**Regular Article**

*Highlighted Paper selected by Editor-in-Chief*

**Prediction of Ligand Binding Affinity to Target Proteins by Molecular Mechanics Theoretical Calculation**

Hideyoshi Fuji, Fei Qi, Liang Qu, Yoshihisa Takaesu, and Tyuji Hoshino*

Graduate School of Pharmaceutical Sciences, Chiba University; 1–8–1 Inohana, Chuo-ku, Chiba 260–8675, Japan.

Received November 24, 2016; accepted February 13, 2017

Accurate estimation of ligand–receptor binding affinity is indispensable for computer-assisted drug discovery and structure-based drug design. Many computational scoring functions for estimating binding affinity have been proposed. Every scoring function reported so far, however, has strengths and weaknesses depending on the chemical properties of ligands and the feature of the binding site of the receptor. Hence, potential functions that can be used for many kinds of target proteins are required. In this work, we developed a software program based on Morse-type potential functions that enables evaluation of binding affinity and geometry optimization. Eight different kinds of proteins were used as test data, and ligand chemicals for which the binding pose to the protein and inhibitory constant are known were selected for evaluation. The calculated binding score and the experimentally measured inhibitory constant showed good compatibilities for six target proteins but poor correlation for one target. These compatibilities were compared with the results obtained by using two other software programs. The comparison suggested that the performance of the software developed in this work is good. Since the software can be handled in a computer facility with a many-core system, the software will be effective for search for an active compound from a chemical database and for assistance in chemical modification of the active compound in the pharmaceutical research field.

**Key words**  binding affinity; potential function; ligand–receptor complex; computer prediction

Three-dimensional structures of many kinds of disease-related proteins have been revealed due to the progress in experimental techniques such as X-ray crystal analysis and NMR spectroscopy and also advances in theoretical methodology for modeling. With the accumulation of protein structures, structure-based drug design (SBDD), a technique to accelerate the process of drug discovery by utilizing structural information of a target protein, has become important. Molecular docking, one of the most popular computer applications in SBDD, enables evaluation of the binding affinity between a ligand and its target protein. Generally, there are two aims of studies using molecular docking. One is structural modeling; i.e., determination of docking pose, and the other is the prediction of activity, i.e., estimation of binding affinity. Several studies have shown that current molecular docking programs can generate experimentally observed conformations of small molecules and reproduce their binding poses in target proteins. However, accurate estimation of the binding affinity of ligand–receptor complex has been difficult compared with the prediction of binding poses.

Scoring functions can be categorized into three groups: force-field-based, empirical, and knowledge-based functions. The force-field-based scoring function is utilized to obtain classical molecular mechanics energy. The function usually quantifies the sum of two energies: ligand–receptor interaction energy and internal ligand energy (such as steric strain induced by binding). The interaction between a ligand and a receptor is often described by using van der Waals and electrostatic energy terms. A drawback of this function is that the energy landscapes associated with force-field potentials are usually rugged and, therefore, energy minimization is required prior to evaluation of the binding score.

The empirical scoring function estimates the binding free energy by summing up interaction terms derived from weighted structural parameters. The weights are obtained by fitting the scoring functions to experimental binding constants for a training set of ligands acting on a specific receptor. The main drawback of empirical scoring functions is that it is unclear whether they are able to predict the binding affinities of ligands for which structures are greatly different from those used in the training set and whether they can be applied to different types of receptors.

The knowledge-based scoring function represents the binding affinity as a sum of ligand–receptor atom pair interactions. These potentials are derived from ligand–receptor complexes with known structures, with the probability distribution of interatomic distances being converted into distance-dependent interaction free energies of ligand–receptor atom pairs using the inverse Boltzmann law. A drawback is that knowledge-based scoring functions are essentially based on information from a limited number of ligand–receptor complex structures.

