J|A|C|S
ARTICLES
Published on Web 09/06/2002

Spatial Localization of Ligand Binding Sites from Electron Current Density Surfaces Calculated from NMR Chemical Shift Perturbations

Mark A. McCoy* and Daniel F. Wyss

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

Received March 11, 2002

Abstract: Rapid, accurate structure determination of protein−ligand complexes is an essential component in structure-based drug design. We have developed a method that uses NMR protein chemical shift perturbations to spatially localize a ligand when it is complexed with a protein. Chemical shift perturbations on the protein arise primarily from the close proximity of electron current density from the ligand. In our approach the location of the center of the electron current density for a ligand aromatic ring was approximated by a point-dipole, and dot densities were used to represent ligand positions that are allowed by the experimental data. The dot density is increased in the region of space that is consistent for the most data. A surface can be formed in regions of the highest dot density that correlates to the center of the ligand aromatic ring. These surfaces allow for the rapid evaluation of ligand binding, which is demonstrated on a model system and on real data from HCV NS3 protease and HCV NS3 helicase, where the location of ligand binding can be compared to that obtained from difference electron density from X-ray crystallography.

Introduction

The rapid calculation of high-quality X-ray structures of protein−ligand complexes is a central tool in rational drug design.[1] These structures are most easily calculated using difference maps of Fourier electron density.[2] For protein−ligand complexes, this requires the collection of diffraction data for protein crystals with and without bound ligands. High-quality electron difference maps clearly show the ligand electron density from which X-ray structures of protein−ligand complexes can be refined. NMR chemical shift perturbations, which are also used to study ligand binding to proteins, are produced by ligand electrons. Rather than electron density, it is the flow of electrons (electron current density, *j*) that is responsible for NMR chemical shifts and chemical shift perturbations.[3,4] The close proximity of electron current density from the ligand causes chemical shift perturbations in protein atoms that can broadly indicate the ligand interaction site on the protein surface. In this article, we explore the use of NMR chemical shift perturbations to localize ligand electron current density from which surfaces can be constructed. We demonstrate that these ligand surfaces can define the Cartesian coordinates of protein-bound ligands.

NMR chemical shifts are produced by electron currents induced in a molecule by a static magnetic field.[3,4] The flow of electrons, measured by the current density, is strongly influenced by the conjugation of the bonds in the molecule. The generation of local magnetic fields by flowing electrons is the source of the dispersion of chemical shifts in a typical one-dimensional NMR spectrum. The electron current density can be calculated for simple organic molecules using ab initio and DFT methods, from which shielding tensors and isotropic chemical shifts can be accurately determined.[5,6] Electrons are also responsible for so-called secondary shifts, or chemical shift perturbations. Chemical shift perturbations (ΔCS) are produced, for example, when small molecules bind to proteins. ΔCS maps have been used extensively to locate ligand interaction sites on protein surfaces.[7−9] Chemical shift perturbations are also fundamental to the SAR-by-NMR[10] method that is used to screen for small-molecule ligands which has become an important tool in structure-based drug discovery and design. ΔCS restraints have been used to calculate structures of protein bound peptides[11] and, recently, to dock structures of protein−ligand[12,13] and protein−protein complexes.[14−16] While methods exist to mini-

* To whom correspondence should be addressed. E-mail: mark.mccoy@spcorp.com.

(1) Blundell, T. L. *Nature* **1996**, *384*, 23−26.
(2) Blundell, T. L.; Johnson, L. N. *Protein Crystallography*; Academic Press: London, 1990; Chapter 14.
(3) Slichter, C. P. *Principles of Magnetic Resonance*, 2nd revised edition; Springer-Verlag: Berlin, 1980; Chapter 4.
(4) Jardetzky, O.; Roberts, G. C. K. *NMR in Molecular Biology;* Academic Press: New York, 1981.
(5) Case, D. A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 197−203.
(6) Case, D. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 624−630.
(7) Zuiderweg, E. R. *Biochemistry* **2002**, *41*, 1−7.
(8) Markus, M. A.; Nakayama, T.; Matsudaira, P.; Wagner, G. *Protein Sci.* **1994**, *3*, 70−81.
(9) Rajagopal, P.; Waygood, E. B.; Reizer, J.; Saier, M. H., Jr.; Klevit, R. E. *Protein Sci.* **1997**, *12*, 2624−2627.
(10) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531−1534.
(11) Moyna, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702−709.
(12) McCoy, M. A.; Wyss, D. F. *J. Biomol. NMR* **2000**, *18*, 189−198.
(13) Lugovskoy, A. A.; Degterev, A. I.; Fahmy, A. F.; Zhou, P.; Gross, J. D.; Yuan J, Wagner, G. *J. Am. Chem. Soc.* **2002**, *124*, 1234−1240.
(14) Morelli, X.; Dolla, A.; Czjzek, M.; Palma, P. N.; Blasco, F.; Krippahl, L.; Moura, J. J.; Guerlesquin, F. *Biochemistry* **2000**, *39*, 2530−2537.
(15) Morelli, X. J.; Palma, P. N.; Guerlesquin, F.; Rigby, A. C. *Protein Sci.* **2001**, *10*, 2131−2137.
(16) McCoy, M. A.; Wyss, D. F. *J. Am. Chem. Soc.* **2002**, *124*, 2104−2105.

