Journal of Biomolecular NMR, **18:** 189–198, 2000.
KLUWER/ESCOM

Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations

Mark A. McCoy* & Daniel F. Wyss
Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033, U.S.A.

Received 7 June 2000; Accepted 10 August 2000

Key words: calmodulin, chemical shift, intermolecular alignment

Abstract

Structural studies of protein-ligand complexes are often limited by low solubility, poor affinity, and interfacial motion and, in NMR structures, by the lack of intermolecular NOEs. In the absence of other structural restraints, we use a procedure that compares simulated chemical shift perturbations to observed perturbations to better define the binding orientation of ligands with respect to protein surfaces.

Introduction

X-ray and NMR structures of proteins interacting with other molecules provide us with a powerful tool for understanding molecular recognition. The study of complexes is often, however, plagued by poor solubility and data quality is diminished by interfacial motion. Most X-ray crystal structures of protein-ligand complexes are obtained for tight binding ligands with affinities typically ranging from nanomolar to picomolar. While NMR structures can be obtained on complexes with affinities ranging from picomolar to millimolar, the difficulty of structure determination increases with decreasing affinity. There are many conditions under which no NOEs are detected between the ligand and protein; poor solubility, low affinity, conformational variation and lack of protons are the most common difficulties that we experience. Even in unfavorable conditions where other inter-nuclear interactions cannot be detected, the nuclear chemical shift is sensitive to local electronic changes. Chemical shift (CS) perturbations have long been used to indicate interaction sites, but their quantitative use has been slowed by a lack of theoretical understanding. Progress in simulating protein chemical shifts (Case, 1998, 1995; Oldfield, 1995; Ösapay and Case, 1994; Le and Oldfield, 1994; de Dios et al., 1993; Laws et al., 1993) and correlating shifts with secondary and tertiary structure has renewed interest in using shift values as structural restraints (Wishart and Sykes, 1994; Williamson et al., 1995; Wishart et al., 1997).

Experimentally, amide proton and nitrogen CS perturbations are easily detected in ^{15}N-edited HSQC spectra of ^{15}N-labeled proteins interacting with unlabeled ligands. Chemical shift differences between free and bound protein are quantified by obtaining a weighted average of ^{1}HN and ^{15}N CS perturbations. The cause of the perturbations may be from direct contact with the ligand (e.g. hydrogen bonds) or indirect effects such as ring current shifts produced by aromatic groups on the ligand.

Experimental chemical shift perturbations can be used to identify a protein-ligand interaction site using a surface mapping approach (recent examples: Shuker et al., 1996; Farmer et al., 1996; Rajagopal et al., 1997; Schmiedeskamp et al., 1997; Freund et al., 1999). In surface mapping, perturbations are typically mapped from the protein backbone (where the perturbation is detected) through the side chain of the perturbed residue. Residues that are perturbed are assumed to be proximal to the interaction site. One problem that we have encountered with surface mapping lies in the fact that experimental CS perturbations are detected in the protein backbone while much of the protein surface is composed of amino acid side chains. It is

