J|A|C|S
COMMUNICATIONS

Published on Web 08/25/2004

Pose Scoring by NMR

Bing Wang, Kaushik Raha, and Kenneth M. Merz, Jr.*

152 Davey Laboratory, Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

Received April 21, 2004; E-mail: merz@psu.edu

Nuclear magnetic resonance (NMR) spectroscopy is developing into a powerful tool for the characterization of protein−ligand interactions.[1−3] The understanding of these interactions is greatly enhanced if structural information (be it NMR or X-ray) on the protein−ligand complex is available. However, high-resolution NMR structure determination of such complexes is laborious and time-consuming. Moreover, this process can be hampered by poor solubility, low affinity, unfavorable dynamics, and signal overlap. When a small molecule is bound to a protein, the chemical shifts of the protein as well as the ligand can display a wide range of variations, which can be easily detected by NMR experiments. These so-called ligand-induced chemical shift perturbations (CSPs) have been utilized to identify the approximate binding site of a ligand and form the cornerstone of the structure−activity relationship (SAR) by NMR technique.[1] Nevertheless, no information on the orientation of the ligand in the binding site is obtained by this procedure. To overcome this deficiency, Medek et al.[4] have used differences in CSPs induced by a series of ligand analogues to determine the precise location of a ligand binding site and the orientation of the ligand in the binding pocket.

Theoretical determination of ligand-induced chemical shift perturbations can provide detailed insights into protein−ligand interaction at the molecular level. McCoy and Wyss[5] have set up a protocol to align a model molecule to a protein surface by simulating proton chemical shift perturbations on the basis of empirical formulas. Recently, we have developed a fast approach[6] that calculates NMR chemical shifts using gauge-including atomic orbitals (GIAO)[7,8] at the semiempirical MNDO[9] level. The perturbed density matrix with respect to the magnetic field is obtained by the diagonalization of the complex Fock matrix using the divide and conquer (D&C) approach.[10−12] An atom pair prescreening scheme was employed to reduce the cost of the magnetic integral calculations. Using semiempirical parameters developed by Patchkovskii and Thiel,[13] this approach was able to carry out NMR chemical shift calculations for systems with thousands of atoms to good accuracy.[6] To demonstrate the ability of this approach to characterize protein/ligand systems, we describe a procedure that uses chemical shift perturbations to score different poses (i.e., different orientations of a molecule in a binding pocket) of GPI-1046 (henceforth GPI; see Figure 1) bound to the FK506 binding protein (FKBP12).

GPI is effective at inhibiting the peptidyl-prolyl cis−trans isomerase (PPIase) activity of FKBP12 and in promoting nerve regeneration both in vitro and in vivo.[14] Ten NMR structures (GPI1 to GPI10) of this complex have been described by Sich et al. (PDB 1F40).[15] The geometries of the free ligand and the complexes were optimized at the AM1[16] level, followed by NMR chemical shift calculations for both the free and bound ligand at the MNDO level.[6] As shown by Figure 2, excellent agreement between the calculated proton chemical shifts with experiment was obtained. Strong upfield shifts for protons on the pyrrolidine ring upon binding are also

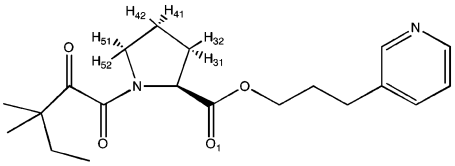

Figure 1. Structure of GPI-1046.

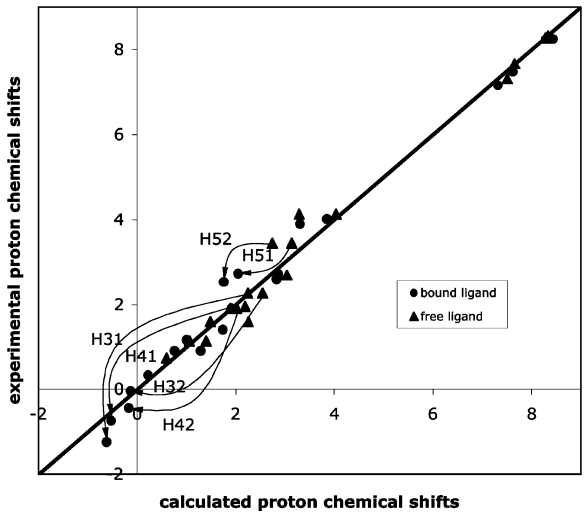

Figure 2. Plot of the calculated versus experimental proton chemical shifts (in ppm) for both the free and bound ligand. Only data for GPI6 are shown, and the correlation coefficients R^2 are 0.981 and 0.986, respectively. The binding-induced upfield-shifted protons in the pyrrolidine ring are labeled.

