring) would improve the enrichment of true positives.[11−13] However, the compounds selected by VLS also have many false-positive docking conformations, and further verification methods are expected.

Recently, Kitaura et al. developed the ab initio fragment molecular orbital (FMO) method,[14−17] which enables the ab initio electronic structure calculation of proteins, and applied it to the human estrogen receptor α ligand-binding domain (hERα LBD; 241 amino acid residues) with an FMO-HF/STO-3G level of theory.[18] In the ab initio FMO method, a molecule or a molecular cluster is divided into *M* fragments, and the MO calculations on the fragments (monomers) and the fragment pairs (dimers) are performed to obtain the total energy that is expressed as a summation of the fragment energies and interfragment interaction energies (IFIEs). IFIEs could be used to analyze the interaction of a ligand with the amino acid residues of a protein and are applied to the scoring function and clustering tools after VLS (Figure 1). Although various clustering methods could be used to usefully organize a table of interaction energies between the amino acid residues of the protein and the ligand, the result of a table ordered by clustering, a still-massive collection of numbers, remains difficult to assimilate. To overcome the problem,

* Corresponding author phone: +81-3-5452-6622; fax: +81-3-5452-6623; e-mail: samari@fsis.iis.u-tokyo.ac.jp.
† Institute of Industrial Science, University of Tokyo.
‡ Revolutionary Simulation Software Project.
§ Mizuho Information & Research Institute, Inc.
|| AdvanceSoft Corporation.
⊥ University of Tokushima.
National Institute of Health Sciences.

10.1021/ci050262q CCC: $33.50
Published on Web 11/12/2005

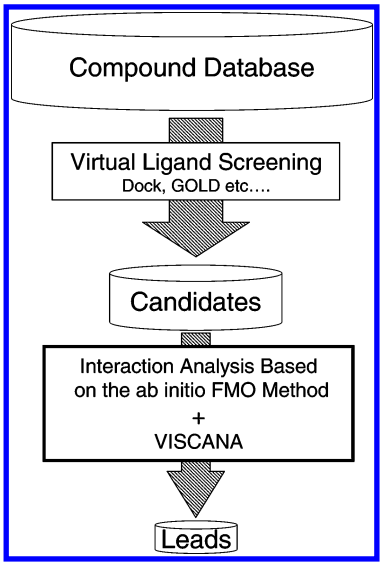

Figure 1. Schematic diagram of the quantum molecular interaction analysis system.

the Biogen group has proposed SIFt (structural interaction fingerprint) for analyzing 3D protein−ligand binding interactions.[19]

In this paper, we have proposed the visualized cluster analysis of protein−ligand interaction (VISCANA) for VLS, by using the dissimilarity that is defined as the squared Euclidean distance between the interaction energy patterns of two ligands and by representing each data point with a color that quantitatively and qualitatively reflects the interaction energy. To show how VISCANA can be used as an effective molecular filter for VLS to select molecules with desirable conformation(s) and interaction patterns with the target protein, we applied VISCANA to a docking study of the pharmacophore of hERα LBD.

METHODS

Visualized Cluster Analysis of Protein−Ligand Interaction Based on the ab Initio FMO Method (VISCANA). A brief description of the ab initio FMO method[14−17] at the Hartree−Fock level is as follows. First, a molecule or a molecular cluster is divided into M fragments, and the ab initio MO calculations on the fragments (monomers) under the electrostatic potential from surrounding $(M - 1)$ monomers V^I are solved repeatedly until all monomer densities become self-consistent. Second, the equations for the dimers are solved under the electrostatic potential from surrounding $(M - 2)$ monomers V^{IJ}. Finally, through the use of the total energies of the monomer E_I and the dimer E_{IJ}, the total energy of the system E is calculated by the following equation:

$$E = \sum_{I>J}^{M} E_{IJ} - (M-2)\sum_{I}^{M} E_I \qquad (1)$$

In the following calculations, the fragmentation is done at α carbon atoms in blocks of one amino acid residue. The ligand molecules were not particularly large and, thus, were treated as a single fragment. The total electron density

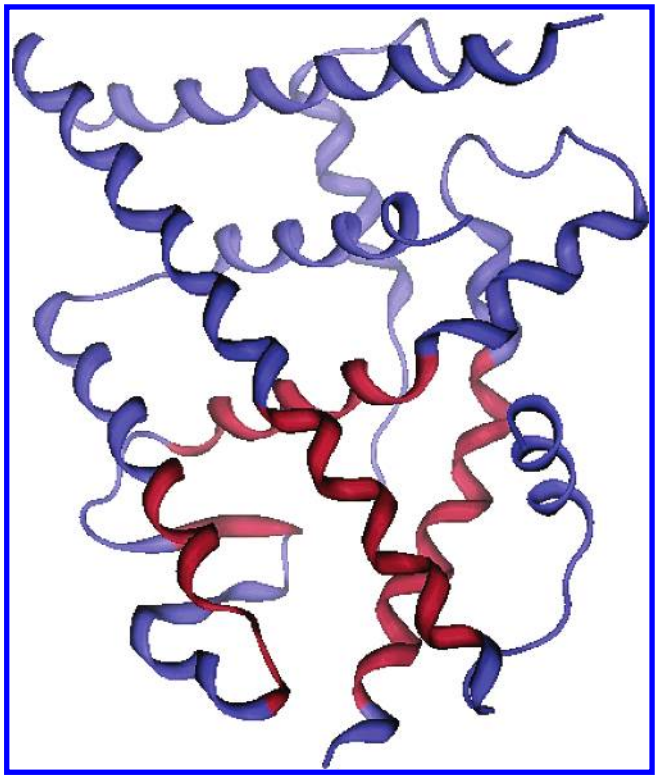

Figure 2. Ribbon display of the hERα LBD. Residues belonging to the pharmacophore are shown in red.

