Journal of Molecular Graphics and Modelling 31 (2011) 20–27

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

# Protein–ligand docking guided by ligand pharmacophore-mapping experiment by NMR

Yoshifumi Fukunishi[a,*], Yumiko Mizukoshi[a,b], Koh Takeuchi[a], Ichio Shimada[a,c], Hideo Takahashi[a,d], Haruki Nakamura[a,e]

[a] Biomedicinal Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26, Aomi, Koto-ku, Tokyo 135-0064, Japan
[b] Japan Biological Informatics Consortium (JBIC), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan
[c] Graduate School of Pharmaceutical Sciences, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
[d] Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, Suehirocho 1-7-29, Tsurumi-ku, Yokohama 230-0045, Japan
[e] Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

ARTICLE INFO

*Article history:*
Received 28 June 2011
Received in revised form 3 August 2011
Accepted 5 August 2011
Available online 3 September 2011

*Keywords:*
Protein–ligand docking
NMR
Ligand epitope-mapping
Protein–compound complex structure
*In silico* docking
Pharmacophore-mapping

ABSTRACT

We developed a new protein–ligand docking calculation method using experimental NMR data. Recently, we proposed a novel ligand epitope-mapping experiment, which utilizes the difference between the longitudinal relaxation rates of ligand protons with and without irradiation of target protein protons (DIRECTION epitope-mapping experiment; Y. Mizukoshi, et al., An accurate pharmacophore mapping method by NMR, submitted for publication). Although the epitope-mapping experiment is simple and rapid, the result should reflect the proximity of ligand protons to the target protein surface. However, it cannot directly provide the protein–ligand complex structure without any other structural information. While the accuracy of protein–ligand docking software is insufficient, the software can provide many candidate complex structures. In many cases, the correct complex structure is included in the set of predicted complex structures and the correct structures could be selected by applying the above experimental result of ligand epitope mapping. In the current study, we combined the protein–ligand docking software with the NMR experimental information so as to improve the prediction of the protein–ligand complex structure. Consequently, the prediction accuracy was improved by 1.3–1.9 times (from ca. 50% to ca. 70%) in a self-docking test for the simulated epitope mapping result. Moreover, this method was applied to actual NMR experiments, and it successfully reconstructed the protein–ligand complex structures.



## 1. Introduction

Protein–ligand docking is a key technology for *in silico* screening and many protein–ligand docking programs have been reported [1–19]. Since 1982, more than 50 docking programs have been developed, but their prediction accuracy remains insufficient [14–19]. In order to improve the prediction accuracy, it is necessary to improve the protein–ligand docking software and its post-processing method. In some recent analyses, the commonly used programs succeeded in hit prediction for about 50% of the target proteins but failed for the other 50% in *in silico* screening [10,13,14]. To overcome this problem, when several docking programs and scoring functions are available, consensus scoring methods are used [20–25]. One such docking program achieved an accuracy of 98% on the self-docking test, but the accuracy of cross-docking test was still insufficient, which is more practical than the self-docking test [14–19]. There is no perfect docking program, and the screening result depends on the combination of the docking program used and the target protein.

Protein–protein docking is a more complicated problem than protein–ligand docking due to the greater flexibility of a protein compared to that of small organic compound [26]. To perform protein–protein docking, supporting information is helpful. Multiple-sequence alignment of amino-acid sequences shows a conserved region that should be a likely protein interface [26]. In our previous studies, we proposed the use of the NMR cross-saturation (CS) signals in protein–protein docking [27–32]. In those studies, molecular dynamic (MD) simulation optimized the theoretical CS signal calculated from the snapshot structure

* Corresponding author. Tel.: +81 3 3599 8290; fax: +81 3 3599 8099.
*E-mail address:* y-fukunishi@aist.go.jp (Y. Fukunishi).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.08.002

to be fitted to the experimental CS signal, and those applications were successfully applied to build protein–protein complex structures.

We tried to apply the same approach of protein–protein docking with NMR experiment to the protein–ligand docking to improve the accuracy of predicted complexes. In the case of protein–ligand docking, a different type of NMR data is necessary. Among various ligand-observed NMR experiments, the saturation transfer difference (STD) experiment is the most widely used for the investigation of protein–ligand interactions [33,34]. However, there exists a potential problem with the STD method: namely, the difference of longitudinal relaxation of each ligand proton severely interferes with the quantitative result and sometimes leads to an erroneous conclusion, especially in the case of the ligand epitope-mapping experiment [35]. Given such difficulty, we proposed an alternative approach, the "Difference of Inversion RECovery rate with and without Target IrradiatiON" (DIRECTION) methodology for the epitope-mapping experiment, which utilizes the difference between the longitudinal relaxation rates of ligand protons with and without irradiation of target protein protons (Fig. 1) [36]. We applied this method to epitope-mapping experiments for several ligand–protein interaction systems and obtained consistent results with the estimated proton density around each ligand proton. The DIRECTION approach is simple and rapid, and could be suitable for high-throughput analysis. In the DIRECTION experiment, only the peak assignment of the ligand atom is needed and the peak assignment of the protein is not necessary. From the viewpoint of rapidness and feasibility, the DIRECTION analysis has been shown to be superior to the intermolecular NOE analysis [37,38]. Also, the DIRECTION analysis should be able to be applied to a mixture of compounds for higher throughput.

Here, we proposed a new ligand-docking method using the experimental DIRECTION data. In the protein–protein docking, the CS signal can be simulated by the Bloch equation based on the protein–complex coordinates. Thus, the experimental and simulated NMR results can be compared directly. On the other hand, the trend of the DIRECTION result can be simulated based on the protein–ligand complex coordinates. Thus, the correlation between the experimental and simulated results was used in the current study. This concept was included in our protein–ligand docking program and it improved the docking accuracy.

## 2. Method

The score ($G$) is the sum of the docking score ($G_0$) and the correlation coefficient between the experimental and theoretical NMR (DIRECTION) results (R):

$$G = G_0 + \lambda \times R, \tag{1}$$

where $\lambda$ is an arbitral parameter that is determined to maximize the prediction accuracy. Hereafter, we call the DIRECTION results "NMR data" or "NMR information". The theoretical NMR results can be calculated from the given protein–ligand complex structure. The current method maximizes the score $G$ to predict the protein–ligand complex structure while moving the ligand coordinates.

