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 λ 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 ith 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 ith atom, the theoretical NMR result of the ith atom, and the number of the observed NMR signals for ligand protons. The correlation coefficient R is given by

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

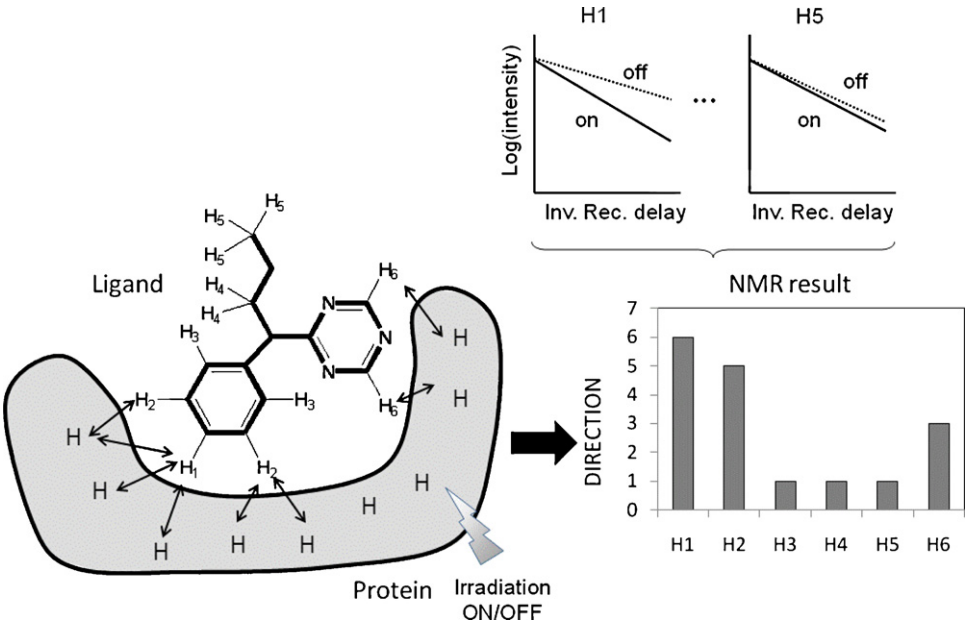

Fig. 1. Schematic representation of DIRECTION NMR experiment and signal intensity.