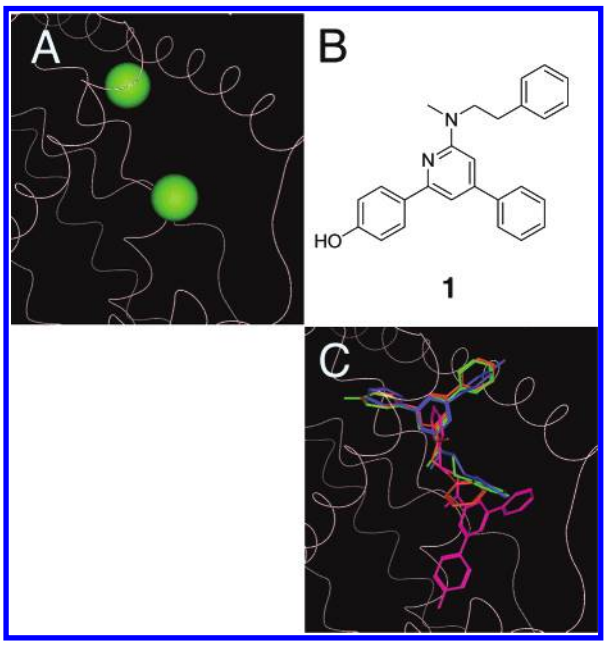

Figure 3. (A) Predicted hERα active site points (ASPs) using PASS. (B) 4-[6-(Methylphenethylamino)-4-phenylpyridin-2-yl]-phenol. (C) Selected conformations of compound **1**. The conformations of GOLD fitness score rankings 1/25, 3/25, 20/25, and 23/25 are shown in pink, green, orange, and blue, respectively.

of the system $\rho(\mathbf{r})$ is also calculated by the following equation:

$$\rho(\mathbf{r}) = \sum_{I>J}^{M} \rho_{IJ}(\mathbf{r}) - (M-2)\sum_{I}^{M} \rho_I(\mathbf{r}) \qquad (2)$$

where $\rho_I(\mathbf{r})$ and $\rho_{IJ}(\mathbf{r})$ are the monomer and dimer electron density, respectively. In this paper, we use the Mulliken population approximation to evaluate net atomic charges.

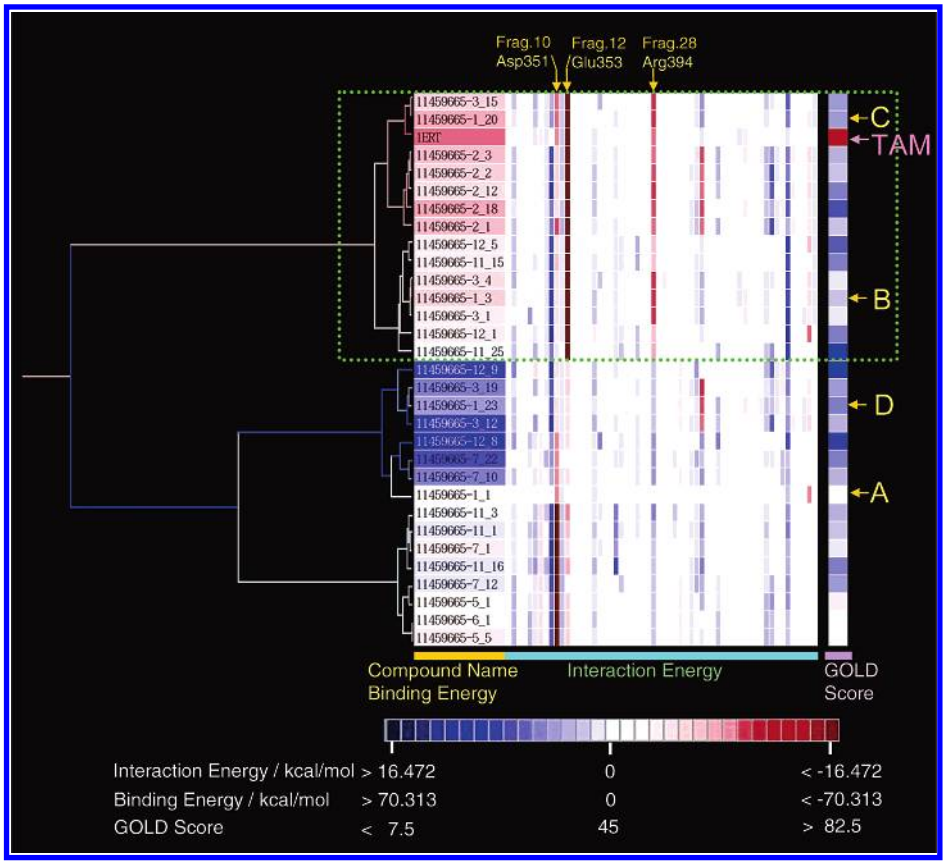

Figure 4. Hierarchical clustering by the interaction patterns of the hERα LBD and antagonists. The binding energies and IFIEs are on a scale of −15 (16.5 kcal/mol; blue) to +15 (−16.5 kcal/mol; red). The important fragments for the binding of the antagonist to hERα are fragment 10 (Asp351), fragment 12 (Glu353), and fragment 28 (Arg394). GOLD fitness score rankings of the conformations of compound **1** are shown on the right side of the table (1/25, A; 3/25, B; 20/25, C; and 23/25, D). TAM denotes the IFIEs of 4-hydroxytamoxifen. The color coding in the compound name column shows the binding energy.

We define IFIE as follows:

$$\Delta E_{IJ} = (E'_{IJ} - E'_I - E'_J) + \mathrm{Tr}(\Delta \mathbf{P}^{IJ}\mathbf{V}^{IJ}) \qquad (3)$$

where $\Delta \mathbf{P}^{IJ}$ is a difference density matrix, $\mathbf{V}^{IJ}$ is the environmental electrostatic potential for dimer IJ, and E'_I and E'_{IJ} are the monomer and dimer energy without an environmental electrostatic potential, respectively. From ΔE_{IJ}, E is calculated by

$$E = \sum_{I>J}^{M} \Delta E_{IJ} + \sum_{I}^{M} E'_I \qquad (4)$$

Without a loss of generality, the dissimilarity of IFIEs between the two compounds I and J (d_{IJ}) is expressed by the following formula

$$d_{IJ} = \sum_{K=1}^{N} (\Delta E_{IK} - \Delta E_{JK})^2 \qquad (5)$$

where N is the number of amino acid residues of the target protein, I and J are fragment indices of the ligand compounds, and K is the fragment index of the residues. VISCANA could be applied to not only the FMO method but also any molecular interaction system which can provide interaction energies or other properties of interest such as charge distribution. For example, we could also define ΔE_{IK} as the ligand−amino acid residue interaction energy by using the energy of decomposition of nonbonding interaction energies (electrostatic and van der Waals) with the classical force field. The receptor−ligand interaction energy table is defined as follows: where L is the number of ligands and N