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

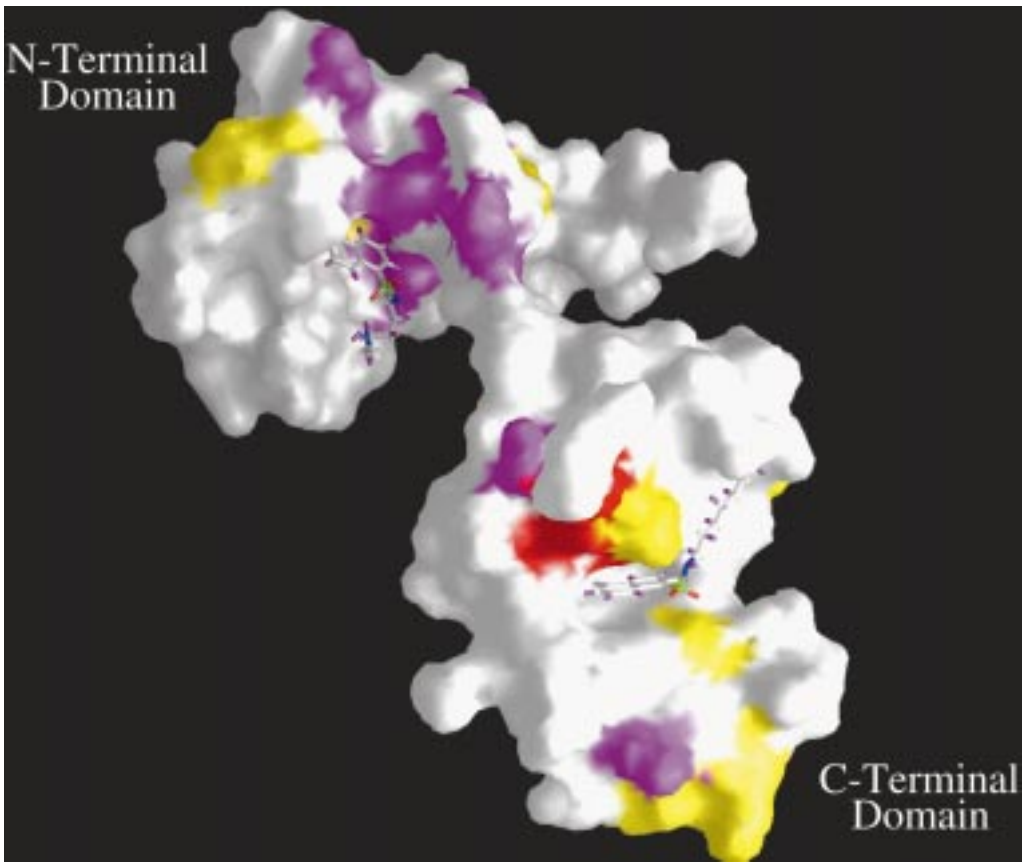

Figure 1. Experimental CS perturbations due to W-7 binding are mapped to the surface of Ca^{2+}-CaM using GRASP (Nichols et al., 1993). Color coding of the Ca^{2+}-CaM surface depicts the weighted average of ^{1}HN and ^{15}N CS perturbations, $\Delta CS = ([\Delta CS(^{1}HN)]^2 + [0.2\Delta CS(^{15}N)]^2)^{1/2}$; red ($\Delta CS = 1.7$ to 0.9 ppm); magenta ($\Delta CS = 0.9$ to 0.5 ppm); yellow ($\Delta CS = 0.5$ to 0.25 ppm). W-7, shown in a bond representation, binds Ca^{2+}-CaM at two sites with approximately equal affinity ($IC_{50} \sim 31\ \mu M$). The correct ligand interaction sites cannot be deduced from this representation.

our experience that CS surface mapping tends to over-emphasize shifts of large, solvent exposed residues (Tyr, Lys, Arg, His) and under-emphasize interactions with smaller residues such as Gly; buried residues are also under-emphasized. Moreover, we have noticed that if only the backbone of a specific residue interacts with the ligand, then surface mapping can lead to the propagation of the interaction site to surface locations that are not involved in ligand binding. Despite these potential problems, chemical shift surface mapping is very useful since it can quickly and easily provide approximate interaction sites on protein surfaces. More detailed structural information would require the detection of intermolecular NOEs to be used as distance restraints between the ligand and the protein.

Figure 1 shows the limitations of the surface mapping approach when it is applied to the Ca^{2+}-CaM-W-7 complex (Osawa et al., 1998). From such a map, the interaction sites of W-7 cannot be located. Even in less problematic CS perturbation maps it is our experience that precise information on the binding mode of the ligand is not available from such data. Our goal was to develop a more quantitative, structural representation of protein/ligand complex interactions under conditions that are not favorable to traditional structural work. To this end, we have developed a procedure to align weakly interacting molecules (few or no NOEs) to a protein surface using information contained in proton chemical shift perturbations. By alignment we refer to the placement of the ligand relative to the protein in the correct binding location and orientation. Recent work (Medek et al., 2000) uses differences in CS perturbations produced by many high affinity, closely related analogs to identify and dock ligands into a binding pocket. Our procedure utilizes a simulation of proton chemical shift perturbations to