Although many scoring functions have been proposed in the past two decades, it seems difficult to develop a scoring function that always shows good performance for many kinds of protein families within the formula categorized in any of the above-mentioned conventional functions. Hence, a scoring function that can be applied to a broad range of target proteins is still needed. The requirements for such a function are the capability to identify active compounds from a chemical database and the capability to distinguish a highly active chemical structure from its analogous ones in the same scaffold. Furthermore, accurate estimation of the binding affinity of a ligand–receptor complex is necessary even if the calculation is time-consuming. Calculation time has become less problematic because of the recent progress in computer technology. A combination of two approaches, that is, rapid prediction of the binding mode of a ligand against a target protein and precise, even if time-consuming, estimation of the binding affinity of

*To whom correspondence should be addressed.  e-mail: hoshino@chiba-u.jp

© 2017 The Pharmaceutical Society of Japan
the binding mode, will be a promising methodology.

In this study, we developed a software program named Ori-entation that calculates the binding affinities of ligand–protein complexes. We focused on hydrophobic and hydrophilic interactions because they are the major factors in ligand–receptor interactions. We implemented hydrophobic and hydrophilic potential functions as non-bonded energy terms instead of van der Waals and electrostatic potentials. As bonded terms, the bond, angle, and torsion energies are estimated by the force-field usually incorporated in a program for molecular dynamics simulation.

The accuracy of estimation of ligand–receptor binding affinity was checked by using data from the PDBbind database. Scoring evaluation was carried out for 8 protein targets for which experimental binding constants were known and the binding modes of the ligands were available in the protein data bank (PDB). Next, we assessed our original scoring function in comparison with some widely used scoring functions. Three functions in the GOLD program (Goldscore, Chemscore, and ASP) and two functions in the AutoDock program (AutoDock and AutoDock Vina) were used for comparison. These functions cover the three groups described above. Goldscore, AutoDock and AutoDock Vina scores are force-field-based scoring functions. ChemScore is an empirical scoring function, and ASP is a knowledge-based scoring function. The results of the comparison suggest that the formula used in our scoring function is effective for predicting the binding affinities of ligands to target proteins.

Experimental

Model Preparation of Ligand–Receptor Complexes

X-Ray crystal structures of ligand–receptor complexes, the \( K_i \) values of which were known, were selected from the PDBbind database. PDBbind is a collection of experimentally measured binding affinities such as \( K_i, K_d \) and IC\(_{50}\) for the ligand–protein complexes, the structures of which are available in PDB. The crystal structure, chemical formula of the ligand, and the binding affinity of a complex are referred by PDB entry code. Eight target proteins, acetylcholinesterase (AChE), cyclin-dependent kinase 2 (CDK2), factor Xa (FXa), FK506 binding protein (FKBP), neuraminidase (NEU), penicillopepsin (PCP), plasmapsin II (PlmII), and purine nucleoside phosphorylase (PNPase), were selected from the database. Structure data except for PlmII were downloaded from the PDBbind refined-set. The data for PlmII were downloaded from the PDBbind general-set. Enzyme classification and the number of complexes for the respective targets are shown in Table 1. The PDB codes of the complexes are shown in Table S1 in Supplementary materials. All the ligand chemical structures for the respective targets are shown with their binding affinity in Fig. S1.

Missing heavy atoms and hydrogen atoms of the proteins were generated by using PDB2PQR. Metal ions that were not related to ligand–receptor interaction and all of the crystal water molecules were removed. Hydrogen atoms of the ligands were generated by using SYBYLX.