$$\begin{pmatrix} \Delta E_{\text{ligand}\,1,\,1} & \Delta E_{\text{ligand}\,1,\,2} & \cdots & \Delta E_{\text{ligand}\,1,\,N} \\ \Delta E_{\text{ligand}\,2,\,1} & \Delta E_{\text{ligand}\,2,\,2} & \cdots & \Delta E_{\text{ligand}\,2,\,N} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta E_{\text{ligand}\,L,\,1} & \Delta E_{\text{ligand}\,L,\,2} & \cdots & \Delta E_{\text{ligand}\,L,\,N} \end{pmatrix} \qquad (6)$$

is that of amino acid residues. The IFIE table is visualized with a color scale from −15 (16.5 kcal/mol; blue) to +15 (−16.5 kcal/mol; red). The order of ligand compounds was sorted by a hierarchical clustering procedure. The process of forming a hierarchical cluster was visualized as a dendrogram. The cluster analysis program was coded in Perl. We compared seven clustering methods:[20,21] the furthest-neighbor, nearest-neighbor, group average, centroid, median, Ward, and flexible methods. In VISCANA, we used the standard hierarchical clustering procedures; however, the novelty of VISCANA is in evaluating the dissimilarity that is defined as the distance between the amino acid residue−ligand interaction energy patterns of two ligands.

Computational Detail. Receptor structures were obtained from the Protein Data Bank (PDB). Protein structures used for the calculations were taken from the X-ray structure of the estrogen receptor with the agonist 17β-estradiol (PDB ID: 1ERE) and the antagonist 4-hydroxytamoxifen (PDB ID: 3ERT). The missing atoms and residues in a PDB file were complemented with the rotamer search[22] function of

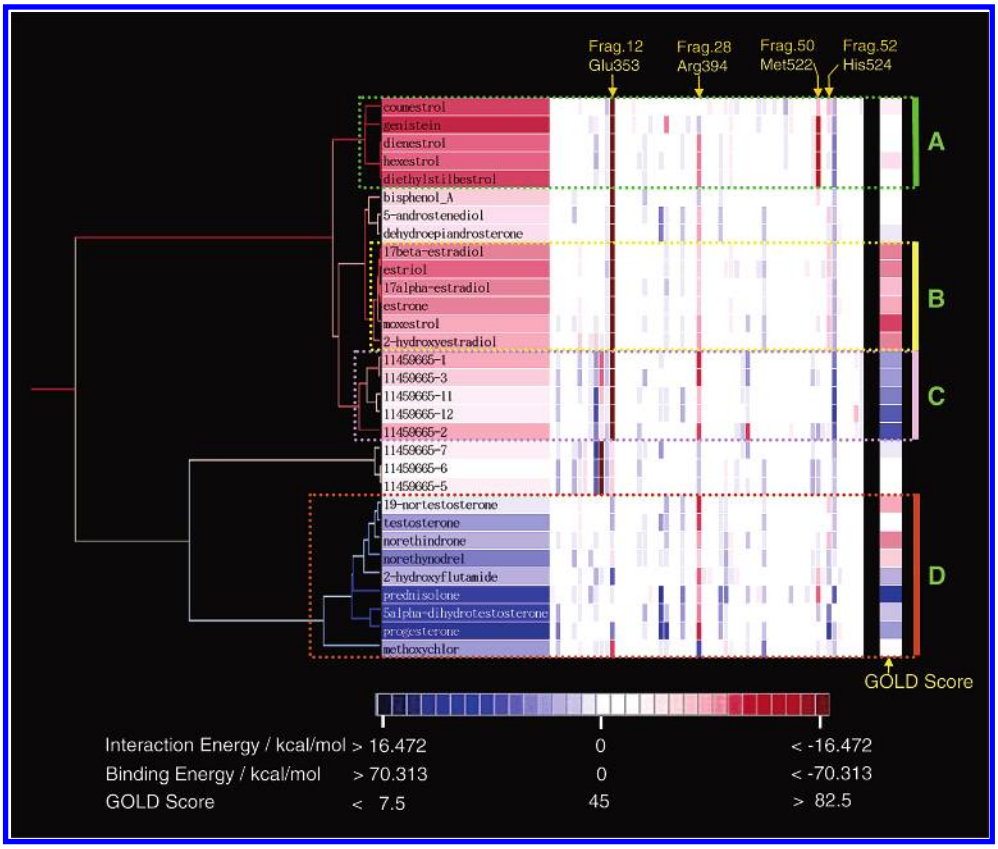

Figure 5. Hierarchical clustering by the interaction patterns of the hERα LBD and 31 compounds from the *Ki*Bank. The binding energies and IFIEs are on a scale of −15 (16.5 kcal/mol; blue) to +15 (−16.5 kcal/mol; red). The important fragments for the binding of the compounds to hERα are fragment 12 (Glu353), fragment 28 (Arg394), fragment 50 (Met 522), and fragment 52 (His524). The color coding in the compound name column shows the binding energy. The right column shows the GOLD fitness score.