revealed, which reconfirms the experimental results that place this ring into the hydrophobic pocket formed by the aromatic residues Tyr26, Phe46, Trp59, and Phe99. This aromatic shielding effect is highly sensitive to the relative positioning of the interacting group(s). Therefore, the chemical shift perturbation values can be used to determine the orientation of the ligand and evaluate the quality of the experimental structures. Because CSPs are relative values, they benefit theoretical approaches by masking systematic error. However, CSPs are very sensitive to the local environment and are relatively small in magnitude, both of which could pose a challenge for a theoretical approach to CSP determination.

Table 1 lists the root-mean-square deviations (RMSDs) between the calculated and experimental proton CSPs for the 10 NMR structures. RMSDs for GPI5, GPI6, GPI8, and GPI10 are below 0.3 ppm, which we interpret as indicating that they are closer to the native structure. Table 1 also lists distances between the O1 carbonyl oxygen and the backbone N−H in Ile56. This clearly shows that there is a hydrogen bond between the structures with low CSP RMSDs (below 0.5 ppm), but not for the structures with high CSP RMSDs (above 0.5 ppm). This hydrogen bond, which was observed experimentally, is not present in all of the NMR structures due to the dearth of NMR restraints. Our results confirm

10.1021/ja047695e CCC: $27.50

Table 1. Chemical Shift Perturbation RMSDs (in ppm) and Distances (in Å) between O1 and HN in Ile56 for the 10 Experimentally Determined NMR Structures

	CSP RMSD	O1···HN(Ile56)		CSP RMSD	O1···HN(Ile56)
GPI1	0.621	4.70	GPI6	0.233	2.12
GPI2	0.660	5.21	GPI7	0.743	4.92
GPI3	0.500	5.08	GPI8	0.276	2.13
GPI4	0.807	5.15	GPI9	0.439	2.16
GPI5	0.211	2.17	GPI10	0.277	2.14

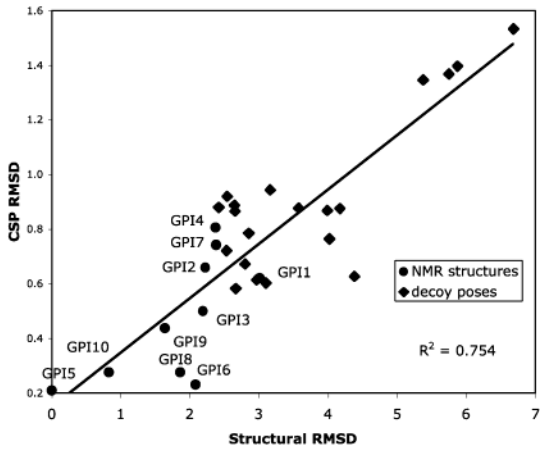

Figure 3. Correlation plot of the chemical shift perturbation RMSDs (in ppm) versus the structural RMSDs (in Å) with respect to GPI5 for nine GPI NMR structures and 20 computationally generated structures.

the presence of this hydrogen bond in the native states of the complex and suggest that it plays a key role in GPI binding. Interestingly, this hydrogen bond interaction has its equivalent in other FKBP12 complexes.[17]