To move the ligand, the derivation of the score $G$ is necessarily.

$$\frac{\partial G}{\partial x_i} = \frac{\partial G_0}{\partial x_i} + \lambda \times \frac{\partial R}{\partial x_i}, \tag{2}$$

where $x_i$ is the coordinate of the $i$th atom of the ligand. Since the derivation of $G_0$ was given in the previous study (the first term), the derivation of the correlation coefficient (the second term) should be described.

Let $s_b^i$, $s_a^i$ and $Nc$ be the experimental NMR result of the $i$th atom, the theoretical NMR result of the $i$th atom, and the number of the observed NMR signals for ligand protons. The correlation coefficient R is given by

$$R = \frac{\sum_i \left(s_b^i - \left(\sum_i s_b^i/Nc\right)\right)\left(s_a^i - \left(\sum_i s_a^i/Nc\right)\right)}{\sqrt{\sum_i \left(s_b^i - \left(\sum_i s_b^i/Nc\right)\right)^2 \cdot \sum_i \left(s_a^i - \left(\sum_i s_a^i/Nc\right)\right)^2}}. \tag{3}$$

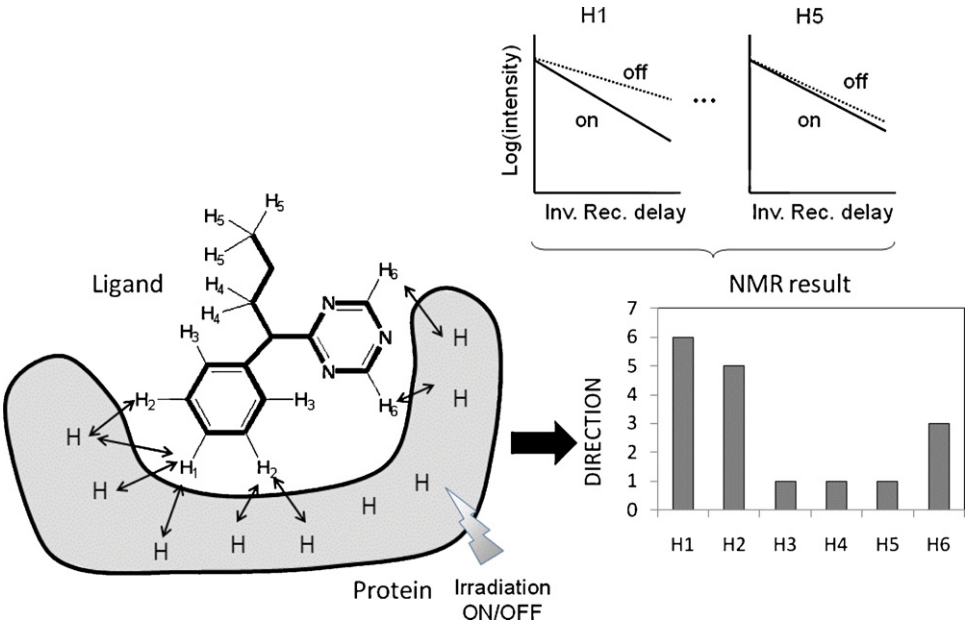


**Fig. 1.** Schematic representation of DIRECTION NMR experiment and signal intensity.

The derivation of the correlation coefficient is given by

$$\frac{\partial R}{\partial x_k} = \frac{1}{\sum_i Bi^2 \sum_i Ai^2}\left(\sqrt{\sum_i Bi^2 \sum_i Ai^2}\,\frac{\partial}{\partial x_k}\sum_i BiAi \sum_j BiAi \frac{\partial}{\partial x_k}\sqrt{\sum_i Bi^2 \sum_i Ai^2}\right)$$

$$= \frac{1}{\sum_i Bi^2 \sum_i Ai^2}\left(\sqrt{\sum_i Bi^2 \sum_i Ai^2}\sum_i Bi\left\{\frac{\partial}{\partial x_k}Ak\right\} - \sum_i BiAi\sqrt{\sum_i Bi^2}\,\frac{1}{\sqrt{\sum_i Ai^2}}\sum_i Ai\left\{\frac{\partial}{\partial x_k}Ak\right\}\right), \quad (4)$$

where

$$Bi = s_b^i - \frac{\sum_i s_b^i}{Nc}, \quad (5)$$

$$Ai = s_a^i - \frac{\sum_i s_a^i}{Nc}, \quad (6)$$

and

$$\frac{\partial}{\partial x_i}Ai = \frac{\partial}{\partial x_i}\sum_j \frac{1}{R_{ij}^6} = -6\sum_j \frac{1}{R_{ij}^7}\frac{\partial}{\partial x_i}\sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2}$$

$$= -6\sum_j \frac{1}{R_{ij}^7}\frac{(x_i-x_j)}{\sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2}}. \quad (7)$$

Since the DIRECTION results are basically related to the intermolecular cross-relaxation terms [36], the results can be substantially calculated from the given protein–ligand complex structure by

$$s_a^i = \sum_j^{NP} \frac{1}{R_{ij}^6}, \quad (8)$$

where $R_{ij}$ is the distance between the $i$th hydrogen atom of the ligand and the $j$th hydrogen atom of the protein, and NP is the number of hydrogen atoms of the protein. The $s_a^i$ value must be averaged for the equivalent hydrogen atoms, for example, the three or two hydrogen atoms of $CH_3$ and $CH_2$ groups. In practical calculation, the $s_a^i$ value was given by

$$s_a^i = \sum_j^{NP} E_{ij}, \quad (9)$$

$$E_{ij} = \begin{cases} 4\left(\dfrac{A^{12}}{R_{ij}^{12}} - \dfrac{A^6}{R_{ij}^6}\right) & : \text{when } R \geq 2^{1/6}A \\ -1 & : \text{when } R < 2^{1/6}A \end{cases}, \quad (10)$$

to avoid the error of division by zero ($R_{ij}=0$). The $R^{-12}$ term was introduced to reduce the sensitivity of E against R, since the docking program provides rough protein–ligand complex structures with atomic collisions. The van der Waals radius of hydrogen atom is 1.1–1.5 Å in AMBER force field and we tried the $A$ values of 0.1, 0.2, 0.5 and 1 Å. The $A$ value was set to 1.0 Å in the current study, since the $A$ value does not change the result so much when the $A$ value is less than the van der Waals radius.