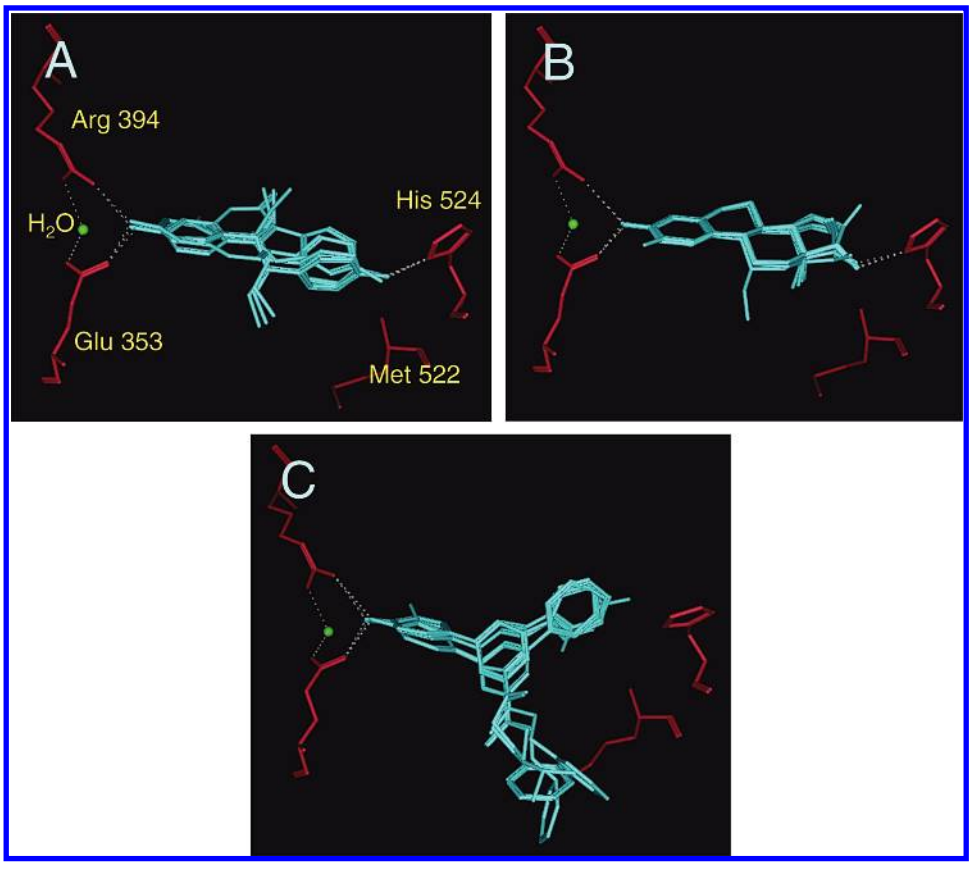

Figure 6. Superposition of the ligands (shown in cyan stick models) in the three clusters of Figure 5.

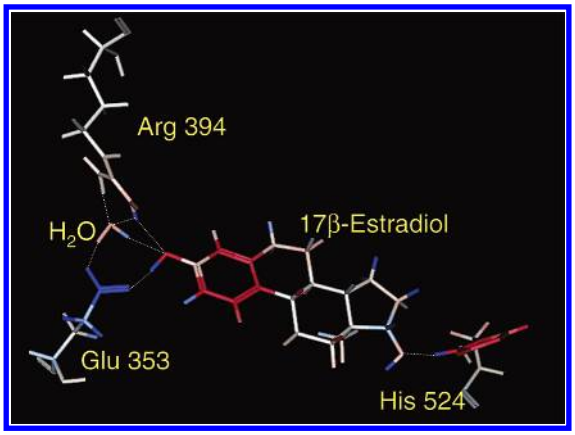

Figure 7. Differences in the net charges between the hERα and 17β-estradiol complex and the individual component molecules (blue, increasing net charge; red, decreasing net charge). The differences of atomic net charge are between −0.149 *e* and +0.104 *e*.

MOE.[23] Hydrogen atoms were added by Reduce.[24] Structures of 31 ligand compounds of hERα with experimental affinity data were taken from the *Ki*Bank,[25,26] and 1135 compounds whose molecular weights are 250 but no more than 300 were chosen from the Kyoto Encyclopedia of Genes and Genomes (KEGG) LIGAND database.[27]

The docked conformations of the ligand to the hERα LBD were predicted using GOLD, version 2.1. The calculation area for an agonist binding pocket was set up in a radius of 10 Å from the center of the 17β-estradiol of the hER crystal structure (1ERE), while that for an antagonist was in a radius of 15 Å from the center of the 4-hydroxytamoxifen of the ER crystal structure (3ERT). MMFF94x (modified MMFF94s)[28] optimized the ligand structure and the positions for hydrogen atoms and a water molecule, which exist less than 4.5 Å from the ligand in the docking conformation predicted by MOE.

The pharmacophore of the hERα LBD (57 amino acid residues; Figure 2) was used for the ab initio FMO calculations at the HF/STO-3G level of theory, and then, the fragmentation was done at α carbon atoms in blocks of one amino acid residue. Binding energies were calculated for the acquired docked conformations by the FMO method. The ligand binding energy ΔE_{ligand} is expressed as the following formula:

$$\Delta E_{\text{ligand}} = E_{\text{complex}} - (E_{\text{receptor}} + E_{\text{ligand}}) \qquad (7)$$

where E_{complex}, E_{receptor}, and E_{ligand} are total energies of the receptor−ligand complex, receptor protein, and ligand compound, respectively. We made calculations with the PC cluster (Intel dual Xeon 3.06 GHz, memory 4 GB, 16 nodes) using the ABINIT-MP program.[16,17,29]

RESULTS AND DISCUSSION

We applied VISCANA to analyze the results of a docking study of nine synthesized antagonists, eight 2-amino-4,6-diarylpyridines[30] and 4-hydroxytamoxifen, to hERα. Two binding site candidates were obtained as a result of the binding site prediction by using PASS[31] (Figure 3A). For example, 25 conformations were obtained as the docking simulation result of 4-[6-(methylphenethylamino)-4-phenylpyridin-2-yl]-phenol[30] (compound **1** in Figure 3B) and clustered by Tanimoto similarity metrics[30] with Type Atom Triangle fingerprints.[33] The conformations of compound **1** were classified into four clusters, and then, the conformation that had the highest GOLD fitness score in each cluster was selected for VISCANA (Figure 3C). The docked conformations of the other seven 2-amino-4,6-diarylpyridines and 4-hydroxytamoxifen used for VISCANA were selected in the same manner as compound **1**. The FMO calculations were performed for the 31 selected conformations. We applied a