10.1021/ja026166c CCC: $22.00

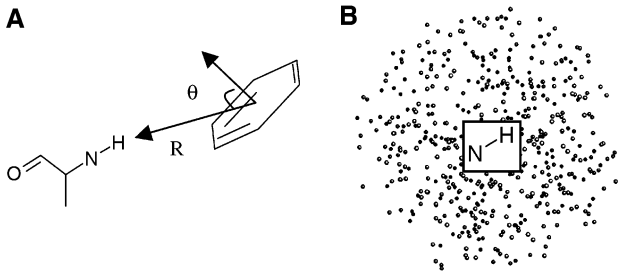

Figure 1. (A) Size and sign of a chemical shift perturbation observed at a HN proton of a protein depend on the distance R and angle θ from the ring center of the ligand. (B) Probability of finding the center of current density from an aromatic ring is represented by a sphere with radius R_{max} of N dots surrounding a HN atom for which a chemical shift perturbation ΔCS is observed. The dot density is distributed randomly throughout the sphere.

mize the ligand position to best fit observed protein chemical shift perturbations,[12] these calculations rely on good data, single-site binding, and little or no protein rearrangement. The ligand surfaces that we describe and calculate in the article are useful in evaluating the consistency and quality of protein-detected chemical shift perturbations. Ligand surfaces can be used to determine which chemical shift perturbations are consistent with the close proximity of a ligand aromatic ring. These calculations are rapid and are intended to identify data that is suitable for minimization using chemical shifts as restraints. Ligand surfaces provide a visual summary of the chemical shift perturbation data used to restrain the protein−ligand complex in analogy with the way that difference electron density is used in X-ray crystallography.

Methods

We will focus our discussion on protein−ligand complexes where frequency shifts of the protein protons are caused by ligand electrons that are in close proximity to NMR-detected nuclei, typically ^{1}HN. If we assume that ΔCS(i), the chemical shift perturbation detected at protein proton atom i, is generated primarily by electron currents from an aromatic ring in the ligand, we can approximate the origin of the ΔCS as a single point-dipole located at the center of the ring.[11]

$$\Delta\mathrm{CS}(i) = \mathrm{CS}(i)_{\mathrm{P+L}} - \mathrm{CS}(i)_{\mathrm{P}} \approx \left(\frac{B_{\mathrm{dip}}}{R_i^{\,3}}\right)(1 - 3\cos^2\theta_i) \qquad (1)$$

In eq 1, P + L and P denote ligand-bound and ligand-free protein, respectively; R_i is the length of the vector from the center of the ligand ring to the perturbed protein atom i; θ_i is the angle made by the vector normal to the ring plane and the vector R_i; B_{dip} is a proportionality constant. R_i and θ_i are shown in Figure 1A. The point-dipole can take on a continuum of values of R and θ that satisfy a specific ΔCS(i) shift. The explicit value of ΔCS(i), however, places constraints on R_i and θ_i that can be exploited. Explicitly, the sign of ΔCS constrains θ; for ΔCS < 0, 0° < θ < 54.7°; 180° < θ < 180°−54.7° and for ΔCS > 0, 54.7° < θ < 180°−54.7°. In eq 2 the size of ΔCS constrains R to a maximum value R_{max}.