Details of the Computation for Estimation of Binding Affinity

The potential energy and the force acting on every atom are computed by estimating the intra- and inter-molecular interactions separately. In other words, the potential energy and the force between the atoms inside the protein or those between the atoms inside the ligand are computed as the intra-molecular interaction. In contrast, the atom pair between the ligand and receptor is treated as the inter-molecular interaction. The intra- and inter-molecular interactions are calculated by the following equations.

\[
E = \sum_{\text{bonds}} K_i (r - r_{0i})^2 + \sum_{\text{angles}} K_d (\theta - \theta_{0i})^2 + \sum_{\text{dihedrals}} \left( \frac{1}{2} \frac{\partial V}{\partial \phi} \right) + \sum_{\text{specific}} \left( \frac{A_{ij}}{r_{ij}^2} - B_{ij} \right)
\]

The first three terms are applied for the atoms connected by covalent bonds. These bond, angle, and torsion energy terms are broadly used in software for molecular dynamics simulation. The last two terms represent the electrostatic and van der Waals potential energies. The van der Waals potential energy is given by the Lennard–Jones type-formula. The summation range of the last two terms can be changed by the computational option. If the ligand–receptor complex is surrounded by solvent, the influence of the solvent atoms on the last two terms is usually included.

\[
E = \sum_{\text{phobic}} D_{\text{phobic}}[(1 - e^{-A_{\text{phobic}}(R - R_{\text{cutoff})}})^2 - 1] + \sum_{\text{repulsive}} \frac{D_{\text{rep}}}{(A_{\text{rep}} R)^n} + \sum_{\text{specific}} D_{\text{sp}}[(1 - e^{-A_{\text{specific}}(R - R_{\text{cutoff})}})^2 - 1]
\]

As shown in Eq. 2, hydrophobic and hydrophilic interac-

---

Table 1. Proteins Used for Evaluation of the Computer Program

| Protein name                          | EC number | Enzyme name                      | Number of data |
|---------------------------------------|-----------|----------------------------------|----------------|
| Acetylcholinesterase (AChE)           | 3.1.1.7   | Carboxy ester hydrolase          | 8              |
| Cyclin-dependent kinase 2 (CDK2)      | 2.7.11.22 | Protein-serine/threonine kinase  | 7              |
| Factor Xa (FXa)                       | 3.4.21.6  | Serine endopeptidase             | 18             |
| FK506 binding protein (FKBP)          | 5.2.1.8   | cisis-trans-Isomerase            | 5              |
| Neuraminidase (NEU)                   | 3.2.1.18  | Glycosidase                      | 6              |
| Penicillopepsin (PCP)                 | 3.4.23.20 | Aspartic endopeptidase           | 8              |
| Plasmapsin II (PlmII)                 | 3.4.23.39 | Aspartic endopeptidase           | 4              |
| Purine nucleoside phosphorylase (PNPase) | 2.4.2.1  | Glycosyltranferase               | 9              |

a) Abbreviation is shown in parenthesis.
tions between a ligand and a receptor are described by Morse-type potential functions. The parameters, \( D, A \) and \( R_{\text{eq}} \), change depending on the combination of atom-types of the pair. The parameter values are shown in Table S2. All of the terms can be deactivated by the computational option. It should be emphasized that the summation is relevant to the atom pairs in which one atom of the pair belongs to the ligand and the other belongs to the receptor. The sum of the first and second terms of Eq. 2 is defined as the ligand–receptor binding energy, hereafter referred to as Orientation score, which is used as an index for binding affinity. The third term in Eq. 2 represents the repulsion of atoms. This term is needed to avoid unfavorable atom collision. The fourth term, specific interaction, is effective only if the designated atom pairs are indicated in the input file. The fourth term is seldom activated.

The inclusion of \( \pi-\pi \) interaction is possible through the terms in Eq. 3. Not only the interaction of two aromatic rings but also the interaction between an aromatic ring and any of the CH, NH, OH, SH groups, which are denoted as XH in Eq. 3, can be taken into account in a manner similar to that in the previous study. \(^{16}\)

\[
E = \sum_{\pi-\pi} \left( D_{\text{vertical}} \left( 1 - e^{-A_{\text{vert}}(R_{\text{eq}} - R_{\text{act}})} \right)^2 - 1 \right) \left[ 1 - (\vec{n}_i \cdot \vec{n}_j) \right] \cos^2(\theta - \theta_0) \\
+ \sum_{\pi-\text{XH}} \left( D_{\text{XH}} \left( 1 - e^{-A_{\text{XH}}(R_{\text{eq}} - R_{\text{act}})} \right)^2 - 1 \right) \cos^2(\theta - \theta_0) \\
+ \sum_{\text{other}} \left( D_{\text{other}} \left( 1 - e^{-A_{\text{other}}(R_{\text{eq}} - R_{\text{act}})} \right)^2 - 1 \right) \cos^2(\theta - \theta_0)
\]

