J|A|C|S
COMMUNICATIONS

Published on Web 02/12/2002

Structures of Protein−Protein Complexes Are Docked Using Only NMR Restraints from Residual Dipolar Coupling and Chemical Shift Perturbations

Mark A. McCoy* and Daniel F. Wyss

Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033

Received October 5, 2001

Research in structural genomics is expected to lead to a wealth of new three-dimensional protein structures.[1] Post-genomic structure determination will likely focus on soluble single proteins and domains. Proteins achieve their biological function by interacting with other molecules; thus, knowledge about the biological function of proteins will be enhanced with structures of protein complexes. Structures of complexes are, however, typically difficult to solve by either X-ray crystallography or NMR. Moreover, details about the interaction between proteins and ligands are difficult to predict. A significant barrier to NMR structure determination of complexes is the reliance on the intermolecular NOE to restrain protein−protein interactions. Even in complexes where intermolecular NOEs are numerous, peaks arising in the filtered dimension are often ambiguous, and artifacts are common. Clearly, it is desirable to use additional non-NOE-based restraints to determine structures of protein complexes.

Using restraints from residual dipolar coupling (RDC) can reduce the reliance on intermolecular NOEs for calculations of protein−protein complexes.[2−4] The measurement of RDC from partially oriented protein complexes allows for the determination of long-range order over the entire complex, but places no restraint on intermolecular translation. Providing that high-resolution structures of the proteins in the unbound state are available and no significant backbone conformational changes occur upon complex formation, structures of the complex can be calculated with RDC and NOE restraints. Since the orientation information contained in the RDC restraints can replace the orientation information in intermolecular NOEs, structures of complexes can be determined with many fewer NOEs which dramatically reduces the time and effort required for structure determination.[4]

Intermolecular restraints derived from chemical shift perturbations offer another alternative to intermolecular NOEs. Structures of protein−ligand complexes have been determined by minimizing the difference between experimental and simulated chemical shift perturbations (DCS: $\Delta CS_{exp} - \Delta CS_{sim}$) that are produced by the ligand when it binds weakly to a protein surface.[5] We initially attempted to use only DCS restraints to calculate structures of protein−protein complexes without using any NOE information. This approach failed when it was found that DCS restraints alone are effective toward restraining intermolecular translation, but less effective at restricting intermolecular orientations. We report here on the use of DCS and RDC to restrain protein−protein complexes. In anticipation of many high-quality X-ray structures of proteins, we apply our method to dock two X-ray structures, but the method will work equally well to dock high-quality NMR structures.

We used the EIN−HPr complex (3EZA)[6] as a model system for our structure calculations. Crystal structures are available for both EIN(1ZYM)[7] and HPr (1POH);[8] but the complex has not been crystallized. There are NMR structures and assignments for EIN,[9] HPr[10] and the EIN−HPr complex.[6] The rmsd fit of the backbone CA atoms for residues 3−230 of EIN(1ZYM) to EIN(3EZA) is 1.15 Å while the fit of the backbone CA atoms for residues 1−85 of HPr(1POH) to HPr(3EZA) is 0.64 Å. The X-ray structures 1ZYM and 1POH were aligned using the program *DIPOCOUP*.[11] The RDC values for the EIN−HPr complex were obtained from the experimental restraint file 3EZAMR. Alignment using RDC data for the complex establishes the relative orientations of EIN and HPr, but does not restrain intermolecular translation. The fit of the RDC data allows for two EIN−HPr orientations,[4] only one of which places the experimental chemical shift perturbations for EIN and HPr in close proximity. The experimental chemical shift perturbations, prior to RDC alignment are shown in Figure 1A.

To restrain the EIN−HPr complex using chemical shifts, we minimize the difference between experimental and simulated chemical shift perturbations. Simulations of chemical shift differences are made by taking the difference between a reference (EIN and HPr separated by 300 Å) and bound complex (EIN and HPr in close proximity). Protonation of the X-ray PDB files and simulation of proton shifts were performed using *SHIFTS 3.0* (David Case, Scripps). For these calculations, 1ZYM was fixed, and 1POH was moved in 1 Å increments on a Cartesian grid. The initial position of 1POH was ∼12 Å from 1ZYM. Calculations of ΔCS_{sim} were made for 945 HPr positions and compared with experimental ΔCS_{expt} measurements. The experimental ΔCS_{expt} data ranges from −0.38 to +0.37 ppm. The calculated ΔCS_{sim} data has a similar range for the best structures. These are well within the accuracy of SHIFTS calculations that use the Haigh−Mallion method to calculate chemical shift perturbations.[5]