Dot Density Representation. From only one ΔCS(i) value, the probability of finding the location of the ligand ring is restricted to a sphere with a radius ranging from 0 to R_{max} centered at the perturbed protein atom i. (note: ΔCS = +0.2 ppm constrains R_{max} to a 4.8 Å sphere, while ΔCS = −0.2 ppm constrains R_{max} to a 6.0 Å sphere). In Figure 1B, we represent N allowed locations for a single ΔCS(i) as a sphere of N dots that surround the perturbed protein HN atom. The dot representation is similar to that which is used for visualizing atomic and molecular orbitals. Each dot represents the coordinates of the ligand ring that is consistent with eq 1 with boundaries placed by eq 2.

$$\Delta\mathrm{CS}(i) > 0, R_{\mathrm{max}}(i) = \left(\frac{B_{\mathrm{dip}}}{\Delta\mathrm{CS}(i)}\right)^{1/3};$$

$$\Delta\mathrm{CS}(i) < 0, R_{\mathrm{max}}(i) = \left(\frac{-2B_{\mathrm{dip}}}{\Delta\mathrm{CS}(i)}\right)^{1/3} \qquad (2)$$

For simplicity, the dots are randomly distributed. For ΔCS(i) for a single protein atom i the probability density is given by

$$\mathrm{PD}_i = \left(\frac{1}{V_i}\right)\sum_N P_i$$

where P_i is the probability that a single dot is the correct ligand location; V_i is the volume of the N dot sphere for protein atom i. It is not possible to define a single ligand ring location from one ΔCS(i). Instead, we rely on redundant information (many atoms are perturbed by the same ligand ring). For multiple sites of perturbation, the dot spheres for each perturbed protein atom i contain the same number of dots. Each sphere has a different volume that depends on R_{max}. Since PD (and dot density) scales with the inverse of the volume, stronger perturbations will have dot spheres that are more localized (and more dense) than weak perturbations. By summing over each perturbed atom, $\sum^{\mathrm{I}}\mathrm{PD}_i$, we increase the dot density in regions of space where the data are consistent and therefore overlap. Note that a normalized probability density is not used since it is likely that (1) not all of the perturbation data is from a single ligand ring and (2) only a fraction of the possible perturbations will actually be observed due to NMR peak overlap, incomplete assignments, and so forth. Instead, we use a dot density approach, which is proportional to an increase in probability density.

Some simple programs (that can be obtained from M.M.) were written to calculate the dot density of $\sum^{\mathrm{I}}\mathrm{PD}_i$. One program randomly distributes N dots in a sphere centered at the coordinates for every ΔCS(i) using the boundaries set by R_{max}. This program requires only a pdb file and a list of chemical shift perturbations as input. It extracts the coordinates of each perturbed protein atom (which are usually HN's) and creates a single pdb file with $I \times N$ dots. Each dot represents an allowed ring position. There are I spheres centered on each perturbed HN. Each sphere has N dots; typically N is 500−2000. Another program calculates dot densities for each point in this pdb file and appends this value (typically 1−100 dots/Å^3) to the 10th field of the pdb file. An awk script is used to calculate the mean and standard deviation (σ) of the dot density and writes the coordinates of dots with densities exceeding certain thresholds (typically 2−4σ) into separate pdb files. Surfaces of dots at various σ levels are easily calculated using GRASP[17] typically with an atom radius of 0.5 Å and surface probe radius of 1.0 Å. Surfaces created from dots with high densities and high σ can localize the ligand ring in a manner that is consistent with the existing data. This is shown graphically in Figure 2 where dot spheres from i = 2 and i = 3 perturbed protein atoms overlap at the ligand ring location. Care must be taken to create surfaces at σ levels that exceed the dot density of nonoverlapping dots. Surfaces created from regions of high dot density are referred to as "j-surfaces", since the electron current density j has the greatest probability of being localized within these surfaces. The accuracy of the j-surfaces will depend on the number and quality of the ΔCS values, the composition of the ligand and the geometry of the protein binding site. Surfaces that are constructed from high density and high σ dots are consistent with the experimental shifts and are spatially localized at different σ. If the data is inconsistent (nonoverlapping dots), sparse (i = 2,3) or weak (|ΔCS| < 0.05) only low σ surfaces can be formed all of which will be diffuse.

The dot density method relies on the overlap of dot spheres. In special cases (when the perturbed protein atoms are at 0°, 90°, 180°, or 270° relative to the ligand ring) the dot density extends only to the center of

(17) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281−296.