Equation 3

Orientation program can perform geometry optimization of both the structures of the ligand and receptor. The optimized structure corresponds to the energy-minimized point on the energy surface of the ligand–receptor complex. The first and second terms in Eq. 3. Not only the interaction of two aromatic rings but also the repulsion of atoms. This term is needed to avoid unfavorable atom collision. The fourth term, specific interaction, is effective only if the designated atom pairs are indicated in the input file. The fourth term is seldom activated.

The inclusion of \( \pi-\pi \) interaction is possible through the terms in Eq. 3. Not only the interaction of two aromatic rings but also the interaction between an aromatic ring and any of the CH, NH, OH, SH groups, which are denoted as XH in Eq. 3, can be taken into account in a manner similar to that in the previous study. \(^{16}\)

\[
E = \sum_{\pi-\pi} \left( D_{\text{vertical}} \left( 1 - e^{-A_{\text{vert}}(R_{\text{eq}} - R_{\text{act}})} \right)^2 - 1 \right) \left[ 1 - (\vec{n}_i \cdot \vec{n}_j) \right] \cos^2(\theta - \theta_0) \\
+ \sum_{\pi-\text{XH}} \left( D_{\text{XH}} \left( 1 - e^{-A_{\text{XH}}(R_{\text{eq}} - R_{\text{act}})} \right)^2 - 1 \right) \cos^2(\theta - \theta_0) \\
+ \sum_{\text{other}} \left( D_{\text{other}} \left( 1 - e^{-A_{\text{other}}(R_{\text{eq}} - R_{\text{act}})} \right)^2 - 1 \right) \cos^2(\theta - \theta_0)
\]

Equation 3

Orientation program can perform geometry optimization of both the structures of the ligand and receptor. The optimized structure corresponds to the energy-minimized point on the energy surface of the ligand–receptor complex. The first and second terms in Eq. 3. Not only the interaction of two aromatic rings but also the repulsion of atoms. This term is needed to avoid unfavorable atom collision. The fourth term, specific interaction, is effective only if the designated atom pairs are indicated in the input file. The fourth term is seldom activated.

The inclusion of \( \pi-\pi \) interaction is possible through the terms in Eq. 3. Not only the interaction of two aromatic rings but also the interaction between an aromatic ring and any of the CH, NH, OH, SH groups, which are denoted as XH in Eq. 3, can be taken into account in a manner similar to that in the previous study. \(^{16}\)
sion of guanidine into amine in 2QWD largely reduces the inhibitory activity. The conversion into hydroxyl group further decreases the $pK_a$ values in 2QWB, 2QWC, and 2SIM.

For PCP, a macrocyclic inhibitor (PDB#: 1BXO) is depicted in Fig. 1(f). This inhibitor contains both polar and non-polar moieties in a good balance and is well fitted to the hollow site of the protein. The other 7 ligands for PCP have no cyclic structure as shown in Fig. S1(f). The ligands of 1APV and 1APW have no aromatic rings and isopropyl and isobutyl side chains are assembled by ester linkages. A high similarity is observed in chemical structures of 1BXQ, 1PPL, 1PPM, 2WEB and 2WEC. Phenylpropionic acid methyl ester moiety is located at one terminal side with connected through a phosphoryl group. Isobutyl is bound to the opposite terminal side in 1BXQ and 1PPL. Benzyl is bound in 1PPM and naphthalene is bound in 2WEB and 2WEC. In spite of the presence
of the aromatic rings in 1PPM, 2WEB, 2WEC, the pKᵢ values of these 3 ligands are less than those of the isobutyl-bound ligands of 1BXQ and 1PPL.

