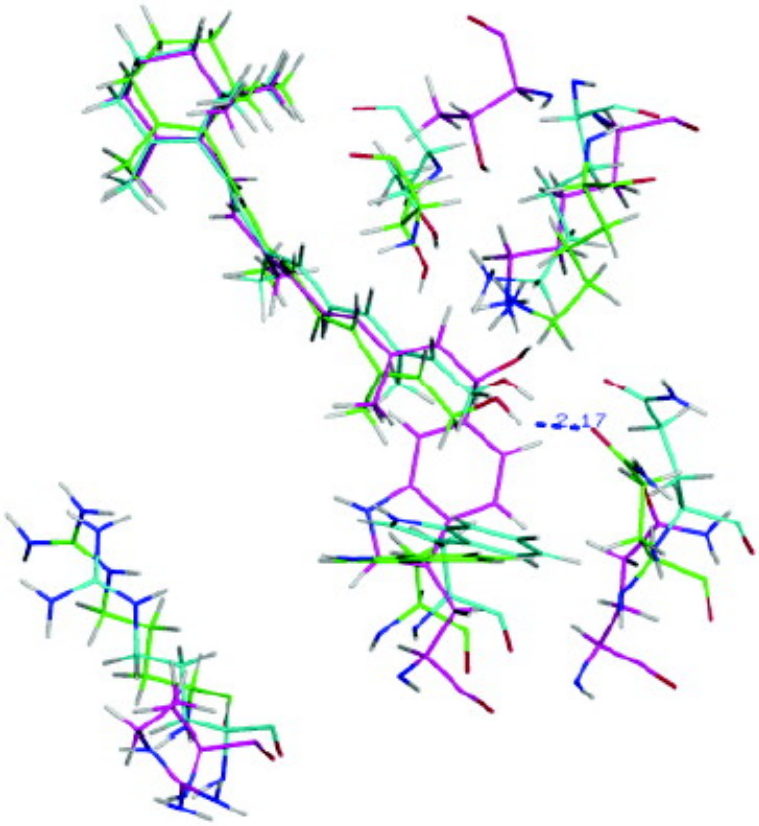
Communication

Validation of the Binding Site Structure of the Cellular Retinol-Binding Protein (CRBP) by Ligand NMR Chemical Shift Perturbations

Bing Wang, and Kenneth M. Merz

J. Am. Chem. Soc., **2005**, 127 (15), 5310-5311• DOI: 10.1021/ja042616k • Publication Date (Web): 26 March 2005



J|A|C|S
COMMUNICATIONS

Published on Web 03/26/2005

Validation of the Binding Site Structure of the Cellular Retinol-Binding Protein (CRBP) by Ligand NMR Chemical Shift Perturbations

Bing Wang and Kenneth M. Merz Jr.*

Department of Chemistry, The Pennsylvania State University, 104 Chemistry Building, University Park, Pennsylvania 16802

Received December 8, 2004; E-mail: merz@psu.edu

The binding of a small molecule to a protein may cause chemical shift changes on the protein as well as on the ligand. These chemical shift perturbations (CSPs) contain valuable information about protein−ligand interactions. For example, the binding surface can be mapped from a weighted average of ^{1}HN and ^{15}N chemical shift changes obtained in the ^{15}N/^{1}H Heteronuclear Single Quantum Coherence (HSQC) spectra. This technique has been routinely used in drug discovery applications such as the SAR by NMR[1] approach. Since NMR chemical shifts are sensitive to variations in molecular structure, theoretical calculation of these CSPs will enhance our understanding of the relationship between chemical shifts and the structure of protein−ligand complexes. McCoy and Wyss[2] have aligned small molecules to protein surfaces by a comparison of the experimental CSPs with the calculated values based on an empirical approach.[3] Recently, we have developed a fast approach[4] to calculate NMR chemical shifts using the divide and conquer (D&C)[5−7] method at the modified neglect of diatomic overlap (MNDO)[8] level of theory. This approach is able to handle thousands of atoms and provides accurate ^{1}H and ^{13}C chemical shifts. Using this approach, we have carried out proton chemical shift calculations for the FK506 binding protein (FKBP) in complex with a small molecule.[9] Excellent agreement between the experimental and calculated ligand CSPs was obtained for the native binding structures. Moreover, we were able to differentiate them from a set of decoy poses by CSP root-mean-square deviations (RMSD) from experiment.

Another potential application of this approach is the validation of protein−ligand complexes obtained from NMR studies. NMR structures are usually an ensemble of structures, which are best fit to experimental restraints derived from Nuclear Overhauser Effects (NOE) and spin−spin coupling constants. However, the quality of these structures is affected by the availability of restraints and the errors in these restraints due to signal mis-assignment and/or ambiguities. It is more difficult to extract these restraints for protein−ligand complexes since their interactions are generally noncovalent. Therefore, it is important to validate these complexes through the use of other NMR parameters, such as chemical shifts. In this paper, we evaluate the quality of NMR structures for the cellular retinol-binding protein (CRBP) based on CSPs.