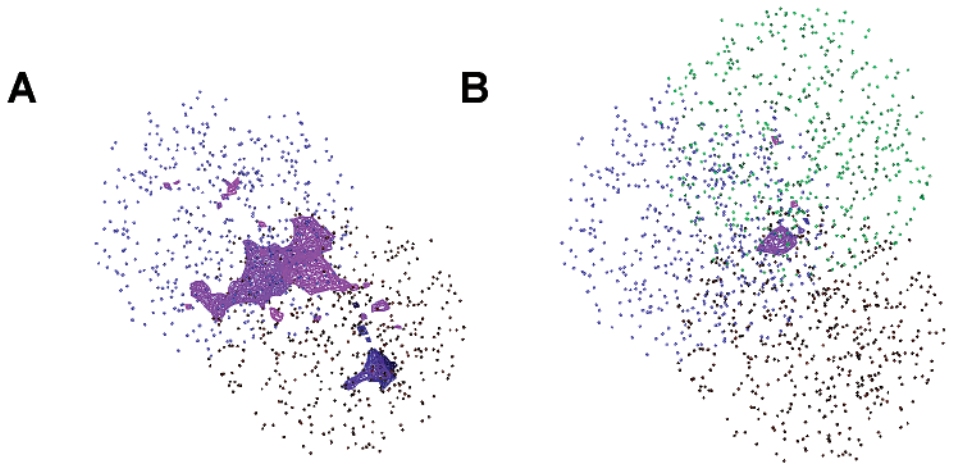

Figure 2. Overlap of dot spheres derived from (A) two and (B) three perturbed atoms localizes the center of the current density that causes the perturbation. Current density surfaces (magenta) are drawn around dots with high density. Ligand current density surfaces (*j*-surfaces) should be localized if the experimental chemical shift data is consistent and if the size and number of perturbations are sufficient.

the ring and may not have extensive overlap. If R_{max} is extended slightly by 1.1 Å, then dot spheres from consistent perturbation data will overlap independent of θ. Even though this adjustment is necessary only for these special cases of $R = R_{max}$, it should be applied to all perturbation data since there is no a priori knowledge of θ.

The method that we have outlined for calculating ligand *j*-surfaces uses a point-dipole approximation for the ring current calculated from a single benzene ring. The effect of ring substitution on the ring current will likely be small and since $R_{max} \approx (B_{dip})^{1/3}$, small changes in B_{dip} will have little consequence to R_{max}. The current methodology can be used on ligands with multiple rings. If the rings are separated by more than ~5Å, then there are no errors as the perturbation patterns are largely independent. If the rings are fused, then errors will result when perturbations have contributions from each ring. In this case, the *j*-surface will be formed primarily from protein atoms that are perturbed by one of the rings. Equation 1 can be modified to reduce errors from perturbations with multiple sources. This can be done by the inclusion of multiple radii, which unfortunately, introduces errors for perturbations from a single source. The method we have outlined in this section relies on the existence of redundant information from multiple points of detection on the protein. The accuracy of the method does not so much rely on the accuracy of the R_{max} calculation, but rather on the redundancy of the chemical shift perturbation data to create regions of high dot density. The diffuse pattern of the aromatic ring current is well suited for such an approach.