The prediction procedure is as follows:

Step 1. The ligand was docked to the target protein by the Sievgene protein–ligand docking program [10] and the top 100 predicted structures were sampled.
Step 2. These top-ranked structures were optimized to maximize the $G$ score of Eq. (1) using Eq. (4) by using the steepest descent method.
Step 3. These optimized structures were re-ranked by Eq. (1). The top-ranked structure was selected as the prediction result.

Some λ values were examined and roughly optimized to maximize the self-docking accuracy using the data of Section 3.1. The λ value was set to $-1.0 \times 10^4$ for structure optimization by Eq. (4), and to −500 for the final scoring of the generated structures. The Sievgene protein–ligand docking program [10] is available on the web site (http://presto.protein.osaka-u.ac.jp/myPresto4/index_e.html).

## 3. Result

### 3.1. Data preparation for protein–ligand docking based on the simulated NMR information

We first examined the self-docking performance for the simulated NMR information. We applied our method to the protein–ligand complex structures derived from the protein data bank (PDB). Here, 108 complexes were selected from the database used in the evaluation of GOLD and FlexX [39]. This data set contains a rich variety of proteins and ligands, whose structures have all been determined by high-quality experiments with a resolution of less than 2.5 Å. Almost all the atom coordinates are supplied, and the all-atomic structures around the ligand pockets are quite reliable. Thus, this data set has so far been used in the clustering analysis of proteins, ligands and *in silico* screening [40–43]. We removed from the original data set those complexes containing a covalent bond between the protein and ligand. The PDB identifiers are summarized in Appendix A. All water molecules and cofactors were removed from the proteins, and all missing hydrogen atoms were added to form the whole-atom models of the proteins. The conformations of all ligands were randomized before the docking simulation.

The size distribution of ligands was as follows: 1–9 atoms, 3.6%; 10–19 atoms, 15.2%; 20–29 atoms, 30.9%; 30–39 atoms, 15.4%; 40–49 atoms, 15.8%; 50–59 atoms, 10.6%; and more than 60 atoms, 16.1%. The average ligand size was 37.1 atoms. For all protein–ligand complexes, the atomic charges of the proteins were the same as the atomic charges in AMBER parm99 [44].

In order to prepare the simulated NMR (DIRECTION) results of these protein–ligand complex structures, the contribution of intermolecular cross-relaxation terms were calculated by Eq. (10) based on the experimental protein–ligand complex structure to mimic the experimental values. Here, it should be noted that the DIRECTION experiment would be conducted for only hydrogen atom connected to carbon atoms. Considering that the actual NMR experiments should be performed in $D_2O$, the hydrogen atoms of OH, SH and NH groups were exchanged to deuterium (D) atoms, which are not involved in the calculation.

Three types of structural models were examined to determine the structural dependence of the current method. In Model 1, the heavy atoms of the protein–ligand complex structures were fixed

**Table 1**
Docking accuracy obtained by self-docking test of Model 1.

| | Original[a] | NMR[b] | 5% Noise[c] | 10% Noise[d] | 15% Noise[e] | 20% Noise[f] |
|---|---|---|---|---|---|---|
| RESP | | | | | | |
| RMSD <1 Å | 21.4 | 32.5 | 33.6 | 33.6 | 38.1 | 38.1 |
| RMSD <2 Å | 60.7 | 67.5 | 69.0 | 67.3 | 65.5 | 68.1 |
| RMSD <3 Å | 68.4 | 77.8 | 77.9 | 76.1 | 76.1 | 81.4 |
| AM1 | | | | | | |
| RMSD <1 Å | 23.1 | 33.3 | 36.3 | 38.1 | 38.9 | 37.2 |
| RMSD <2 Å | 57.3 | 70.1 | 70.8 | 73.5 | 69.9 | 69.9 |
| RMSD <3 Å | 65.0 | 76.9 | 77.0 | 77.0 | 74.3 | 80.5 |
| Gasteiger | | | | | | |
| RMSD <1 Å | 17.1 | 37.6 | 37.2 | 38.9 | 38.9 | 37.2 |
| RMSD <2 Å | 50.4 | 70.1 | 70.8 | 70.8 | 68.1 | 68.1 |
| RMSD <3 Å | 57.3 | 78.6 | 78.8 | 79.6 | 78.8 | 78.8 |

[a] The original Sievgene.
[b] The current method with simulated NMR information without noise.
[c] The current method with simulated NMR information with 5% noise.
[d] The current method with simulated NMR information with 10% noise.
[e] The current method with simulated NMR information with 15% noise.
[f] The current method with simulated NMR information with 20% noise.

to their original coordinates and hydrogen atoms were added to the complex structures arbitrarily. In Model 2, the coordinates of the hydrogen atoms of Model 1 were optimized by energy optimization in vacuum. In Model 3, all coordinates of the ligand atoms were optimized in addition to the protein hydrogen atoms. The energy optimizations were performed using the cosgene/myPresto program [45], which is available on the web site (http://presto.protein.osaka-u.ac.jp/myPresto4/index_e.html). The dielectric constant was set to 4R where R is the inter-atomic distance. In Models 2 and 3, an AMBER param99 force field was applied to the proteins [44], and the general AMBER force field was applied to the ligand atoms [46].

### 3.2. Protein–ligand docking based on the simulated NMR information

The docking results are summarized in Tables 1–3. The numbers in Tables 1–3 are the probabilities (%). The docking accuracy did not depend on which model was used (Models 1, 2 and 3). The results summarized in Tables 1–3 show the similar trends of improvement by the current method. The docking accuracy slightly depended on the accuracy of the atomic charge. The restricted electrostatic point charges (RESP charges with RHF/6-31G*) were the most precise among these three charges [47]. We used Gaussian98 to perform the quantum chemical calculations [48]. The Mulliken populations of MOPAC AM1 (Quantum Chemistry Program Exchange (QCPE), Indiana University, Bloomington, Indiana, USA, http://openmopac.net/index.html) gave better results than those by the Gasteiger charges [49,50]. On average, the docking accuracy obtained by the RESP charges was the best among these three results, and the accuracy obtained by MOPAC AM1 was better than that obtained by the Gasteiger charges. The differences between these results were about 3%.