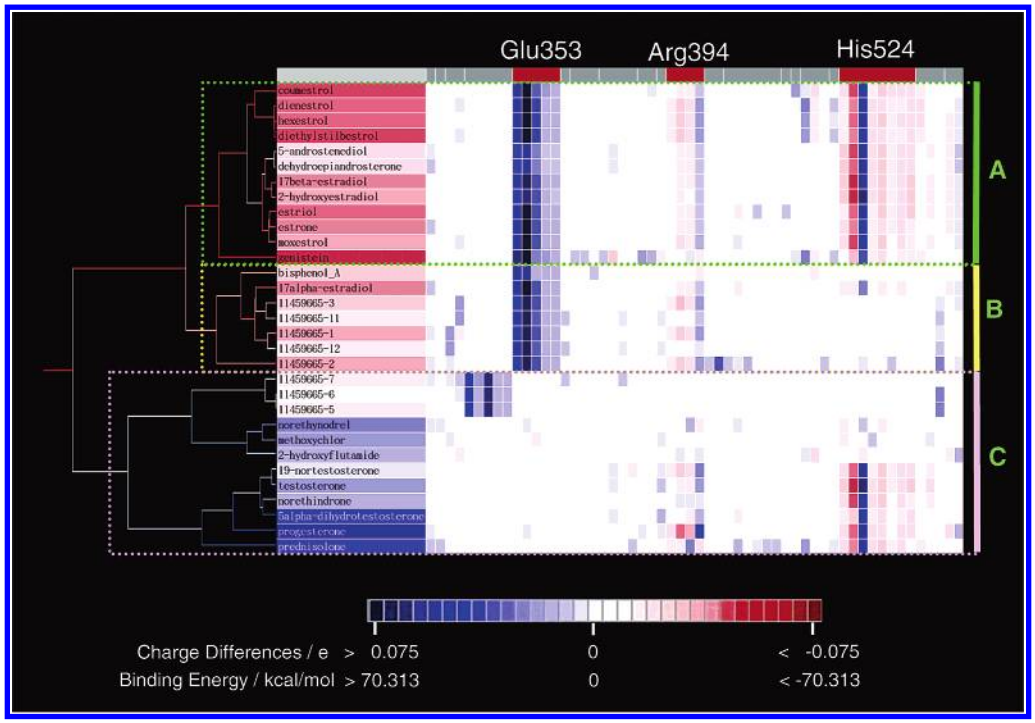

Figure 8. VISCANA using the differences in the net charges between the hERα−ligand complex and the individual component molecules. The table elements are on a scale of −15 (0.075 *e*; blue) to +15 (−0.075 *e*; red). The color coding in the compound name column shows the binding energy.

pairwise furthest-linkage cluster analysis to the IFIEs with ab initio FMO calculations. VISCANA is a form of hierarchical clustering, and the relationships among the conformations are represented by a tree (dendrogram) whose branch lengths reflect the degree of dissimilarity that is defined by the receptor−ligand interaction patterns (eq 5) between the fused clusters. We compared seven clustering methods—the furthest-neighbor, nearest-neighbor, group average, centroid, median, Ward, and flexible methods—and we found that the furthest-neighbor and flexible methods could clearly make up the cluster. However, the nearest-neighbor method could not provide clear-cut clusters for ligand−receptor interaction patterns (data not shown). In the following cluster analysis, we adopted the furthest-neighbor method, which used the dissimilarity of the furthest pair of ligands in a pair of fused clusters. The ordered interaction table (eq 6) was displayed graphically with a representation of the dendrogram to indicate the relationships among the conformations (Figure 4). As shown in Figure 4, the conformations were classified into two clusters with high and low binding energies. The feature of a cluster with a high binding energy (green dashed line in Figure 4) showed a strong interaction with a ligand in Glu353 (fragment 12) and Arg394 (fragment 28), while that of a cluster with a low binding energy showed the interaction between ligands and these residues as a weak attraction or repulsion. The importance of hydrogen bonds among Glu353, Arg394, and the ligand is shown by the experiments,[34] and the results of doing FMO calculations would be consistent with those from the experimental measurement. Furthermore, the feature of an antagonist-type ligand is the hydrogen bond with Asp351,[34] which we could verify from the calculation results. As shown by the cluster with the higher binding energies that contains 4-hydroxytamoxifen in Figure 4, Glu353 is crucial for binding the compounds to hERα, while Arg394 and Asp351 (fragment 10) are substantially important. The GOLD fitness score rankings of the conformations of compound **1** are also shown in Figure 4 (1/25, conformation A; 3/25, conformation B; 20/25, conformation C; and 23/25, conformation D). While conformation A has the highest GOLD fitness score, conformation C has the highest binding energy at the FMO-HF/STO-3G level of theory. The interaction pattern of conformation C is almost the same as that of 4-hydroxytamoxifen. These results suggest that conformation C is the most reliable conformation of compound **1**. It is shown that VISCANA could be used as a useful computational filter for reducing the false-positive docked conformations.

The next application of VISCANA was to analyze a docking study of the pharmacophore of the hERα LBD with 31 compounds for which experimental binding affinities were known. As shown in Figure 5, VISCANA classified the docked compounds into three high-affinity clusters—steroidal agonists (cluster B), nonsteroidal agonists (cluster A), and antagonists (cluster C)—and into low-affinity compounds (cluster D). The GOLD fitness scores of the steroidal ligands in cluster B that have high experimental binding affinities were high (55.37−67.10), whereas those of nonsteroidal ligands such as diethylstilbestrol in cluster A that bind to hERα were lower (44.41−51.66). In fact, a steroid-type ligand is selected according to the GOLD fitness score. The binding energies obtained by the FMO calculations show that the interactions of the steroidal and nonsteroidal ligands

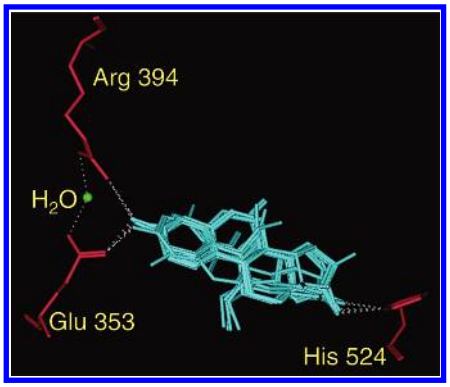

Figure 9. Superposition of the ligands (shown in cyan stick models) in cluster A of Figure 8.

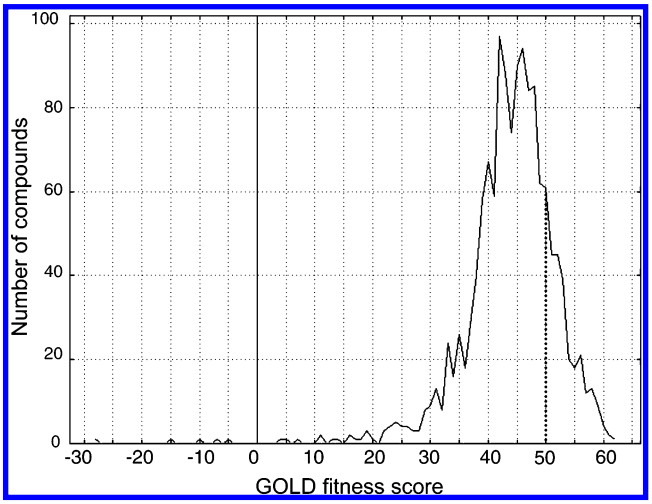

Figure 10. Histogram of the GOLD fitness scores of 1135 compounds for the hERα LBD−ligand docking.