Results and Discussion

Simulated Ligand Binding. The accuracy of the calculation of ligand *j*-surfaces was first assessed on a synthetic data set. All simulations were performed with SHIFTS 3.0[18] using the Haigh−Mallion method. A single phenylalanine was used as a ligand mimic and was placed in the center of a protein mimic; a $4 \times 4 \times 5$ three-dimensional grid of glycine residues separated by 4 Å. 4 Gly residues from the grid center were removed so as not to overlap the Phe. Amide proton chemical shifts of the remaining 76 Gly residues were calculated with and without the Phe to simulate bound and unbound chemical shifts. The nine largest chemical shift perturbations, shown in Figure 3A, ranged from −0.295 to +0.238 ppm. The nine perturbations in Figure 3A were used to calculate the ligand *j*-surface for Phe binding that is shown in Figure 3B. 2000 dots were calculated for each perturbation; the dot density was calculated over a 1 Å^3 volume element. For this data set of $I \times N = 18000$ dots, the maximum, mean and 1 standard deviation of the density are, respectively, 105, 27, and 16 dots/Å^3. Figure 3B shows the "bound" location of the Phe with a surface drawn around dots that have a density of greater than 80 dots/Å^3 which is the density at 3.3σ from the mean. We calculated the "noise ceiling", the highest dot density of nonoverlapped dot spheres, to be 25 dots/Å^3. As discussed in the Methods section, the current methodology can be used on ligands with multiple rings with no modifications. This is demonstrated in Figure 3C which is similar to Figure 3B, only with a Trp as the ligand mimic. *j*-Surfaces for both Phe and Trp accurately locate the ligand location using ΔCS from "protein" HN atoms. The *j*-surfaces in Figure 3, parts B and C, are asymmetric. A symmetric surface should result if a fine grid spacing were used, the Phe were perfectly centered and rotated, and the number of dots in each sphere were large. Instead, we have used a practical grid spacing of 4 Å which is roughly the HN−HN spacing in a protein, and we have chosen to do the calculation on only nine perturbed atoms which is the most that you would typically see in a protein. In addition, we used a relatively small number of randomly generated points to simplify and speed up the calculations and the Phe has been manually docked. These results suggest that (1) ΔCS data can be used to obtain a spatially localized ligand position and since the simulated ΔCS data was calculated with the quantum mechanics based Haigh−Mallion method (2) the point-dipole approximation is accurate for this purpose. The sensitivity of the *j*-surface was investigated by reducing the initial nine ΔCS values to as few as three; using only four positive or five negative ΔCS values; we even distorted the ΔCS data by reducing it by as much as 40%. In all cases accurate, high-quality ligand *j*-surfaces could be constructed that were correctly centered on or near the aromatic ring. We next calculated *j*-surfaces for several protein−ligand complexes to demonstrate their usefulness. All calculations were performed on ligand-free X-ray crystal structures.

Ligand Binding to HCV NS3 Protease. In Figure 4A we show the chemical shift mapping of residues that are perturbed upon ligand binding to (hepatitis C virus) HCV NS3 protease.[19,20] Perturbed residues (as determined by a ^{1}H−^{15}N HSQC spectrum) are given a unique color that is mapped to the protein van der Waals surface using GRASP.[17] The resulting map is

(18) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436−9444; Xu, P. X.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321−333.

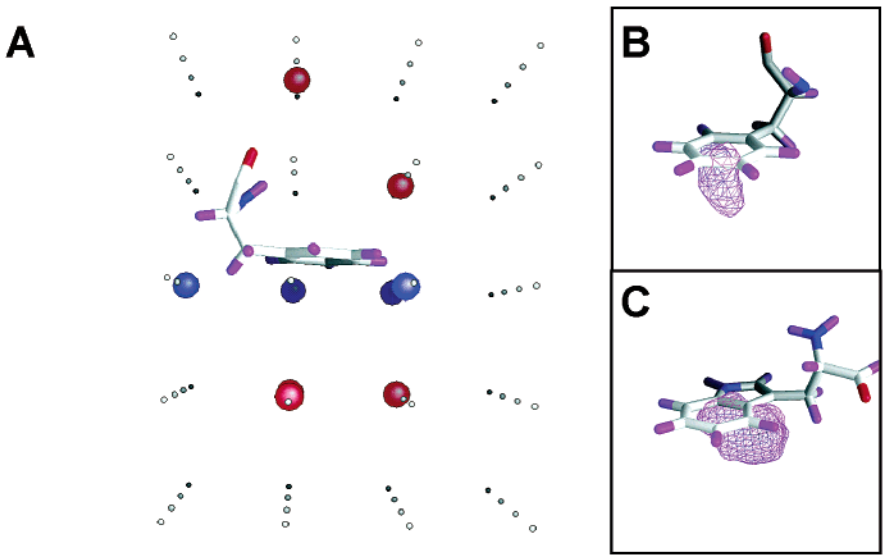

Figure 3. (A) Phenylalanine placed inside a grid of 76 glycines spaced at 4 Å is used to simulate ligand binding. Chemical shift perturbations for glycine HN amide protons were calculated (SHIFTS) with and without a Phe. Glycine HN atoms with negative or positive perturbations are shown by red or blue spheres, respectively; gray spheres show glycine HN atoms (with reduced radii) that were not strongly perturbed by the Phe and were not used in the calculation. (B) 3σ *j*-surface (magenta) for the Phe in the Gly grid is created using GRASP.[17] (C) 3σ *j*-surface for a Trp residue in the Gly grid is shown. In both (B) and (C), *j*-surface is localized and runs through the aromatic ring.