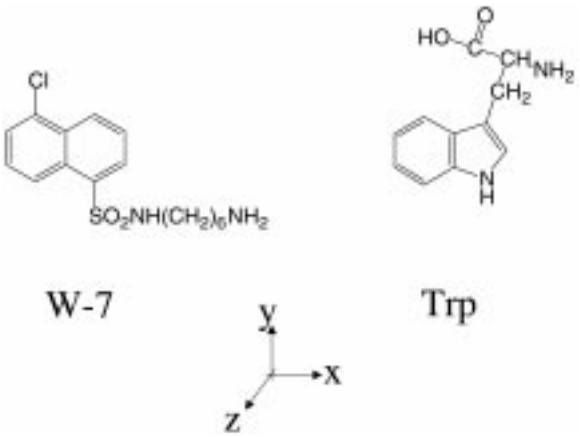

Figure 2. The chemical structure of the W-7 ligand. A tryptophan residue (Trp probe) is used to mimic the W-7 naphthalene group for the purpose of simulating the ring current shifts. Axes are defined for the rotations/translations used in Figure 4.

match experimentally observed CS perturbations; it does not require analogs, nor do we need high affinity ligands. The current implementation of our procedure uses the aromatic ring current from the ligand as the only source of CS perturbations; therefore, the ligand must have at least one aromatic ring. Our procedure is, however, general and is easily modified to use non-aromatic contributions as well. Non-aromatic contributions are in fact calculated, but are usually very small compared with aromatic ring current. This manuscript describes the alignment protocol that we have developed and applies it to align the W-7 ligand (Figure 2) to Ca^{2+}-bound calmodulin (Ca^{2+}-CaM) (Osawa et al., 1998) using ^{1}HA-CS perturbations.

Materials and methods

Alignment procedure

Our alignment procedure consists of three steps; collection of experimental data, calculation of simulated data and iterative alignment of the ligand position with respect to the protein. Below, we describe some of the details of our procedure.

Experimental shift data

In general, our ligand alignment procedure can use experimental ^{1}HN- or ^{1}HA-CS perturbation data (or both). Experimental ^{1}HN-CS perturbations can be obtained from ^{15}N-edited HSQC spectra of ^{15}N-labeled protein with and without ligand. Experimental ^{1}HA-CS perturbations can be obtained from HSQC-TOCSY or HNHA spectra that require only the use of ^{15}N-labeled protein. All shift data for this work were obtained from published chemical shifts of Ca^{2+}-CaM (BMRB accession number-547; supplementary shifts from M. Ikura) and from the Ca^{2+}-CaM-W-7 complex (BMRB accession number-4056). The 1MUX.pdb structure file of the Ca^{2+}-CaM-W-7 complex was used for all CS simulations of ligand binding.

Atom representation

Experimental and simulated CS perturbations are mapped onto an atom representation rather than to a surface. Figure 3A shows experimental ^{1}HA CS perturbations upon W-7 binding to calmodulin in an atom representation. In an atom representation, the atom size is linearly scaled according to relative intensity of the CS perturbations; only those atoms that are perturbed are shown. The atom color indicates the sign of the shift. Negative CS perturbations (bound PPM − reference PPM) are indicated by red atoms while positive CS perturbations are indicated by blue atoms. Note that the atom representation of CS perturbations eliminates problems such as interaction propagation that is often seen in the surface mapping approach.

Calculated shift data

Simulated proton CS perturbations are obtained from comparing changes in calculated shifts of the protein with and without ligand. SHIFTS 4.01b (D. Case, The Scripps Research Institute, La Jolla, CA) is used to calculate, from a pdb file, the expected ^{1}H shifts of a protein with a ligand in close proximity and the ligand far removed from the protein (reference calculation).