The DCS calculation is a summation of $\Delta CS_{expt} - \Delta CS_{sim}$ for all HN protons of EIN. It is a single number describing the total fit of all of the ΔCS data. DCS values for 945 EIN−HPr structures ranged from 1.71 (best fit) to 6.17 (worst fit). The mean DCS value for all 945 structures is 3.37; 1 standard deviation from the mean DCS value is 0.59; 97% of the DCS values are from 2 to 5. The DCS values for the 12 best structures range from 1.71 to 1.94; the mean structure of the 12 best structures has a DCS of 1.76. A simulated complex where HPr is completely off the EIN surface ($\Delta CS_{sim} = 0$) has a DCS of 3.3. Simulated EIN−HPr complexes that have clear van der Waals violations have typical DCS values of 4−6.

Figure 1, B and C, demonstrates the potential of our approach to rapidly dock protein crystal structures. Figure 1B shows the relative position of the HPr crystal structure in the 12 docked complexes that best fit the experimental perturbation data (DCS is smallest). These 12 "best fit" structures have a 2.1 Å rmsd relative to the mean structure of the ensemble. The mean structure of the ensemble is 2.5 Å from the position determined from the superposi-

* Address correspondence to this author: E-mail: mark.mccoy@spcorp.com.

10.1021/ja017242z CCC: $22.00

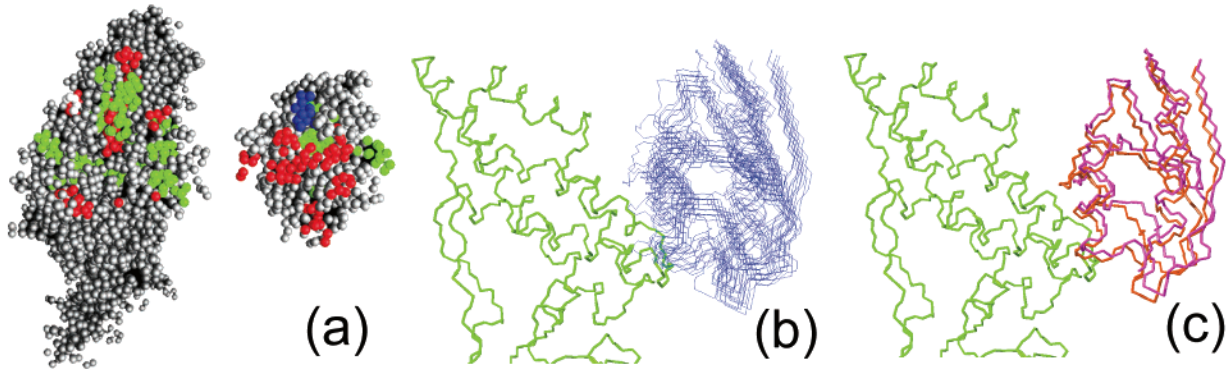

Figure 1. X-ray crystal structures of EIN and HPr are docked using NMR restraints. (A) Experimental proton chemical shift perturbations ($\Delta CS_{expt} = CS_{complex} - CS_{free}$) are mapped onto the structures of EIN (left) and HPr (right). Red and green atoms indicate negative and positive shifts, respectively. Blue atoms are very large shifts; they were not used in these calculations. van der Waals spheres for protons are shown (GRASP[13]). EIN and HPr are neither docked nor aligned with RDC data in Figure A. (B) The 12 HPr structures that best fit the DCS and RDC data are shown in blue using *molmol*.[14] The RDC restraints are satisfied by keeping the relative orientation of EIN (green tube) to HPr fixed. The DCS values are evaluated by taking the normalized difference between experimental and simulated HN chemical shift perturbations of EIN. DCS values were calculated for 945 structures over a (5 Å × 9 Å × 21 Å) grid. The rmsd of the "best fit" ensemble to the mean is 2.1 Å. (C) The rmsd of the mean HPr structure (orange) to the average minimized HPr location from the EIN−HPr complex (magenta) is 2.5 Å. These results demonstrate that RDC and DCS are complementary and, together, can provide restraints in structures of complexes in which intermolecular NOEs may be difficult to obtain.

tion of CA atoms of the EIN and HPr crystal structures on the solution NMR structure of the EIN−HPr complex. In previous models and structures of the EIN−HPr complex, the proximity of His189 (EIN) to His15 (HPR) was restrained on the basis of activity data.[7,11] Significantly, no restraints other than RDC and DCS were used in our calculations.