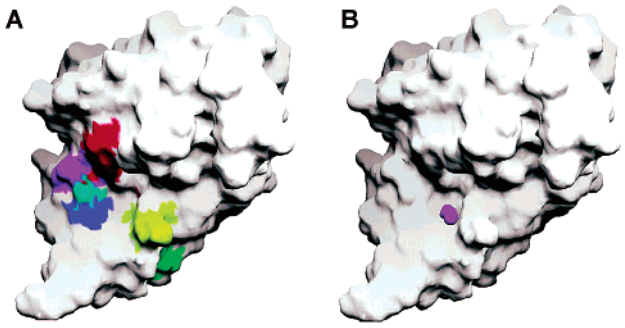

Figure 4. (A) ΔCS map is used to determine the interaction site of an aromatic ligand binding to HCV NS3 protease.[19,20] This map was created by coloring protein residues for which chemical shift perturbations were observed in $^{1}H-^{15}N$ HSQC[21] spectra; each color represents a different residue. The surface area of the perturbed residues covers ~350 Å^2. (B) Ligand *j*-surface (magenta), created from the same data used in Figure 4A localizes the ligand binding site to the S_1-pocket over a volume of 19 Å^3.

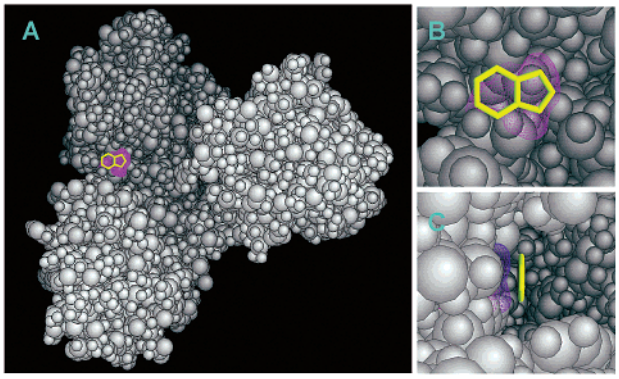

Figure 5. Crystal structure of HCV NS3 helicase[22] (gray atoms) complexed with an aromatic ligand (yellow bonds) is shown together with the ligand *j*-surface (magenta surface). The surface is drawn around dot density at 2σ (magenta) relative to the mean of the dot density distribution.

diffuse; it suggests that the ligand binds the unprimed side of the substrate channel, but no particular site is suggested by these data. Figure 4B shows the *j*-surface using the same chemical shift data that is mapped to the van der Waals surface in Figure 4A. Here the ligand *j*-surface (magenta) is localized off the van der Waals surface of the protein and is centered completely within the S1 pocket. Figure 4 illustrates two important points. First, the van der Waals surface of the protein can provide a significant restraint on the allowed values of the *j*-surface. If the ligand *j*-surface is not located off the protein van der Waals surface, then there may be inconsistencies in the data, or perhaps the amount and quality of the data is insufficient to clearly define a surface. *j*-Surfaces can be localized within the van der Waals surface of the protein, which can be an indication of the rearrangement of protein side chains. Second, this method allows us to evaluate the consistency of the chemical shift perturbation data. That is, we can determine which of the residues that are mapped in Figure 4A contribute to the surface in 4B.

Ligand Binding to HCV NS3 Helicase. In Figure 5 we demonstrate the calculation of a *j*-surface for a ligand binding to the 49 kDa HCV NS3 helicase.[22,23] This example is particularly useful since the X-ray crystal structure[22] (gray atoms) with the ligand (yellow bonds) is known and assignments for many residues are available.[24] Unlike the model systems, the data is not perfect. The aromatic ligand is substituted at several positions; there are only a few ΔCS values to work with, and the shifts and assignments were observed on an engineered subdomain 1 construct rather than the whole enzyme. Nonetheless, the *j*-surface is above the van der Waals surface of the protein, and at the 2σ level, comes into close contact with the ligand binding orientation determined from the crystal structure.

(19) Kim J. L.; Morgenstern, K. A.; Lin, C.; Fox, T.; Dwyer, M. D.; Landro, J. A.; Chambers, S. P.; Markland, W.; Lepre, C. A.; O'Malley, E. T.; Harbeson, S. L.; Rice, C. M.; Murcko, M. A.; Caron, P. R.; Thomson, J. A. *Cell* **1996**, *87*, 343−355.
(20) McCoy, M. A.; Senior, M. M.; Gesell, J. J.; Ramanathan, L.; Wyss, D. F. *J. Mol. Biol.* **2001**, *305*, 1099−1110.
(21) Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185−189.
(22) Yao, N.; Hesson, T.; Cable, M.; Hong, Z.; Kwong, A. D.; Le, H. V.; Weber, P. C. *Nat. Struct. Biol.* **1997**, *6*, 463−467.
(23) Gesell, J. J.; Liu, D.; Madison, V. S.; Hesson, T.; Wang, Y. S.; Weber, P. C.; Wyss, D. F. *Protein Eng.* **2001**, *148*, 573−582.
(24) Liu, D.; Wyss, D. F. *J Biomol. NMR* **2001**, *19*, 283−284.