CRBP is one of the intracellular carriers of the hydrophobic molecule retinol (see Figure 1) within the aqueous environment of the cytosol. A 2.0 Å X-ray structure of its *holo* form was solved in 1993 (PDB code 1CRB[10]). Recently, two groups have independently reported NMR solution structures of CRBP (PDB code 1KGL[11] and 1MX8[12], respectively). All of these structures are 10-stranded antiparallel β-barrels with a large binding cavity. Two α-helices cap the open end of the barrel, making it nearly solvent inaccessible. However, the structure of retinol and its orientation relative to residues inside the binding site for the NMR structures show a wide spectrum of deviations from that of the X-ray structure. Note that the measured proton chemical shifts for bound retinol are almost identical in 1KGL and 1MX8 (BioMagResBank accession numbers 5319 and 5578), although different experimental conditions were used. The experimental values of free retinol were taken from the previously published data.[13] To validate these NMR structures, we have calculated proton chemical shifts of the free and bound retinol using our approach after AM1[14] geometry optimization.

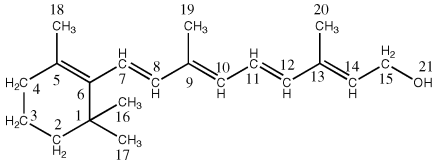


Figure 1. The structure of retinol.

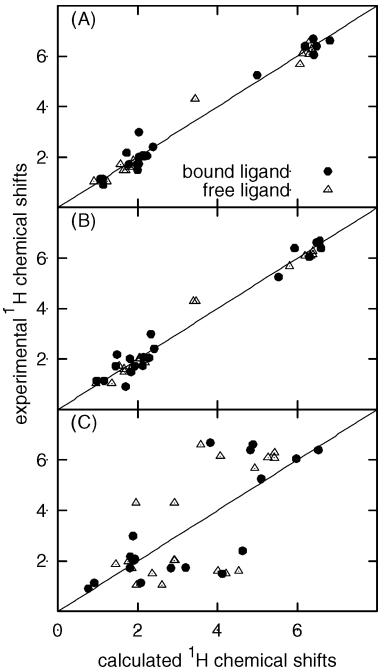


Figure 2. Plot of the measured versus the calculated proton chemical shifts (in ppm) of bound and free retinol for three different CRBP structures: (A) 1CRB, (B) 1KGL, and (C) 1MX8. The correlation coefficients for bound and free retinol are (A) 0.977 and 0.974; (B) 0.971 and 0.972; and (C) 0.582 and 0.384, respectively.

Figure 2A plots the correlation between the calculated and measured proton chemical shifts (excluding hydroxyl proton) for free and bound retinol in the X-ray structure. Figure 2B,C plots one representative model for the two NMR structure ensembles. Experimentally, the end of the isoprene tail experiences large chemical shift perturbations: −2.1 and −1.3 ppm for H15a and H15b, respectively, −0.94 ppm for H20, and −0.42 ppm for H14, which are mainly induced by Trp106. This aromatic shielding effect is extremely sensitive to the relative position of the interacting groups and, hence, provides a useful tool to validate the quality of

10.1021/ja042616k CCC: $30.25

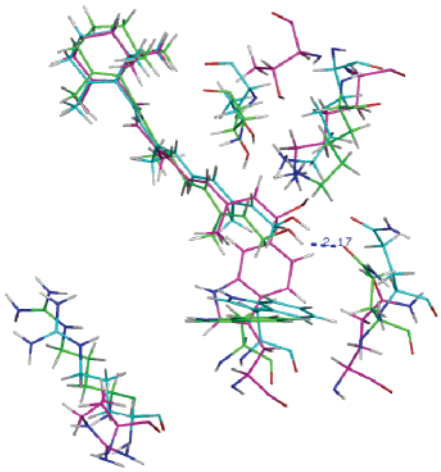


Figure 3. Comparison of retinol and the side chain orientations of residues inside the binding pocket for the three CRBP structures: 1CRB (green), 1KGL (cyan), and 1MX8 (magenta).

NMR structures for the binding pocket. Our theoretical results capture all of these effects very well for the X-ray structure; the correlation coefficients for both the bound and free ligand are above 0.97, and the RMSD between calculated and experimental CSPs for all protons is 0.23 ppm. This excellent agreement demonstrates that our approach is capable of accurately calculating proton chemical shifts for protein−ligand complexes and confirms that the X-ray structure of the binding site is relevant in aqueous solution. For 1KGL, similar agreement was obtained with a CSP RMSD of 0.26 ppm. However, a very poor correlation was found for 1MX8 with a CSP RMSD of 1.29 ppm.