To test our approach's ability to select native binding states from non-native poses, we generated 20 decoy poses by manual rotation around torsional angles of GPI and translation inside the binding pocket and again calculated the NMR chemical shifts. From this analysis we obtained a good correlation between the CSP RMSDs and the structural RMSDs for all atoms with respect to GPI5 (selected as the "best" NMR structure due to its lowest CSP RMSD relative to experiment). The result of this analysis is shown in Figure 3, where the non-native poses and the NMR structures ($R^2 = 0.754$) are correlated relative to GPI5. The relatively large structural RMSD between GPI5 and GPI6 (the second lowest CSP RMSD) arises due to the orientation of the pyridine ring, which is simply rotated along the $-CH_2-C$(aromatic) bond in the two structures. Another low CSP RMSD structure, GPI8, also adopts GPI6's conformation, as indicated by the low structural RMSD between these two structures (1.18 Å). These findings indicate that the pyridine ring can adopt two conformations due to the absence of a hydrogen bond interaction between the pyridine N and the protein surface that would add specificity to the pyridine/protein interaction. Regardless of the pyridine conformation, the NMR observations are reproduced.[15] However, the placement of the rest of the molecule is much more sensitive to the local environment. Overall, the NMR structures (which were fit to satisfy the observed NOE constraints) generally gave better agreement with the observed CSPs than did the decoy poses. The five structures (GPI1−GPI4 and GPI7) that did not have the O1···HN hydrogen bond with Ile56 were clustered with the decoy poses, while the five that showed the presence of this hydrogen bond were clearly differentiated from the decoy poses. Hence, we conclude that our approach based on computed CSPs can readily differentiate between correctly docked versus incorrectly docked structures.

The ability to compare experimental and computed CSPs to differentiate between correctly positioned versus incorrectly positioned ligands in a binding pocket has significant ramifications. Thus, one could envision carrying out a computational docking study of a molecule in a binding pocket to generate a set of candidate poses. We can then computationally determine the CSPs for all of the poses, which when compared with experimental would allow for the selection of the correct binding mode of the ligand. The key aspect of this analysis is that we are able to obtain the correct binding mode just through the examination of CSPs of the ligand.

In conclusion, our results have demonstrated that the correct binding orientation of a ligand can be determined by a simple comparison of calculated and experimental chemical shift perturbations for the ligand. We have also shown that deviations in the computed CSPs from experiment offer a straightforward manner in which to score different poses for the ligand.

Acknowledgment. We thank Dr. Christian Sich for providing us with the proton chemical shifts of free GPI. We also thank the NCSA for supercomputer time, the NSF (MCB-0211639), and the NIH (GM44974) for support.

References

(1) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531−1534.
(2) Stockman, B. J. *Prog. NMR Spectrosc.* **1998**, *33*, 109−151.
(3) Zuiderweg, E. R. *Biochemistry* **2002**, *41*, 1−7.
(4) Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W. *J. Am. Chem. Soc.* **2000**, *122*, 1241−1242.
(5) McCoy, M. A.; Wyss, D. F. *J. Biomol. NMR* **2000**, *18*, 189−198.
(6) Wang, B.; Brothers, E.; van der Vaart, A.; Merz, K. M. *J. Chem. Phys.* **2004**, *120*, 11392−11400.
(7) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789−807.
(8) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260.
(9) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(10) Yang, W. T.; Lee, T. S. *J. Chem. Phys.* **1995**, *103*, 5674−5678.
(11) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1996**, *104*, 6643−6649.
(12) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1997**, *107*, 879−893.
(13) Patchkovskii, S.; Thiel, W. *J. Comput. Chem.* **1999**, *20*, 1220−1245.
(14) Steiner, J. P.; Hamilton, G. S.; Ross, D. T.; Valentine, H. L.; Guo, H. Z.; Connolly, M. A.; Liang, S.; Ramsey, C.; Li, J. H. J.; Huang, W.; Howorth, P.; Soni, R.; Fuller, M.; Sauer, H.; Nowotnik, A. C.; Suzdak, P. D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2019−2024.
(15) Sich, C.; Improta, S.; Cowley, D. J.; Guenet, C.; Merly, J. P.; Teufel, M.; Saudek, V. *Eur. J. Biochem.* **2000**, *267*, 5342−5354.
(16) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(17) Sirockin, F.; Sich, C.; Improta, S.; Schaefer, M.; Saudek, V.; Froloff, N.; Karplus, M.; Dejaegere, A. *J. Am. Chem. Soc.* **2002**, *124*, 11073−11084.

JA047695E