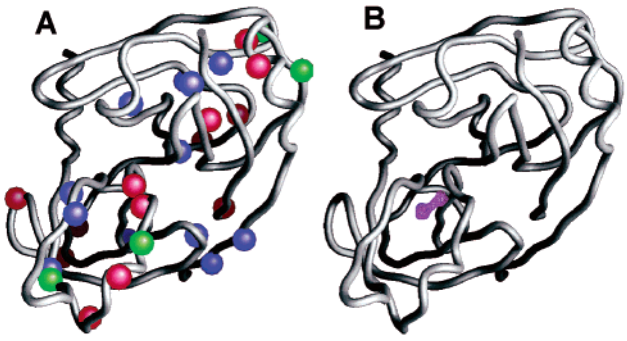

Figure 6. (A) Amide protons of all serines (red), lysines (green), and alanines (blue) of HCV NS3 protease are shown. Only Ser138, Lys136, and Ala157 are perturbed in the real experimental data. If the assignments are not known, then selective ^{15}N labels can be used to identify the residue type that is perturbed. ΔCS values for all residues of that type must be used in the calculation of the ligand *j*-surface. (B) *j*-Surface calculated using 30 ΔCS values (13 Ser, 4 Lys, 13 Ala) that would be available from selective ^{15}N-labeling of Ser, Lys, and Ala in HCV NS3 protease. This *j*-surface correctly identifies the ligand binding site without the use of sequence specific assignments.

While the quality of the data may not be as good as the simulated data, the van der Waals surface of the protein provides a significant restraint on the allowed values for the ligand *j*-surface.

Proteins Lacking Sequence-Specific Assignments. SAR-by-NMR and structured-based NMR screening approaches require sequence-specific backbone assignments of at least the active-site residues which becomes increasingly more difficult with increasing size of the drug target. We tested if we could potentially use the *j*-surface method to localize the ligand binding site without the need for sequence-specific resonance assignments if we knew the 3D structure of the drug target and the ΔCS from amino acid-type selective labeled protein samples. We used the experimental data from the previous example of the HCV protease (Figure 4A) where we observed three perturbations, one each for a Ser, Lys, and Ala residue upon binding of a ring-containing ligand. We then assumed that we did not have sequence-specific resonance assignments but would know those three perturbations from three amino acid-type selective labeled NMR samples where only one type of amino acid is ^{15}N-labeled (Ser, Lys, or Ala). Finally, we allowed all 13 Ser, 4 Lys, and 13 Ala residues in the HCV protease to have the experimentally determined ΔCS for Ser138, Lys136, and Ala157, respectively, in the calculation of the *j*-surface (Figure 6A). Significantly, the *j*-surface for this data set correctly locates the ligand binding site (Figure 6B) without the need of sequence specific assignments. While in this example we assume to know the ΔCS from amino acid-type selective ^{15}N-labeled NMR samples, it demonstrates the potential use of *j*-surfaces for large proteins with amino acid-type selective samples.