On average, the original method without the simulated NMR information reconstructed about 20%, 55%, and 62% of a total of 108 complexes with RMSD values of <1 Å, 2 Å, and 3 Å, respectively. Here, the RMSDs were calculated between all atom positions, with the exception of the hydrogen atoms of each docked ligand and the corresponding atom positions in the complex crystal structure. The current method drastically improved the prediction accuracy. Namely, the current method with the simulated NMR information reconstructed about 37%, 70%, and 80% of a total of 108 complexes with RMSD values of <1 Å, 2 Å, and 3 Å, respectively, thereby improving the accuracy of the original method by about 1.9, 1.3 and 1.3 times, respectively. Using almost the same dataset, DOCK [1], FlexX [2], and GOLD [3] were reported to reconstruct the complexes of 39%, 51%, and 56%, with RMSDs <2 Å, respectively, without any experimental information [4] on 2002, while the docking accuracy has been improved [14–19]. Thus, the accuracy of the current method provided much higher performances that those of the popular docking programs. The average CPU time was 130 s on a single CPU of a Xeon 5570.

**Table 2**
Docking accuracy obtained by self-docking test of Model 2.

| | Original[a] | NMR[b] | 5% Noise[c] | 10% Noise[d] | 15% Noise[e] | 20% Noise[f] |
|---|---|---|---|---|---|---|
| RESP | | | | | | |
| RMSD <1 Å | 22.1 | 37.2 | 40.7 | 41.6 | 43.4 | 40.7 |
| RMSD <2 Å | 58.4 | 76.1 | 76.1 | 75.2 | 77.0 | 75.2 |
| RMSD <3 Å | 62.8 | 85.0 | 83.2 | 85.0 | 85.0 | 84.1 |
| AM1 | | | | | | |
| RMSD <1 Å | 18.6 | 41.1 | 42.5 | 45.1 | 37.2 | 37.2 |
| RMSD <2 Å | 58.4 | 74.1 | 74.3 | 73.5 | 72.6 | 73.5 |
| RMSD <3 Å | 61.9 | 79.5 | 80.5 | 78.8 | 78.8 | 78.8 |
| Gasteiger | | | | | | |
| RMSD <1 Å | 19.5 | 38.1 | 38.9 | 45.1 | 37.2 | 40.7 |
| RMSD <2 Å | 55.8 | 70.8 | 69.0 | 71.7 | 69.0 | 69.9 |
| RMSD <3 Å | 61.9 | 81.4 | 78.8 | 78.8 | 79.6 | 78.8 |

[a] The original Sievgene.
[b] The current method with simulated NMR information without noise.
[c] The current method with simulated NMR information with 5% noise.
[d] The current method with simulated NMR information with 10% noise.
[e] The current method with simulated NMR information with 15% noise.
[f] The current method with simulated NMR information with 20% noise.

**Table 3**
Docking accuracy obtained by self-docking test of Model 3.

| | Original[a] | NMR[b] | 5%noise[c] | 10%noise[d] | 15%noise[e] | 20%noise[f] |
|---|---|---|---|---|---|---|
| RESP | | | | | | |
| RMSD <1 Å | 22.1 | 38.1 | 38.9 | 38.1 | 41.6 | 34.5 |
| RMSD <2 Å | 54.0 | 73.5 | 73.5 | 71.7 | 72.6 | 68.1 |
| RMSD <3 Å | 61.1 | 80.5 | 81.4 | 79.6 | 80.5 | 77.0 |
| AM1 | | | | | | |
| RMSD <1 Å | 19.5 | 38.1 | 37.2 | 38.9 | 37.2 | 34.5 |
| RMSD <2 Å | 54.9 | 68.1 | 67.3 | 69.9 | 68.1 | 69.0 |
| RMSD <3 Å | 62.8 | 78.8 | 77.9 | 77.0 | 78.8 | 77.9 |
| Gasteiger | | | | | | |
| RMSD <1 Å | 19.5 | 31.0 | 30.1 | 34.5 | 36.3 | 28.3 |
| RMSD <2 Å | 53.1 | 66.4 | 66.4 | 64.6 | 71.7 | 62.8 |
| RMSD <3 Å | 60.2 | 75.2 | 79.6 | 73.5 | 78.8 | 74.3 |

[a] The original Sievgene.
[b] The current method with simulated NMR information without noise.
[c] The current method with simulated NMR information with 5% noise.
[d] The current method with simulated NMR information with 10% noise.
[e] The current method with simulated NMR information with 15% noise.
[f] The current method with simulated NMR information with 20% noise.

To examine the robustness of the current method, we added 5%, 10%, 15% or 20% noise to the simulated NMR information. The prediction results were not disturbed very much by the noise; the prediction accuracy was close to that obtained by the simulated NMR data without noise.

### 3.3. Protein–ligand docking based on the experimental DIRECTION data

We applied the current method to a practical problem. The target protein was p38 MAP kinase. To evaluate our result, the docking results were compared with the protein–ligand complex structures registered in the PDB (PDBID: 1KV1 and 1A9U). The two ligand molecules (1-(5-tert-butyl-2-methyl-2H-pyrazol-3-yl)-3-(4-chlorophenyl)-urea, with BMU as the PDB Ligand code and 4-[5-(4-fluorophenyl)-2-[4-[(S)-methylsulfinyl]phenyl]-3H-imidazol-4-yl]pyridine; SB2) were the same ligands as those of 1KV1 and 1A9U, respectively. The atomic charges of the proteins were the same as the atomic charges in AMBER parm99 [44], and the atomic charges of these two ligands were calculated by both the Gasteiger method [49,50] and the RESP with RHF/6-31G* method [47].