For PlmII, Pepstatin (PDB#: 1SME) is depicted in Fig. 1(g). Pepstatin has a high hydrophilicity and a large conformational flexibility due to many amide linkages, and it attaches to the protein surface with good shape complementarity. Pepstatin has no aromatic rings while other 3 ligands of 1LEE, 1LF2, and 1LF3 contain 3 or 5 phenyls as shown in Fig. S1(g). Those aromatic rings are mainly connected through amide linkage in a similar manner to Pepstatin in 1SME. The pKᵢ values of the ligands are, however, not so high as Pepstatin.

For PNPase, Formycin B (PDB#: 1A69) is depicted in Fig. 1(h). Formycin B has many polar atoms, several of which effectively make hydrogen bonds with the protein. In the data set for PNPase, 7 ligands, 1B8N, 1B8O, 1G2O, 1K9S, 1N3I, 1N3J, 1N3K, and 1X6J are shown in Fig. 2.

Fig. 2. Correlation between Experimental Inhibitory Constant and Computational Affinity Score

The inhibitory constant is presented in logarithmic scale in the abscissa and the affinity score obtained by Orientation is presented in the ordinate. Since Orientation score is given as a negative value, the score is converted to a positive value for comparison. (a) AChE, (b) CDK2, (c) FXa, (d) FKBP, (e) NEU, (f) PCP, (g) PlmII, (h) PNPase. The numbers of plots in the respective proteins correspond to those of data in Table 1.
The correlation coefficients between calculated score and experimental activity were compared among software programs as shown in Table 2. The relationships between the scores by the respective programs and the experimentally measured \( K_i \) values for all the ligand–receptor complexes were shown in Figs. S2–S6. Compatibility between calculation and experimental measurement are very different among the programs and the target proteins. The correlations for AChE are good except for that shown by AutoDock. AutoDock, however, gives the best correlation coefficient for FXa. All of the programs show good correlations for CDK2, while the correlations for PlmII are poor except for that shown by Orientation. All of the software programs give good correlations for NEU, PCP, and PNPase. The correlations for FKBP show large differences among the programs. As a whole, Orientation shows good correlations for many target proteins.

### Discussion

Orientation program can perform geometry optimization of ligand–receptor complex structures. We tested the optimizing calculation for all of the complex structures used as a dataset in this study. The optimized structures were similar to the original crystal structures for almost all of the complexes. Judging from the changes in total energy obtained as a sum of Eqs. 1 and 2, 30000 is a sufficient number of optimization cycles to reach a stable structure for small complex models. Even for large complex models, 50000 cycles are sufficient for ordinary computational analysis of the ligand–receptor binding mode. Orientation scores became better for all of the complex structures by optimization. The correlation coefficients between the calculated scores and experimentally measured \( K_i \) values were increased for NEU and decreased for PCP and PlmII. For other target proteins, the correlation coefficient was almost unchanged. Both in PCP and PlmII, the score of one ligand chemical was greatly shifted by the optimization compared to the other chemicals. As a whole, the optimizing routine in Orientation program works properly.