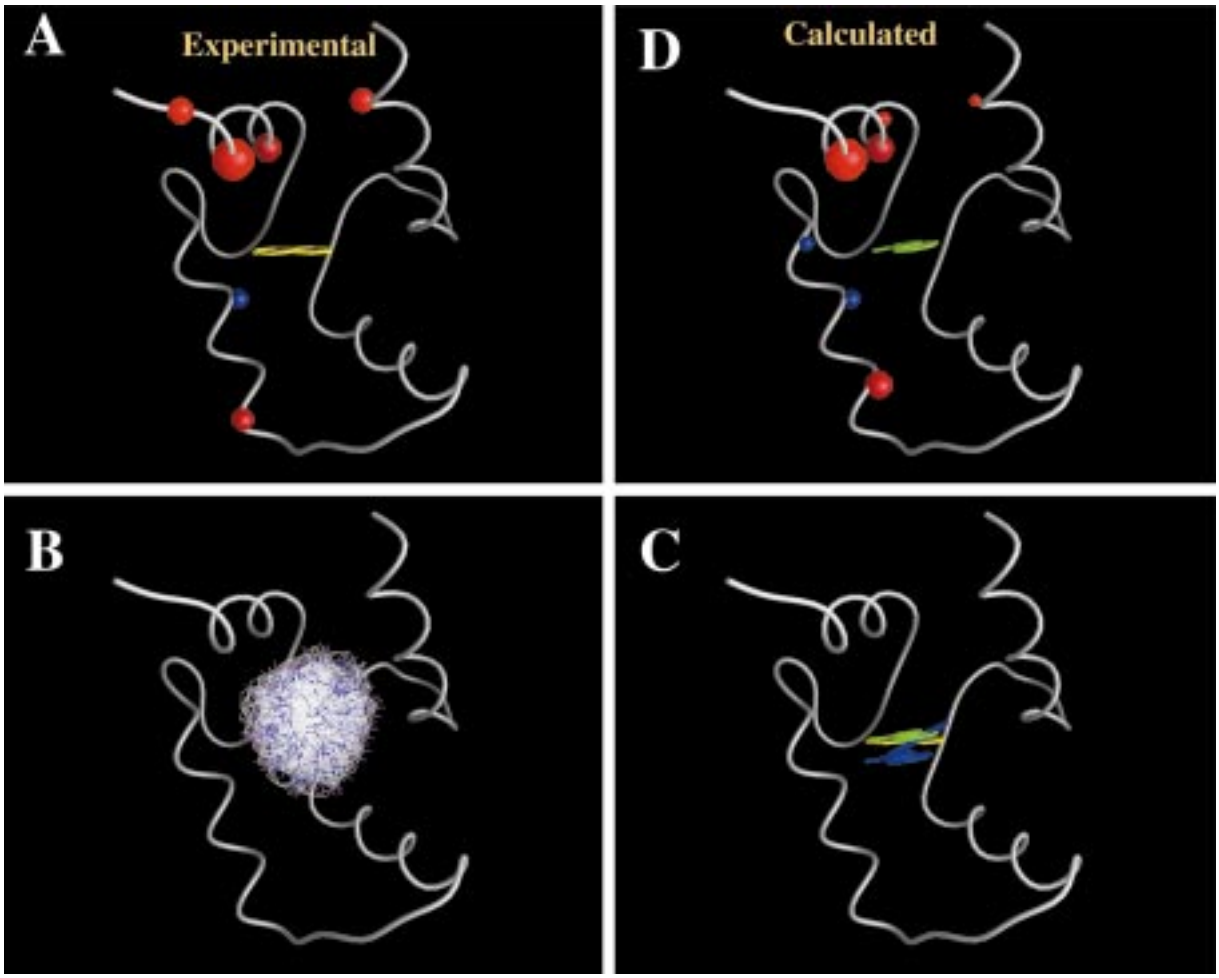

Figure 3. Chemical shift perturbations for HA atoms are plotted in an atom representation using GRASP (Nichols et al., 1993). The size of the spheres is scaled relative to the largest CS perturbation while the sign of the perturbation determines the sphere color. Only the ring structures of W-7 and the Trp probe are shown for clarity. (A) Experimental ^{1}HA chemical shift perturbations caused by W-7 binding to the C-terminal domain of Ca^{2+}-CaM; W-7 is shown in yellow. (B) ^{1}HA-CS perturbations were calculated for 2160 Trp orientations and compared to observed CS perturbations in Figure 3A using a Q-score (see text). The collection of Trp probes is placed in the center of the CS perturbations which coincides with the W-7 binding site. No Trp probe was more than ±5 Å from the W-7 binding site since this produces Q-scores that are too large. (C) Superposition of the W-7 orientation (yellow), the Trp orientation with the best Q-score (green) and the next two best scores (blue). (D) Calculated CS perturbations from a 'best fit' Trp probe interacting with Ca^{2+}-CaM. A comparison of (D) with (A) reveals that the CS calculations place the Trp probe in the correct binding site and in the correct orientation. The agreement between the experimental and calculated data (sphere size and color) is very good.