Advantages of *j*-Surface Calculations. The use of a *j*-surface to locate a ligand binding is more accurate than a ΔCS map (see Figure 4), but potentially less accurate than energy minimization of a target function derived from chemical shift perturbation data.[12] The target function approach can be used accurately if (1) most of the ΔCS data is from ligand aromatic rings, (2) the data is consistent, (3) there is a single binding site, and (4) there is no rearrangement of the protein surface. Even with ideal data, the target function calculation is time-consuming to set up and run; it is not appropriate for rapid evaluation of many different ligands. In addition, there is no a priori knowledge of the data quality, consistency, or origin of the observed chemical shift perturbation data. Ligand *j*-surfaces provide a way to check the data quality and analyze the origin of the perturbations and are, therefore, complementary with and should precede target function minimization. In addition, ligand *j*-surfaces have a number of significant technical and practical advantages that make them useful in analyzing the chemical shift perturbation data for many compounds prior to the use of a target function approach. (1) Ligand *j*-surfaces can spatially localize the ligand position relative to the protein surface. (2) Small chemical shift perturbations that arise from conformational changes of protein aromatic rings can be distinguished from those that arise from ligand aromatic rings. In the case of ΔCS data that is consistent with ligand binding with no structural rearrangement of the protein, the *j*-surface will be localized off the van der Waals surface of the protein. If the data are consistent with rearrangements of protein residues, the *j*-surface can be located within the van der Waals surface of the protein, centered near an aromatic ring. In the case of widespread protein conformational change, it may not be possible to distinguish the ligand binding from protein rearrangements. (3) Ligand *j*-surfaces can be calculated with data from unassigned proteins. (4) Multisite binding can be detected. (5) No probe or model building is necessary so that binding locations can be determined rapidly. Typical calculations take about 1 min on a DEC alpha. (6) Data consistency can be checked. The probabilistic approach gives us a way to judge the reliability of the *j*-surface. (7) Analyzing the density of overlapping dots can filter out nonaromatic ΔCS contributions that tend to be local. Only the placement of the aromatic ring will be evident from *j*-surfaces at high σ values. (8) Ligand *j*-surfaces that arise from aromatic rings can be sensitive to nonaromatic ligand substitutions when viewed at lower σ. This information is invaluable because it localizes chemical modifications on the ring structure that can be exploited by medicinal or combinatorial chemistry.

Ligand *j*-surfaces can be calculated from any perturbation data. Whether these surfaces have meaning depends on the quality and quantity of the perturbation data. In our experience, poor data usually will not result in a *j*-surface above $\sim 2\sigma$. When the data are weak, sparse, or inconsistent, the *j*-surface can appear discontinuous and delocalized, implying that multiple sites are consistent with the data set. If, however, the data are consistent, part of the *j*-surface will engulf the ligand at values of $2\sigma-3\sigma$. The *j*-surface for some ligands may lie completely off the protein van der Waals surface. While this is an ideal response, the localization of the *j*-surface depends on many factors and may penetrate deeply into the protein's Van der Waal's surface. In such cases, the van der Waals surface can provide an important restraint for the allowed ligand binding positions. Aside from localizing and visualizing ligand binding sites for well behaved protein−ligand complexes, we have found that one useful aspect of these calculations is to point out situations where the data is not consistent with simple binding.

Ligands typically bind to concave pockets in the protein surface. The shape of the ring current distribution (from a single source) makes the *j*-surface calculation possible. The shape of the ligand binding site on the protein surface, however, strongly influences the quality of the *j*-surface. A deep site will typically have more perturbed atoms that are better distributed which

enhances the accuracy of the *j*-surface. However, even the perturbations from a shallow site must converge off the protein surface if the data is consistent. If the data is not plentiful or consistent the calculation of a ligand *j*-surface becomes problematic.

Conclusions

NMR is currently the only screening method where ligand binding location and binding constants (K_d) are present in the same data. The NMR detection of ligand binding to a protein is manifested by small frequency shifts in protein nuclei that are close to the ligand. These frequency shifts are primarily due to magnetic fields that are induced by electronic currents in the ligand. A conversion of NMR frequency that shifts directly into spatial restraints would permit the spatial location of the ligand bearing some resemblance to difference maps of electron density used in X-ray crystallography. This article has introduced the calculation of ligand *j*-surfaces, which carry out such a transformation using a point-dipole approximation parametrized for aromatic rings. Although the method requires that the ligand possesses an aromatic ring, it is widely applicable since >95% of compounds in the MDDR[26] contain at least one five- or six-membered aromatic ring. Ligand *j*-surfaces can be calculated in seconds after the data is acquired with the only requirements of having a previous structure (X-ray, NMR, model) and assigned HSQC spectra of the bound and free protein (with observable ΔCS). These requirements are much less demanding than for ligand-bound protein structures determined by X-ray crystallography or NOE-based NMR. A single *j*-surface can spatially localize the ligand aromatic rings relative to the protein surface. This information is immediately useful for the localization of the ligand interaction site and assessing the quality of the chemical shift perturbation data. Ligand *j*-surfaces become a more powerful tool for structure-based drug design when surfaces of closely related analogues can be calculated and used together with information from computational methods and X-ray crystallography.

JA026166C

(25) Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W. *J. Am. Chem. Soc.* **2000**, *122*, 1241−1242.
(26) MACSS-II Drug Data Report, MDL Information Systems Inc., San Leandro CA 94577.