### Table 2. Comparison of Correlation Coefficients among Software Programs and Scoring Functions

| Protein name                        | Orientation | AutoDock | AutoDock Vina | GOLD Goldscore | GOLD Chemscore | GOLD ASP |
|-------------------------------------|-------------|----------|----------------|----------------|----------------|----------|
| Acetylcholinesterase (AChE) \(^a\)  \(^b\)  | 0.86        | −0.26    | 0.72           | 0.72           | 0.88           | 0.90     |
| Cyclin-dependent kinase 2 (CDK2) \(^b\) | 0.93        | 0.98     | 0.92           | 0.80           | 0.82           | 0.96     |
| Factor Xa (FXa) \(^a\)             | 0.05        | 0.55     | 0.39           | 0.38           | −0.13          | −0.05    |
| FK506 binding protein (FKBP) \(^b\) | 0.87        | 0.35     | −0.15          | 0.09           | 0.19           | 0.28     |
| Neuraminidase (NEU) \(^b\)         | 0.89        | 0.75     | 0.52           | 0.85           | 0.60           | 0.74     |
| Penicillopepsin (PCP) \(^b\)       | 0.60        | 0.73     | 0.55           | 0.68           | 0.54           | 0.90     |
| Plasmapesin II (PlmII) \(^b\)      | 0.97        | −0.79    | −0.98          | −0.88          | −0.83          | −0.41    |
| Purine nucleoside phosphorylase (PNPase) \(^b\) | 0.84 | 0.70 | 0.78 | 0.46 | 0.94 | 0.41 |

\(^a\) Abbreviation is shown in parenthesis. \(^b\) The correlation coefficient for AChE of Orientation was drawn by graph (a) in Fig. 2. Those for CDK2, FXa, FKBP, NEU, PCP, PlmII, and PNPase of Orientation column were by (b), (c), (d), (e), (f), (g), and (h) in Fig. 2. The correlation coefficients for AutoDock correspond to the graphs in Fig. S2. Similarly, those for AutoDock Vina, GOLD Goldscore, GOLD Chemscore, and GOLD ASP correspond to Figs. S3, S4, S5, and S6, respectively.
AutoDock, the water molecules, even crystal water molecules, around the ligand–receptor complex are usually removed before docking calculation. It is, however, reasonable to incorporate water molecules for the evaluation of binding affinity because water molecules sometimes play an important role in enhancement of the inter-molecular interaction. The presence of water molecules has been reported to be essential in antigen–antibody recognition to support the formation of a complementary molecular interface. Orientation program can take into account the influence of one water molecule-mediated or two-water molecules-mediated hydrogen bonds connecting the ligand and receptor in estimation of the hydrophilic interaction in Eq. 2. If the computational model for a ligand–receptor complex is solvated with water, the water-mediated interaction is included in the calculation.

For comparison of Orientation program with other docking software programs in terms of accuracy for predicting ligand–receptor binding affinity, computations with AutoDock and GOLD software programs were carried out in this study. To avoid mis-evaluation due to the difference in binding modes, the scores in AutoDock and GOLD were obtained for X-ray crystal structures without executing docking calculation. In AutoDock, epdb entry was used in dpf file. In AutoDock Vina, the –score-only option was activated in execution. In GOLD, the rescoring routine was utilized with setting of the crystal structure as a docking pose. In Orientation, the calculation was carried out using crystal structures without geometry optimization. Although the assessment of correlation was done with the crystal structures in this study, the crystal structures rarely contain unfavorably short inter-atomic distances between the ligand and receptor. Accordingly, geometry optimization is favorable for reliable prediction of their binding affinity.

In the binding affinity prediction of AChE and CDK2 inhibitors, the scores calculated by almost all of the software programs showed good correlations with the reported \( K_i \) values. When the binding pocket is located at a deep hollow site of the target protein, computer programs can provide an accurate prediction of binding affinity. AChE inhibitors contain relatively large hydrophobic groups, and the hydrophobic interaction was estimated well by most of the programs. In the case of NEU, Orientation score was compatible with the reported \( K_i \) values (\( R=0.89 \)). NEU inhibitors contain many hydrophilic groups that can make hydrogen bonds with the protein. When only the hydrophobic term of Eq. 2 was plotted against the p\( K_i \) value, the correlation coefficient for NEU dropped to 0.23. This result indicates that Orientation can also accurately estimate the hydrogen bond contribution. In contrast, Orientation could not correctly predict the binding affinity for FXa. Although almost all of the programs showed little correlation with the experimental measurement for FXa, only AutoDock gave a fair correlation. The shape complementarity between the ligand and receptor is a key factor in predicting the binding affinity for FXa. Furthermore, the addition of an entropic contribution may improve the accuracy of binding affinity prediction.