in clusters A and B with hERα were attractive (−45.36 to −26.46 kcal/mol). The IFIEs between each fragment—fragment 12 (Glu353), fragment 28 (Arg394), and fragment 52 (His524)—and a ligand were attractive, and these results correspond to the hydrogen bond of steroidal agonists with Glu353, Arg394, and His524.[34] In the case of nonsteroidal ligands in cluster A such as diethylstilbestrol, the interaction between a ligand and Met522 (fragment 50) is also crucial. The GOLD fitness scores of the antagonists in cluster C were lower (26.91−34.95), while the total binding energies in cluster C using the FMO calculations showed weak attractions. In cluster D, 19-nortestosterone and norethindrone, which weakly bind to hERα, had higher GOLD fitness scores (57.75 and 61.45, respectively), while the binding energies of 19-nortestosterone and norethindrone were repulsive (8.13 and 17.97 kcal/mol, respectively). The receptor−ligand interaction patterns of the compounds in cluster D whose binding energies, determined by the FMO calculations, were repulsive showed that the interaction with Arg394 was attractive and that of Glu353 was repulsive or weakly attractive, except for the case of methoxychlor. The interaction patterns in cluster D were quite different from the patterns in clusters A, B, and C. Figure 6 shows a superposition of the docked conformations in the three clusters (A, B, and C), and the interaction patterns of the ligand binding to hERα directly reflect the ligand structures. These results suggest that VISCANA is a useful tool for separating ligands with activity (cluster A, B, and C) from those without activity (cluster D).

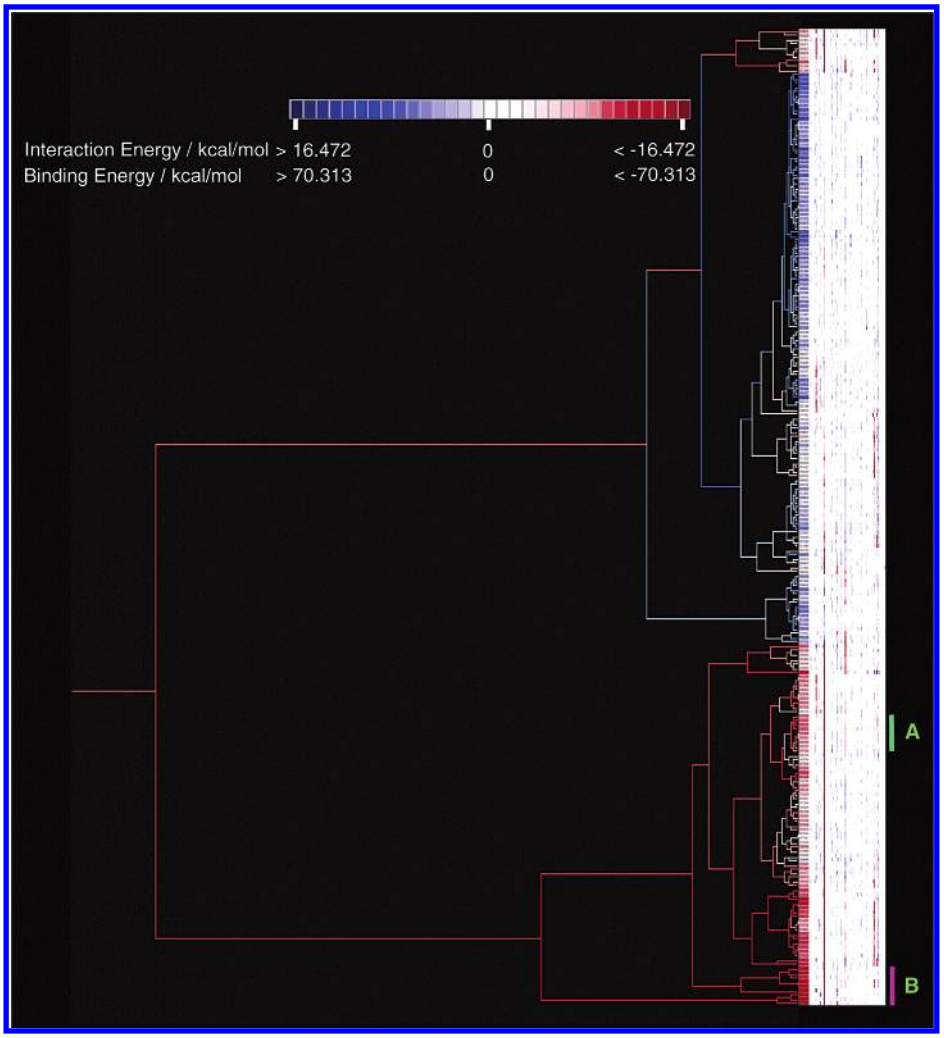

Figure 11. VISCANA for the hERα LBD−ligand docking study with 224 random compounds from the KEGG LIGAND database.

The FMO method has the advantage of describing the charge-transfer between a receptor and a ligand in comparison to a conventional force field method using fixed atomic charges. The difference in the net charges between the hERα and 17β-estradiol complex and the individual component molecules (ΔQ) on a scale of 0.104 *e* (blue) to −0.149 *e* (red) is shown for the important residues (Glu353, Arg394, and His524), the water molecule, and the ligand (17β-estradiol) in Figure 7. In the hERα and 17β-estradiol complex, the net charge difference of the ligand was negative by −0.145 *e*; that of Glu353 was positive by 0.229 *e*, and that of His524 was negative by −0.080 *e*. The changes in the total net charges of Arg394 and the water molecule are smaller than those of Glu353 and His524. In Figure 7, the color-scaled net charge difference shows that the phenolic hydroxyl of the A ring makes direct hydrogen bonds with the carboxylate of Glu353, the guanidinium group of Arg394, and the water molecule and that the 17β-hydroxyl of the D ring makes a single hydrogen bond with His524. Figure 8 shows VISCANA with the difference of the net charges ΔQ instead of IFIEs for the 31 compounds to the pharmacophore of hERα. Using ΔQ_{IJ}, d_{IJ} is calculated by

$$d_{IJ} = \sum_{K=1}^{N} (\Delta Q_{IK} - \Delta Q_{JK})^2 \qquad (8)$$

When ΔQ_{IJ} is used, the receptor−ligand charge difference table is defined as follows

$$\begin{pmatrix} \Delta Q_{\text{ligand}\,1,1} & \Delta Q_{\text{ligand}\,1,2} & \cdots & \Delta Q_{\text{ligand}\,1,N} \\ \Delta Q_{\text{ligand}\,2,1} & \Delta Q_{\text{ligand}\,2,2} & \cdots & \Delta Q_{\text{ligand}\,2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta Q_{\text{ligand}\,L,1} & \Delta Q_{\text{ligand}\,L,2} & \cdots & \Delta Q_{\text{ligand}\,L,N} \end{pmatrix} \qquad (9)$$