The DIRECTION experimental data were exactly the same as those reported in our previous paper [36]. The docking results with and without the NMR information are summarized in Table 4. The intensity of simulated data was normalized ($\Sigma s_a^i = \Sigma s_b^i$). Without using the DIRECTION data, a complex structure with BMU was successfully reconstructed, but the structure with SB2 failed, when the Gasteiger charges were used. In the latter case with SB2, the correct structure was the 5th structure. Using the DIRECTION data, both complex structures with BMU and SB2 respectively were correctly reconstructed. As same as the results in Section 3.1, the results obtained by the RESP charge were more precise than those obtained by the Gasteiger charge. Figs. 2 and 3 show the docking poses of BMU and SB2, respectively. The docking poses of BMU with and

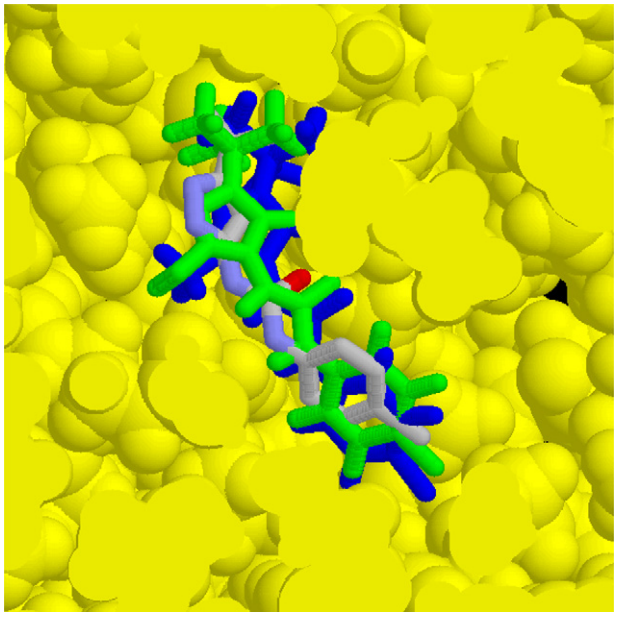

**Fig. 2.** Docking pose of BMU (PDBID: 1KV1). CPK, blue and green stick figures represent the crystal structure of BMU, the docking pose without NMR data, and the docking pose with NMR data while the yellow spheres represent the target protein (PDBID: 1KV1).

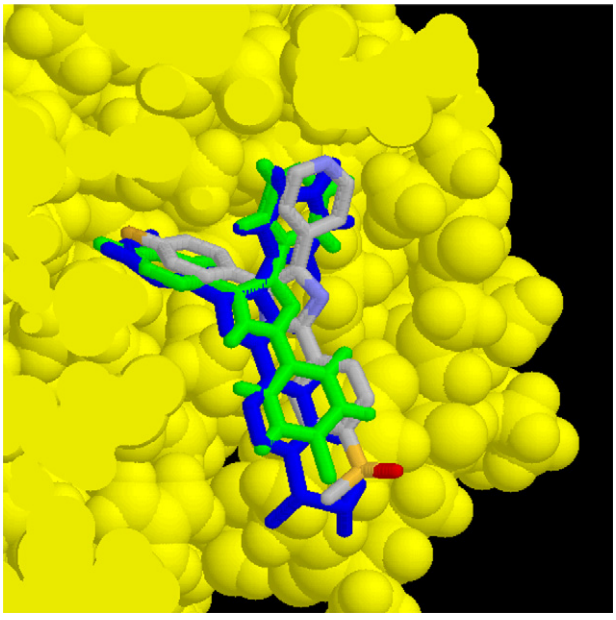

**Fig. 3.** Docking pose of SB2 (PDBID: 1A9U). CPK, blue and green stick figures represent the crystal structure of SB2, docking pose without NMR data, and docking pose with NMR data while the yellow spheres represent the target protein (PDBID: 1A9U). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 4**
Docking accuracy obtained by experimental DIRECTION results for p38 MAP kinase.

| Ligand | Rank | Without NMR information | | | | With NMR information | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | With Gasteiger charge | | With RESP charge | | With Gasteiger charge | | With RESP charge | |
| | | Sievgene score | RMSD (Å) | Sievgene score | RMSD (Å) | Score | RMSD (Å) | Score | RMSD (Å) |
| BMU | 1 | −3.99 | 1.60 | −3.99 | 1.28 | -7.12 | 0.84 | -7.09 | 0.66 |
| | 2 | −3.99 | 1.59 | −3.99 | 1.27 | −7.10 | 0.85 | −7.09 | 0.66 |
| | 3 | −3.99 | 1.59 | −3.98 | 1.26 | −7.08 | 0.86 | −7.06 | 0.65 |
| | 4 | −3.99 | 1.59 | −3.98 | 1.24 | −7.04 | 0.79 | −6.27 | 0.64 |
| | 5 | −3.95 | 2.16 | −3.97 | 1.27 | −7.03 | 0.78 | −5.80 | 0.97 |
| SB2 | 1 | −3.83 | 4.79 | −3.40 | 0.60 | −5.80 | 0.84 | −5.09 | 0.51 |
| | 2 | −3.83 | 4.79 | −3.34 | 0.43 | −5.65 | 0.72 | −5.03 | 0.50 |
| | 3 | −3.82 | 4.79 | −3.34 | 0.43 | −5.56 | 4.90 | −5.00 | 0.58 |
| | 4 | −3.82 | 4.79 | −3.29 | 0.89 | −5.49 | 0.59 | -4.94 | 0.58 |
| | 5 | −3.78 | 0.94 | −3.21 | 1.16 | −5.48 | 2.41 | -4.86 | 0.43 |

without the DIRECTION data were qualitatively the same as those of the crystal structure. The docking pose of SB2 without the DIRECTION data was qualitatively different from the crystal structure; the positions of phenyl and pyridine groups were exchanged. On the contrary, the docking pose of SB2 was improved by the consideration of the DIRECTION data, and the result was the same as that of the crystal structure. Figs. 4 and 5 show the experimental and calculated NMR information for the complex structures with BMU and SB2, respectively. The trends of the experimental DIRECTION values were successfully reproduced by the calculation. Namely, the correlation coefficients between the experimental and simulated NMR results for the complex structures with BMU and SB2 were 0.91 and 0.99, respectively.