There are a number of important assumptions that are made in performing our calculations: (1) We assume that there is no substantial change in the backbone structure of EIN or HPr upon complex formation. This can easily be assessed by fitting the RDC data from the EIN−HPr complex to the individual X-ray structures. (2) A simplification in these simulations is that we do not account for any conformational changes in aromatic side chains that may occur upon binding. As a consequence, we minimize $|\Delta CS|_{expt} - |\Delta CS|_{sim}$ to reduce the influence of side chain conformational changes on the sign of the DCS simulations. (3) The mean structure in Figure 1C is calculated on the basis of HN-DCS. DCS values for Hα protons produced similar results. (4) Only ΔCS_{expt} data for EIN residues were used to calculate the structures in Figure 1, B and C. The assignments for the free HPr were obtained by a different laboratory under different conditions and were therefore excluded from our calculations. (5) In all simulations, the major cause of ΔCS is assumed to be the aromatic ring current. Thus, ΔCS(EIN) is caused by aromatic residues on HPr that are in close proximity to EIN. While this is only going to be true for a fraction of the EIN protons, we can reduce the severity of this assumption by restricting DCS restraints to specific atom types. For example HN protons may be problematic since they are strongly perturbed in hydrogen-bond formation. In contrast, selecting aliphatic protons whose dispersion is dictated primarily by aromatic ring current may produce better results.

Here, we have presented evidence that RDC and DCS data alone can be used to dock crystal structures. Better structures will be obtained by including DCS restraints from heteronuclei and when DCS and RDC restraints can be minimized with an accurate force field.[12] At first it may seem to be simple to construct a model of the EIN−HPr complex. One would start by docking the two existing crystal structures (1POH with 1ZYM) using the available biochemical data that requires the His189 of EIN to be in close proximity to the His15 of HPr. In fact, a model was created in just this way, and it was incorrect.[7] The use of chemical shift perturbation and RDC data to restrain protein−protein complexes has significant advantages over traditional approaches; the data is easy to collect and more easily assigned than a large number of NOEs. RDC and DCS data can be obtained even for large protein complexes, and with the exception of intermediate chemical exchange, chemical shift changes are almost always present in complex formation.

Acknowledgment. We acknowledge helpful conversations with David Case, Christian Griesinger, and Marius Clore.

References

(1) Structural Genomics Supplement. *Nat. Struct. Biol.* **2000**, Nov. 7. Teichmann, S. A.; Murzin, A. G.; Chothia, C. *Curr. Opin. Struct. Biol.* **2001**, *11*(3), 354−363.
(2) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732−738.
(3) Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1997**, *4 (Suppl. S)*, 849−853. Wang, G.; Louis, J. M.; Sondej, M.; Seok, Y. J.; Peterkofsky, A.; Clore, G. M. *EMBO J.* **2000**, *19*, 5635−5649.
(4) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021−9025.
(5) McCoy, M. A.; Wyss, D. F. *J. Biomol. NMR* **2000**, *18*, 189−198.
(6) Garrett, D. S.; Seok, Y.-J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1999**, *6*, 166−173
(7) Liao, D.-I.; Silverton, E.; Seok; Y.-J., Peterkofsky, A.; Davies, D. R. *Structure* **1996**, *4*, 861−872.
(8) Jia, Z.; Quail, J. W.; Waygood, E. B.; Delbaere, L. T. *J. Biol. Chem.* **1993**, *268*, 22490−22501.
(9) (a) Garrett, D. S.; Seok, Y.-J.; Liao, D.-I.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1997**, *36*, 6, 2517−2530. (b) Garrett, D. S.; Seok, Y.-J.; Peterkofsky A.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1997**, *36*, 4393−4398.
(10) Van Nuland, N. A. J.; Boelens, R.; Scheek, R. M.; Robillard, G. T. *J. Mol. Biol.* **1995**, *246*, 180−193.
(11) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283−294.
(12) Moyna, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702−709.
(13) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281−296.
(14) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, 51−5, 29−32.

JA017242Z