In the case of FKBP and PlmII, Orientation was the only software that could correctly predict the affinities of the highest and lowest active molecules in the dataset. As for FKBP, the highest and lowest active molecules are both macrocyclic molecules, and the reported \( K_i \) values of these molecules are 0.9 nM (PDB#: 1PBK) and 100 nM (PDB#: 1FK1), respectively. In all of the scoring functions except for Orientation, the lowest active macrocyclic molecule was estimated to have the highest affinity. Those scoring functions may therefore overestimate the binding affinity of the macrocyclic skeleton, which directly makes contact with a large area of the protein surface. As for PlmII, the highest active molecule is a hexapeptide (Pepstatin) and the reported \( K_i \) value is 6 pm (PDB#: ISME). Orientation successfully distinguished the peptide ligand with the highest affinity. This result indicates that the scoring function can adequately estimate the hydrogen-bond energies between the peptide and receptor. It is notable that the molecular weights of the ligands for PlmII are considerably large as shown in Fig. S7. Taken together, these results suggest that Orientation is effective for predicting the binding affinity of not only low molecular-weight compounds (AChE, CDK2, NEU, and PNPase) but also macrocycles (FKBP) and peptides (PCP and PlmII). Accordingly, Orientation can also be used for prediction of binding affinities of peptide–protein and antigen–antibody complexes.

A huge amount of computations for a variety of compounds to the same kind of protein are sometimes needed in the research field of drug discovery. In in silico screening, every compound from a chemical database is bound to a target protein and its binding affinity is evaluated to search for hit molecules. In SBDD, many derivatives from an active compound are bound to the target and their affinities are estimated. This type of computation requires multi-processing units for better performance in a search from a diverse range of chemical structures. Recent progress in many-core computer facility has enabled large-scale computation to be carried out to satisfy such demands. Orientation program is suitable for a many-core facility. A shell program enabling parallel computation with message passing interface (MPI) technology can handle the execution of Orientation. In a test calculation with an FX10 super-computer system at Information Technology Center, The University of Tokyo, parallel computation of 384 complexes controlled by a shell program showed a parallel efficiency of 99.7%. The computation time depends on the size of the model as well as the cycle of optimization. If 192 different complexes, the sizes of which are less than any of the 8 examples in this study, are executed for 30000 optimization cycles, the computation will be finished within one day with the FX10 system. Therefore, the execution of Orientation program in a many-core facility is effective for identifying hit molecules and for optimizing the chemical structures of active compounds.

In this study, we used the Morse-type function to describe the inter-atomic potential between a ligand and a receptor. The Morse-type function has three parameters, \( D, A, \) and \( R_{eq} \), as shown in Eq. 2. On the other hand, the Lennard–Jones potential function is characterized by two parameters. Therefore, the Morse-type function can represent the inter-atomic interaction in detail and can be fitted to broad potential curves compared to the Lennard–Jones formula. This is the primary reason why we selected the Morse-type function in this study. The advantage of the Lennard–Jones function is the speed of computation. Hence, the Lennard–Jones function is adopted in most of the programs for molecular dynamics simulation. If the number of independent parameters is reduced to two by setting a constraint of \( A \times R_{eq}=6 \), the first and second de-
derivatives of the potential curve at the equilibration distance, $R_{eq}$, of the Morse-type function become equal to those of the Lennard–Jones function. Hence, the Morse-type function can trace the potential curves that the Lennard–Jones function represents.