The binding sites for these three structures are shown in Figure 3. The 1KGL structure resembles the 1CRB structure; the bound retinol has a planar conformation with its hydroxyl group hydrogen-bonded to the side chain of Gln108. However, the 1MX8 structure deviates significantly from the X-ray and 1KGL structures, with distortion of the planarity of the isoprene unit and the lack of a hydrogen bond with Gln108. Moreover, the side chains of the residues within the binding pocket of 1MX8 have different orientations from those observed in the other two structures. These structural deviations in 1MX8[12] are responsible for the large deviations of calculated CSPs from experiment, especially at protons on the isoprene chain and the β-ionone ring. The experimental NMR measurables for 1MX8 did not include any restraints for the Gln108 hydrogen bond, and the minimalist penalty function did not enforce the isoprene unit planarity. However, the distance geometry/simulated annealing algorithm, which was used in the process of 1MX8 structure determination, provides improved sampling of the protein−ligand conformational space consistent with the input NMR data. The 1MX8 structures may be a more realistic representation of the input NMR NOE data. Nonetheless, our CSP analysis shows that these structures would benefit from additional restraints and/or further refinement from chemical shift information.

To further test our approach, we also calculated CSPs for all structural models in the 1KGL ensemble. Figure 4A plots the

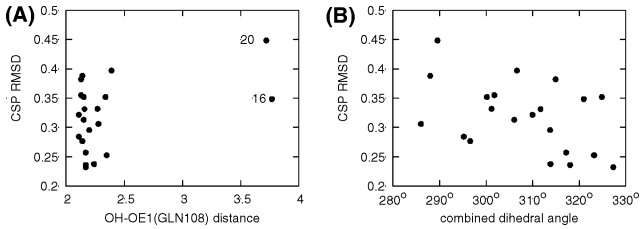


Figure 4. Plot of CSP RMSDs versus the distance (in Å) between the hydroxyl group of retinol and the OE1 atom of Gln108 (A) and the combined dihedral angles (see text) (B).

distance between the hydroxyl hydrogen atom of retinol and the OE1 atom of Gln108 against the CSP RMSDs for the 20 models of 1KGL. The lack of this key hydrogen bond in models 16 and 20 leads to an increase in the computed CSP RMSDs. To test the planarity of the isoprene chain, we selected two dihedral angles: C7−C8−C9−C10 and C11−C12−C13−C14. As shown in Figure 4B, there is a tendency toward lower CSP RMSDs as these two dihedrals become more planar. Overall, via a comparison between experimental and computed CSPs of the ligand only, we were able to identify NMR structural models that were less likely than others. If we averaged proton chemical shifts of retinol for all 1KGL models (except models 16 and 20), the final CSP RMSD is 0.23.

In summary, our approach was able to accurately calculate proton chemical shifts for free and bound retinol in CRBP. Similar results have been obtained for CRBP II, which will be presented in a future publication. These CSPs are very sensitive to the planarity and orientation of retinol within the binding pocket of CRBP. In general, this information is not included in the NMR structure refinement process; however, we find that the CSP RMSD offers a straightforward metric to evaluate the quality or degree of refinement of NMR structures for a ligand found within a protein binding site.

Acknowledgment. We thank the NSF (MCB-0211639) and the NIH (GM 44974) for support. We also thank Dr. David Cistola, Dr. Jianyun Lu, and Dr. Ellen Li for their comments.

References

(1) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531−1534.
(2) McCoy, M. A.; Wyss, D. F. *J. Biomol. NMR* **2000**, *18*, 189−198.
(3) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436−9444.
(4) Wang, B.; Brothers, E. N.; Van Der Vaart, A.; Merz, K. M. *J. Chem. Phys.* **2004**, *120*, 11392−11400.
(5) Yang, W. T.; Lee, T. S. *J. Chem. Phys.* **1995**, *103*, 5674−5678.
(6) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1996**, *104*, 6643−6649.
(7) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1997**, *107*, 879−893.
(8) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(9) Wang, B.; Raha, K.; Merz, K. M. *J. Am. Chem. Soc.* **2004**, *126*, 11430−11431.
(10) Cowan, S. W.; Newcomer, M. E.; Jones, T. A. *J. Mol. Biol.* **1993**, *230*, 1225−1246.
(11) Franzoni, L.; Lucke, C.; Perez, C.; Cavazzini, D.; Rademacher, M.; Ludwig, C.; Spisni, A.; Rossi, G. L.; Ruterjans, H. *J. Biol. Chem.* **2002**, *277*, 21983−21997.
(12) Lu, J.; Cistola, D. P.; Li, E. *J. Mol. Biol.* **2003**, *330*, 799−812.
(13) Liu, R. S. H.; Asato, A. E. *Tetrahedron* **1984**, *40*, 1931−1969.
(14) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

JA042616K