## 4. Discussion

For the investigation of protein–ligand interactions, STD experiment is a widely used NMR approach [33,34] and the result is often utilized as experimental constraints for the protein–ligand docking [51]. In the case where the longitudinal relaxation rates of each ligand proton do not show considerable variation, the result obtained from STD could directly reflect the effective intermolecular cross-relaxation rate constant. However, protons of small compounds usually exhibit a variety of longitudinal relaxation rates, which depends on the proton moiety considered (e.g. methyl, methylene, methine), local proton density, mobility, and the ligand conformations. The difference of the longitudinal relaxation of each ligand proton severely interferes with the derived epitope mapping result from STD [35].

In order to exclude such relaxation-dependent ambiguity, we recently developed a simple and efficient experimental approach

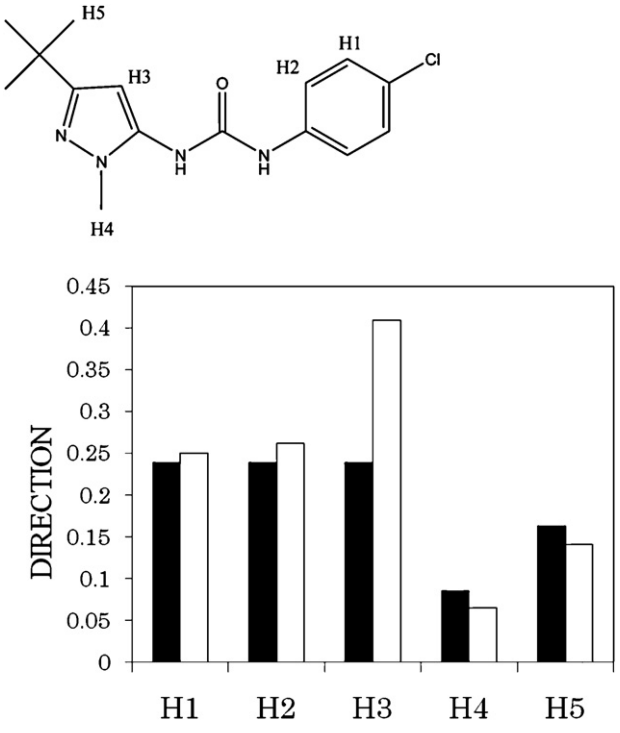


**Fig. 4.** Experimental and calculated DIRECTION results for 1KV1. Black bars and white bars represent the results obtained by the NMR experiment and the docking calculation, respectively.

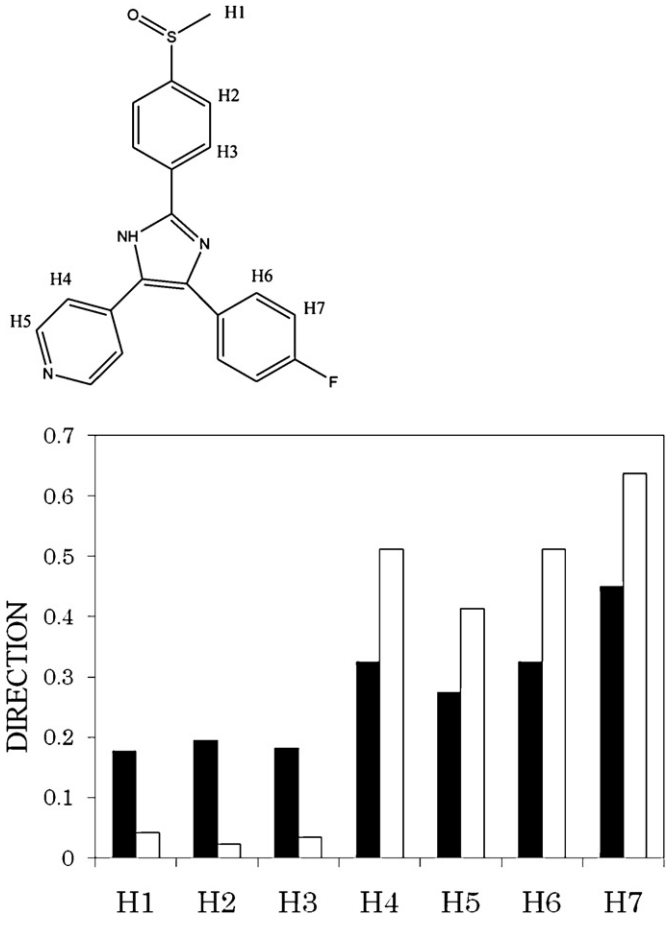


**Fig. 5.** Experimental and calculated DIRECTION results for 1AU9. Black bars and white bars represent the results obtained by the NMR experiment and the docking calculation, respectively.

to epitope mapping that utilizes the difference between the longitudinal relaxation rates of ligand protons with and without irradiation of target protein protons. We applied our approach to identifying the binding portions of ligand molecules for a number of ligand–protein interaction systems and obtained consistent results with the estimated proton density around each ligand proton [36]. However, the results obtained by the DIRECTION experiment were not perfectly coincident with the theoretical intermolecular cross-relaxation terms, since other factors including the spin diffusion effect [52] between adjacent protons might affect the experimental results. Therefore, instead of using the absolute value of the experimental results by NMR, the current method utilizes the trend of NMR results, which is well provided by the correlation coefficient (Eq. (3)). This point is different from our previous study on protein–protein docking [30], and the current method should be robust against such dynamic propagation phenomena.

For fragment-based drug development (FBDD), the seed and lead compounds are relatively small and show weak affinities. Although NMR experiments are frequently applied in the FBDD process [53], protein–ligand docking studies are difficult for small fragments, making the design of fragment growth difficult [53]. The DIRECTION epitope-mapping experiment is based on a ligand-observed NMR approach and thus can intrinsically adapt to weak protein–ligand interaction systems. Therefore, the current docking method coupled with the ligand epitope-mapping experiment could provide us a guideline for fragment growing and/or linking in an FBDD project.