The elements of the color-scale table are such that one or more absolute values of the charge difference of an atom are larger than the threshold value 0.01 *e*. In Figure 8, VISCANA, using ΔQ, classified the ligands into three clusters: a positive ΔQs on Glu353 and His524 (cluster A), a positive ΔQ on Glu353 (cluster B), and a small ΔQ on Glu353 (cluster C). The ligands in cluster A that contain 17β-estradiol have large binding energies (Figures 8 and 9). Most compounds in cluster A were overlapped with the compounds in clusters A and B of Figure 5. These results suggest that the charge redistribution caused by making hydrogen bonds is very important for receptor−ligand binding.

To demonstrate the applicability of VISCANA, VISCANA was applied to a docking study of random compounds to the hERα LBD. The docking calculations using GOLD were performed for 1135 selected compounds from the KEGG LIGAND database. The phosphate groups of ligand mol-

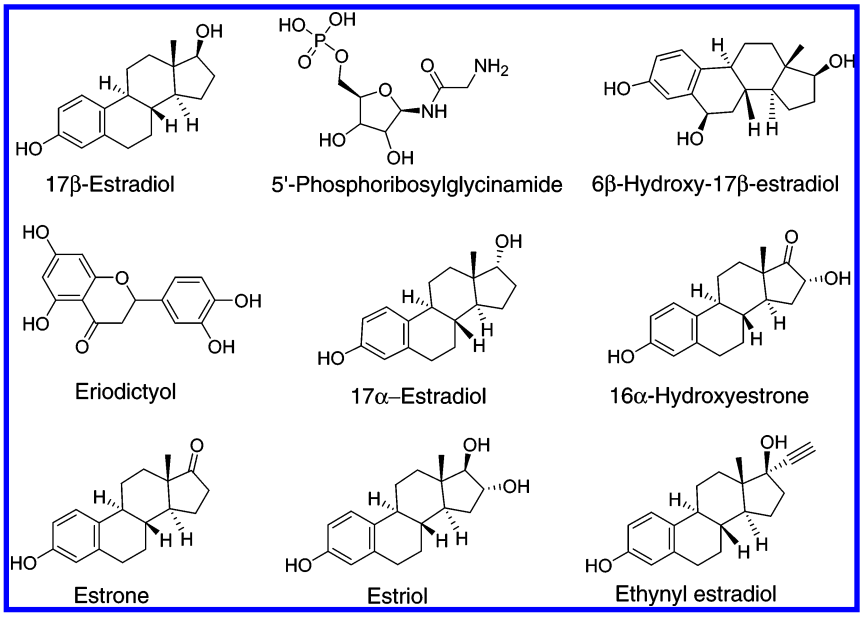

Figure 12. Nine compounds contained in cluster A of Figure 11.

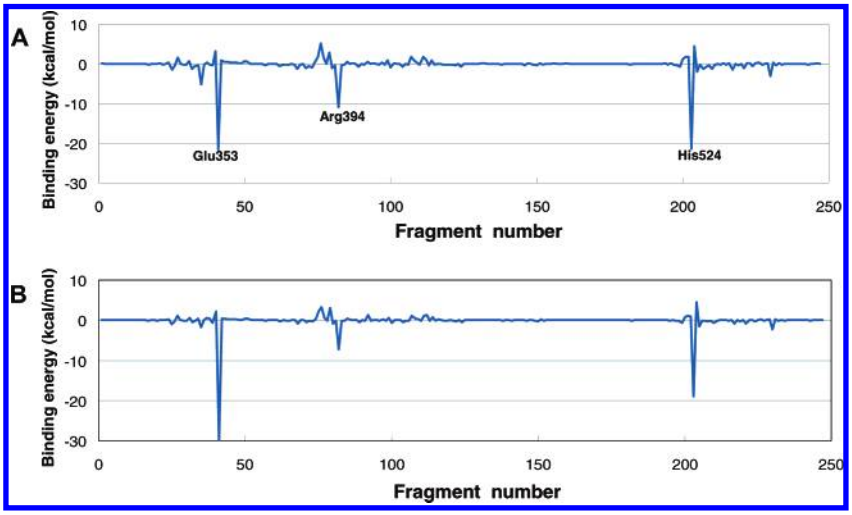

Figure 13. Basis set dependency of IFIEs. (A) FMO-HF/6-31G. (B) FMO-HF/STO-3G.

ecules were treated as protonated. The FMO calculations were performed for 249 compounds having GOLD fitness scores higher than 50 (Figure 10). The FMO calculations were completed for 224 compounds, but those of the remaining 25 compounds were not completed because the self-consistent charge procedure did not converge in the calculations. As shown in Figure 11, VISCANA classified the 224 compounds into two clusters with high and low binding energies. In the cluster with the higher binding energies, the compound group having a high experimental binding affinity was contained in cluster A. Figure 12 shows nine of the compounds contained in cluster A of Figure 11. Cluster B contains phosphorus-rich compounds that have higher binding energies than the compounds in cluster A. These results show that VISCANA is useful for reducing the size of a chemical library and finding new seed compounds. Our results demonstrate that VISCANA is a useful tool for VLS. The receptor−ligand interaction patterns are important for reducing false-positive compounds, and VISCANA is different from the conventional screening methods, which choose the higher rank of a docking score on the point. Moreover, even if a compound whose docking score or binding energy is low belongs to the same cluster that contains compounds with activity, the compound would be a candidate for a lead compound.