**Conclusion**  
A software program to estimate the binding affinity of a ligand to a target protein was developed on the basis of the theory of molecular mechanics. The Morse-type potential function was used for evaluating the inter-molecular interaction, while electrostatic and Lennard–Jones potential functions were used for evaluating the non-bonded intra-molecular force and the interaction with solvent molecules. Bond, angle, and torsion force-fields were applied for the bonded terms. The binding scores obtained by the software were compared with experimentally measured $pK_i$ values. For the assessment of computational accuracy, 8 kinds of proteins and ligand molecules bound to the respective proteins were used in this study, and the correlations between computational prediction and experimental measurement were examined. The binding scores obtained by the software showed good compatibilities for the test dataset except for one target protein. To evaluate the performance of the developed software, other docking simulation programs, AutoDock and GOLD, were tested with the same dataset. Every software program showed good compatibility with the experimental measurement for some targets but showed poor correlations for other proteins. The developed software provides relatively reliable scores for prediction of molecular binding affinity.

**Acknowledgments**  
A part of this work was supported by a Grant for Scientific Research C from Japan Society for the Promotion of Science. This work was also supported by a Grant from Japanese Agency for Medical Research and Development. Large-scale computation was performed at Information Technology Center of the University of Tokyo. Calculations were also tested at Research Center for Computational Science, Okazaki, Japan.

**Conflict of Interest**  
The authors declare no conflict of interest.

**Supplementary Materials**  
The online version of this article contains supplementary materials.

**References**

1) Kitchen D. B., Decornez H., Furr J. R., Bajorath J., *Nat. Rev. Drug Discov.*, 3, 935–949 (2004).
2) Warren G. L., Andrews C. W., Capelli A. M., Clarke B., LaLonde J., Lambert M. H., Lindvall M., Nevins N., Semius S. F., Senger S., Tedesco G., Wall J. D., Woolven J. M., Peischoff C. E., Head M. S., *J. Med. Chem.*, 49, 5912–5931 (2006).
3) Ferrara P., Grohle H., Price D. J., Klebe G., Brooks C. L. III, *J. Med. Chem.*, 47, 3032–3047 (2004).
4) Perola E., Walters W. P., Charrifson P. S., *Proteins*, 56, 235–249 (2004).
5) Kim R., Skolnick J., *J. Comput. Chem.*, 29, 1316–1331 (2008).
6) Grohle H., Klebe G., *Angew. Chem. Int. Ed.*, 41, 2644–2676 (2002).
7) Sippl M. J., *J. Mol. Biol.*, 235, 859–883 (1990).
8) Wang R., Fang X., Lu Y., Wang S., *J. Med. Chem.*, 47, 2977–2980 (2004).
9) Wang R., Fang X., Lu Y., Yang C. Y., Wang S., *J. Med. Chem.*, 48, 4111–4119 (2005).
10) Jones G., Willett P., Glen R. C., Leach A. R., Taylor R., *J. Mol. Biol.*, 267, 727–748 (1997).
11) Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R., Goodsell D. S., Olson A. J., *J. Comput. Chem.*, 30, 2785–2791 (2009).
12) Trott O., Olson A. J., *J. Comput. Chem.*, 31, 455–461 (2010).
13) Dolinsky T. J., Czodrowski P., Li H., Nielsen J. E., Jensen J. H., Klebe G., Baker N. A., *Nucleic Acids Res.*, 35 (Suppl. 2), W522–W525 (2007).
14) Dolinsky T. J., Nielsen J. E., McCammon J. A., Baker N. A., *Nucleic Acids Res.*, 32 (Suppl. 2), W665–W667 (2004).
15) OpenEye Scientific Software, “SYBYL.”: <http://www.eyesopen.com>.
16) Yuki H., Tanaka Y., Hata M., Ishikawa H., Neya S., Hoshino T., *J. Comput. Chem.*, 28, 1091–1099 (2007).
17) Osajima T., Suzuki M., Neya S., Hoshino T., *J. Mol. Graph. Model.*, 53, 128–139 (2014).
18) Osajima T., Hoshino T., *Comput. Biol. Chem.*, 64, 368–383 (2016).
19) Mathematics and Computer Science Division, Argonne National Laboratory, “MPICH.”: <https://www.mpich.org/>.
20) Lim T. C., *Z. Naturforsch.*, 58a, 615–617 (2003).