This method could be combined with any kind of protein–ligand docking program, since Eq. (1) is a general form and it is not specialized for the Sievgene docking program. Sievgene has been developed for large-scale *in silico* drug screening, so that it is fast but its precision is not very high. If a docking program has greater accuracy than the Sievgene program, more precise results would be expected.

## 5. Conclusion

We developed a new protein–ligand docking method guided by the ligand epitope-mapping experiment by NMR. The protein–ligand complex structures generated by the ordinary protein–ligand docking software were optimized to be consistent with the experimental NMR information of the compound, and the most suitable structure was selected as the prediction result. This method was applied to the self-docking test of 108 known protein–ligand complex structures derived from the PDB. The current method with the information of simulated ligand epitope mapping by NMR reconstructed about 37%, 70%, and 80% of a total of 108 complexes with RMSD values of <1 Å, 2 Å, and 3 Å, respectively. These accuracies represent significant improvement of about 1.9, 1.3 and 1.3 times, respectively, compared to the ordinary docking results without the simulated NMR information. The current method was applied to the docking of two ligands onto p38 MAP kinase using the experimental NMR (DIRECTION) data. Both ligands were successfully docked to the target protein by the current method, while one of the two ligands could not be docked by the original docking software without the NMR information.

## Acknowledgements

This work was supported by grants from the New Energy and Industrial Technology Development Organization of Japan (NEDO) and the Ministry of Economy, Trade, and Industry (METI) of Japan.

## Appendix A.

The following PDB identifier list of complexes was used: 1a28, 1a42, 1a4g, 1a4q, 1abe, 1abf, 1aoe, 1apt, 1apu, 1aqw, 1atl, 1b58, 1b9v, 1bma, 1byb, 1byg, 1c5c, 1cbs, 1cbx, 1ckp, 1com, 1coy, 1cps, 1cvu, 1d0l, 1d3h, 1dd7, 1dg5, 1dhf, 1dog, 1dr1, 1ebg, 1eed, 1ejn, 1epb, 1epo, 1ets, 1f0r, 1f0s, 1f3d, 1fen, 1fkg, 1fki, 1fl3, 1glp, 1hdc, 1hfc, 1hos, 1hpv, 1hsl, 1htf, 1hyt, 1ida, 1ivb, 1jap, 1lah, 1lcp, 1lic, 1lst, 1mdr, 1mmq, 1mts, 1mup, 1nco, 1okl, 1phd, 1phg, 1poc, 1ppc, 1pph, 1pso, 1qbr, 1qbu, 1rne, 1rnt, 1rob, 1snc, 1srj, 1tng, 1tnh, 1tni, 1tnl, 1tyl, 1xid, 1xie, 1yee, 2aad, 2ack, 2ada, 2cht, 2cpp, 2ctc, 2fox, 2gbp, 2ifb, 2pk4, 2qwk, 2tmn, 3cla, 3cpa, 3erd, 3ert, 4lbd, 4phv, 5abp, 5cpp, 5er1, and 6rnt.

## References

[1] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule–ligand interactions, J. Mol. Biol. 161 (1982) 269–288.
[2] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J. Mol. Biol. 261 (1996) 470–489.
[3] G. Jones, P. Willet, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[4] N. Paul, D. Rognan, ConsDock: a new program for the consensus analysis of protein–ligand interactions, Proteins 47 (2002) 521–533.
[5] C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, M.D. Eldridge, Flexible docking using tabu search and an empirical estimate of binding affinity, Proteins 33 (1998) 367–382.
[6] D.S. Goodsell, A.J. Olson, Automated docking of substrates to proteins by simulated annealing, Proteins 8 (1990) 195–202.
[7] R. Abagyan, M. Totrov, D. Kuznetsov, ICM: a new method for structure modeling and design: application to docking and structure prediction from the disordered native conformation, J. Comput. Chem. 15 (1994) 488–506.
[8] P.M. Colman, Structure-based drug design, Curr. Opin. Struct. Biol. 4 (1994) 868–874.
[9] A. Kramer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, LigScore: a novel scoring function for predicting binding affinities, J. Mol. Graph. Model. 23 (2005) 395–407.
[10] Y. Fukunishi, Y. Mikami, H. Nakamura, Similarities among receptor pockets and among compounds: analysis and application to in silico ligand screening, J. Mol. Graph. Model. 24 (2005) 34–45.
[11] C. Zhang, S. Liu, Q. Zhu, Y. Zhou, A knowledge-based energy function for protein–ligand, protein–protein, and protein–DNA complexes, J. Med. Chem. 48 (2005) 2325–2335.
[12] I. Muegge, Y.C. Martin, A general and fast scoring function for protein–ligand interactions: a simplified potential approach, J. Med. Chem. 42 (1999) 791–804.
[13] G.L. Warren, C. Webster Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring functions, J. Med. Chem. 49 (2006) 5912–5931.
[14] J.B. Cross, D.C. Thompson, B.K. Rai, J.C. Baber, K.Y. Fan, Y. Hu, C. Humblet, Comparison of several molecular docking programs: pose prediction and virtual screening accuracy, J. Chem. Inf. Model. 49 (2009) 1455–1474.
[15] S. Mukherjee, T.E. Balius, R.C. Rizzo, Docking validation resources: protein family and ligand flexibility experiments, J. Chem. Inf. Model. 50 (2010) 1986–2000.
[16] M. Sandor, R. Kiss, G.M. Keseru, Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes, J. Chem. Inf. Model. 50 (2010) 1165–1172.
[17] R. Thilagavathi, R.L. Mancera, Ligand–protein cross-docking with water molecules, J. Chem. Inf. Model. 50 (2010) 415–421.
[18] J.A. Feng, G.R. Marshall, SKATE: a docking program that decouples systematic sampling from scoring, J. Comput. Chem. 31 (2010) 2540–2554.
[19] A. Grosdidier, V. Zoete, O. Michielin, Fast docking using the CHARMM force field with EADock DSS, J. Comput. Chem. 32 (2011) 2149–2159.
[20] P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters, Consensus scoring: a method for obtaining improved hit rates from docking database of three-dimensional structures into proteins, J. Med. Chem. 42 (1999) 5100–5109.
[21] R. Wang, S. Wang, How does consensus scoring work for virtual library screening? An idealized computer experiment, J. Chem. Inf. Comput. Sci. 41 (2001) 1422–1426.
[22] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew, Consensus scoring for ligand/protein interactions, J. Mol. Graph. Model. 20 (2002) 281–295.
[23] R. Wang, Y. Lu, S. Wang, Comparative evaluation of 11 scoring functions for molecular docking, J. Med. Chem. 46 (2003) 2287–2303.
[24] J.M. Yang, Y.F. Chen, T.W. Shen, B.S. Kristal, D.F. Hsu, Consensus scoring criteria for improving enrichment in virtual screening, J. Chem. Inf. Model. 45 (2005) 1134–1146.
[25] R. Teramoto, H. Fukunishi, Supervised consensus scoring for docking and virtual screening, J. Chem. Inf. Model. 47 (2007) 526–534.
[26] J. Janin, Assessing predictions of protein–protein interaction: the CAPRI experiment, Protein. Sci. 14 (2005) 278–283.