Computation time is one of the most important issues for analysis tools of VLS. In the case of the pharmacophore of the hERα LBD with 17β-estradiol, it is 699.8 s at the FMO-HF/STO-3G level of theory and 3568.5 s at FMO-HF/6-31G when using a PC cluster with Intel Xeon 60 processors. Figure 13 shows IFIEs of the hERα LBD with 17β-estradiol at the FMO-HF/STO-3G and FMO-HF/6-31G levels of theory, and the receptor−ligand binding energies show almost the same profile. Moreover, when the FMO-HF/STO-3G method is used, 100 ligand−protein complex structures were calculated in about 20 h, which shows that the FMO-HF/STO-3G method is appropriate to evaluate IFIEs for VISCANA.

CONCLUSIONS

We have developed a visualized cluster analysis of protein−ligand interaction that carries out a cluster analysis on the basis of IFIEs with the ab initio FMO method

calculations for VLS. VISCANA makes it possible to classify structurally similar ligand molecules that could not bind a receptor, which is different from the conventional clustering method based on ligand structures, through the interaction pattern of a ligand and amino acid residues of the receptor protein as well as the interaction energy of the ligands and the protein. Next, VISCANA could estimate the correct docking conformation by analyzing patterns of the receptor−ligand interactions of some conformations through the docking calculation. In addition, VISCANA provides a powerful tool for assessing the VLS of xenobiotic chemicals prior to biological experimentation. Our results show that ab initio quantum mechanical calculations based on the FMO method are efficient and provide reasonable binding energies and binding patterns of ligand−protein interactions. However, further development of quantum mechanical methods such as the second-order Møller−Plesset perturbation theory based on the FMO method[35,36] will be needed to obtain more reliable descriptions of van der Waals interactions and hydrogen bonds that are important for receptor−ligand binding. It is also necessary to allow optimization of the geometry for entire complexes or pharmacophores, particularly induced-fit complexes. Such projects are in progress. VISCANA is also applicable to interaction analysis by a classical force field and enables high-throughput analysis, but the accuracy of VISCANA is inferior to that of the quantum chemical calculations. The visualization software for molecule and molecular-interaction analyses based on FMO calculations with ABINIT-MP, named BioStation Viewer, was developed by our research group and is freely available at http://www.fsis.iis.u-tokyo.ac.jp/en/. We are scheduled to release a new version of BioStation Viewer, which includes the VISCANA function with a supporting CSV format of the receptor−ligand interaction energy table as well, soon.

ACKNOWLEDGMENT

T.N. is grateful to Dr. Shigeru Koikegami for assistance. The work reported here was supported primarily by the Revolutionary Simulation Software (RSS21) project operated by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and partially by the Core Research for Evolutional Science and Technology (CREST) project of the Japan Science and Technology Agency (JST) and the Toxico-proteomics project fund from the Ministry of Health, Labour and Welfare.

Supporting Information Available: The 31 ligand compounds obtained from *Ki*Bank, and their calculated values. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Walters, P. W.; Stahl, M. T.; Murcko, M. A. Virtual screening − an over view. *Drug Discovery Today* **1998**, *3*, 160−178.

(2) Lyne, P. D. Structure based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.

(3) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439−446.

(4) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(5) Yang, J. M.; Shen, T. W. A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators. *Proteins* **2005**, *59*, 205−220.

(6) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.

(7) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(8) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1999**, *19*, 1639−1662.

(9) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(10) Yang, J. M.; Chen, C. C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* **2004**, *55*, 288−304.

(11) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model* **2002**, *20*, 281−295.

(12) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(13) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.

(14) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. Pair interaction molecular orbital method: an approximate computational method for molecular interactions. *Chem. Phys. Lett.* **1999**, *312*, 319−324.

(15) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.

(16) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: application to polypeptides. *Chem. Phys. Lett.* **2000**, *318*, 614−618.

(17) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayashi, M.; Kitaura, K. Fragment molecular orbital method: use of approximate electrostatic potential. *Chem. Phys. Lett.* **2002**, *351*, 475−480.

(18) Fukuzawa, K.; Kitaura, K.; Uebayasi, M.; Nakata, K.; Kaminuma, T.; Nakano, T. Ab initio quantum mechanical study of the binding energies of human estrogen receptor alpha with its ligands: an application of fragment molecular orbital method. *J. Comput. Chem.* **2005**, *26*, 1−10.

(19) Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein−Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47*, 337−344.

(20) Lance, G. N.; Wiliams, W. T. A general theory of classificatory sorting strategies 1. Hierarchical systems. *Comput. J.* **1967**, *9*, 373−380.

(21) Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236−244.

(22) Bower, M. J.; Cohen, F. E.; Dunbrack, R. L., Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* **1997**, *267*, 1268−1282.

(23) *MOE* (*Molecular Operating Environment*), version 2004.04; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2004.

(24) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735−1747.

(25) Zhang, J.; Aizawa, M.; Amari, S.; Iwasawa, Y.; Nakano, T.; Nakata, K. Development of KiBank, a database supporting structure-based drug design. *Comput. Biol. Chem.* **2004**, *28*, 401−407.

(26) Aizawa, M.; Onodera, K.; Zhang, J.; Amari, S.; Iwasawa, Y.; Nakata, K.; Nakano, T. KiBank: a database for computer-aided drug design based on protein-chemical interaction analysis. *Yakugaku Zasshi* **2004**, *124*, 613−619.

(27) Goto, S.; Nishioka, T.; Kanehisa, M. LIGAND: chemical database for enzyme reactions. *Bioinformatics* **1998**, *14*, 591−599.

(28) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.

(29) *ABINIT-MP*. Available at http://www.fsis.iis.u-tokyo.ac.jp/en/result/software/.

(30) Henke, B. R.; Drewry, D. H.; Jones, S. A.; Stewart, E. L.; Weaver, S. L.; Wiethe, R. W. 2-Amino-4,6-diarylpyridines as novel ligands for the estrogen receptor. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1939−1942.

(31) Brady, G. P., Jr.; Stouten, P. F. W. Fast prediction and visualization of protein binding pockets with PASS, *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383−401.

(32) Willett, P. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(33) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(34) 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.
(35) Mochizuki, Y.; Nakano, T.; Koikegami, S.; Tanimori, S.; Abe, Y.; Nagashima, U.; Kitaura, K. A parallelized integral-direct MP2 method with fragment molecular orbital scheme. *Theor. Chem. Acc.* **2004**, *112*, 442−452.
(36) Mochizuki, Y.; Koikegami, S.; Nakano, T.; Amari, S.; Kitaura, K. Large scale MP2 calculations with fragment molecular orbital scheme. *Chem. Phys. Lett.* **2004**, *396*, 473−479.

CI050262Q