[27] I. Shimada, NMR techniques for identifying the interface of a larger protein–protein complex: cross-saturation and transferred cross-saturation experiments, Methods Enzymol. 394 (2005) 483–506.
[28] H. Takahashi, T. Nakanishi, K. Kami, Y. Arata, I. Shimada, A novel NMR method for determining the interfaces of large protein–protein complexes, Nat. Struct. Biol. 7 (2000) 220–223.
[29] T. Matsuda, T. Ikegami, N. Nakajima, T. Yamazaki, H. Nakamura, Model building of a protein–protein complexed structure using saturation transfer and residual dipolar coupling without paired intermolecular NOE, J. Biomol. NMR 29 (2004) 325–338.
[30] E. Kanamori, S. Igarashi, M. Osawa, Y. Fukunishi, I. Shimada, H. Nakamura, Structure determination of a protein assembly by amino acid selective cross-saturation, Proteins 79 (2011) 179–190.
[31] M. Matsumoto, T. Ueda, I. Shimada, Theoretical analyses of the transferred cross-saturation method, J. Magn. Reson. 205 (2010) 114–124.
[32] I. Shimada, T. Ueda, M. Matsumoto, M. Sakakura, M. Osawa, K. Takeuchi, N. Nishida, H. Takahashi, Cross-saturation and transferred cross-saturation experiments, Prog. Nucl. Magn. Reson. Spectrosc. 54 (2009) 123–140.
[33] M. Mayer, B. Meyer, Characterization of ligand binding by saturation transfer difference NMR spectroscopy, Angew. Chem. Int. Ed. 38 (1999) 1784–1788.
[34] M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contract with a protein receptor, J. Am. Chem. Soc. 123 (2001) 6108–6117.
[35] J. Yan, A.D. Kline, H. Mo, M.J. Shapiro, E.R. Zart ler, The effect of relaxation on the epitope mapping by saturation transfer difference NMR, J. Magn. Reson. 163 (2003) 270–276.
[36] Y. Mizukoshi, A. Abe, T. Takizawa, H. Hanzawa, Y. Fukunishi, I. Shimada, H. Takahashi, An accurate pharmacophore mapping method by NMR, submitted for publication.
[37] I. Bertini, M. Fragai, A. Giachetti, C. Luchinat, M. Maletta, G. Parigi, K.J. Yen, Combining in silico tools and NMR data to validate protein–ligand structural models: application to matrix metalloproteinases, J. Med. Chem. 48 (2005) 7544–7559.
[38] K. Ono, H. Ueda, Y. Yoshizawa, D. Akazawa, R. Tanimura, I. Shimada, H. Takahashi, Structural basis for platelet anti-aggregation by angiotensin II type 1-receptor antagonist losartan (DuP-753) via glycoprotein VI, J. Med. Chem. 53 (2010) 2087–2093.
[39] J.W.M. Nissink, C. Murray, M. Hartshorn, M.L. Verdonk, J.C. Cole, R. Taylor, A new test set for validating predictions of protein–ligand interaction, Proteins 49 (2002) 457–471.
[40] Y. Fukunishi, Y. Mikami, S. Kubota, H. Nakamura, Multiple target screening method for robust and accurate in silico ligand screening, J. Mol. Graph. Model. 25 (2005) 61–70.
[41] Y. Fukunishi, S. Kubota, H. Nakamura, Noise reduction method for molecular interaction energy: application to in silico drug screening and in silico target protein screening, J. Chem. Inf. Model. 46 (2006) 2071–2084.
[42] Y. Fukunishi, Y. Mikami, K. Takedomi, M. Yamanouchi, H. Shima, H. Nakamura, Classification of chemical compounds by protein-compound docking for use in designing a focused library, J. Med. Chem. 49 (2006) 523–533.
[43] Y. Fukunishi, S. Hojo, H. Nakamura, An efficient in silico screening method based on the protein-compound affinity matrix and its application to the design of a focused library for cytochrome P450 (CYP) ligands, J. Chem. Inf. Model. 46 (2006) 2610–2622.
[44] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, P.A. Kollman, AMBER 8, UCSF, 2004.
[45] Y. Fukunishi, Y. Mikami, H. Nakamura, The filling potential method: a method for estimating the free energy surface for protein–ligand docking, J. Phys. Chem. B 107 (2003) 13201–13210.
[46] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[47] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (2000) 1049–1074.
[48] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann Jr., J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Revision A.9, Gaussian, Inc., Pittsburgh, PA, 1998.
[49] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.
[50] J. Gasteiger, M. Marsili, A new model for calculating atomic charges in molecules, Tetrahedron Lett. (1978) 3181–3184.
[51] V. Jayalakshmi, N.R. Krishna, Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand–receptor complexes, J. Magn. Reson. 155 (2002) 106–118.
[52] A. Kalk, H.J.C. Berendsen, Proton magnetic relaxation and spin diffusion in proteins, J. Magn. Reson. 24 (1976) 343–366.
[53] M. Orita, K. Ohno, M. Warizaya, Y. Amano, T. Niimi, Lead generation and examples: opinion regarding how to follow up hits, Methods Enzymol. 493 (2011) 383–419.