The difference between the two calculations simulates the CS perturbations. Only proton chemical shifts are used in our alignment procedure. Even though experimental ^{15}N CS perturbations can be large, the understanding of ^{15}N shifts is limited (Le and Oldfield, 1994) and there is no appropriate simulation currently available. For weakly interacting ligands, we assume that the aromatic ring current from the ligand is the most significant source of chemical shift change. While SHIFTS can calculate other contributions to the chemical shift (electrostatic, bond anisotropy, etc.) we currently use only the ring current contribution. For perturbations caused by aromatic ring currents, the sign of the chemical shift changes as a function of ring orientation and the intensity of the shift changes as a function of distance from the ring. A single ring on the surface of a protein produces significant perturbations (∼0.1 ppm) for protons 7–10 Å from the ring. Out-of-plane orientations with respect to the aromatic ring give rise to negative upfield shifts and in-plane orien-

tations give positive downfield shifts. The change of the sign of the aromatic ring current as a function of ring orientation typically produces a pattern of positive and negative proton shifts when aromatic ligands interact with a protein surface. These patterns are very sensitive to translation and rotation of the ring.

It is important to point out that the accuracy of the 'protein only' CS calculations is less important in our applications than, for example, in using CS calculations to predict chemical shifts of folded proteins. In our work we calculate differences in CS perturbations; only the shift differences caused by the interaction of ligand with the protein are important. As long as these interactions are dominated by ring current shifts produced by the ligand, we feel that this procedure is accurate.

Ligand alignment

The differences between experimental and simulated CS perturbations are minimized to produce the ligand alignment. GRASP (Nichols et al., 1993) is used to place ligands in proximity to the protein and vary distances and angles of the ligand to best fit the NMR data. The calculated and observed data can be directly imported into GRASP for comparison. Since SHIFTS is calibrated primarily for proteins, amino acid fragments were used as probes of chemical shift perturbation. In most simulations, only aromatic ring currents are used for alignment. Phenylalanine is used as a probe for fragments containing benzene rings. Multiple Phe probes are used for compounds containing multiple aromatic rings. More refined models, for example of ^{1}HN proton donation to a carbonyl, can be created using glycine as a probe.

Results

Ca^{2+}-CaM-W-7 complex

The Ca^{2+}-CaM-W-7 complex is used as a model system to test our alignment protocol. This system is used because there are assignments available for Ca^{2+}-CaM with and without the W-7 ligand, and W-7 interacts weakly yet it is soluble enough to permit a high resolution solution structure to be determined with W-7 bound (Osawa et al., 1998). Since there is no secondary or tertiary structural rearrangement of Ca^{2+}-CaM upon W-7 binding, we can use either ^{1}HN or ^{1}HA data for our ligand alignment procedure. While the pattern of experimental ^{1}HN shifts is in agreement with the pattern of ^{1}HA shifts, there are many non-specific ^{1}HN-CS perturbations outside the W-7 binding site that cannot be modeled with a single probe. In our experience non-specific CS perturbations can result from high ligand-to-protein ratios. The experimental ^{1}HN data for the Ca^{2+}-CaM-W-7 complex was obtained with a 5:1 ligand-to-protein ratio which was necessary to obtain good NOE data. ^{1}HA shifts are, however, less affected and can be reliably fit even at high ligand-to-protein ratios. Due to the aforementioned limitations in the experimental data, we used only ^{1}HA shifts (calculated and experimental) for ligand alignment to the Ca^{2+}-CaM-W-7 model system.

HA shifts pinpoint W-7 binding sites

The Ca^{2+}-CaM-W-7 structure was determined on a sample of 1.5 mM Ca^{2+}-CaM and 7.5 mM W-7 (Osawa et al., 1998). No structural rearrangement of Ca^{2+}-CaM was observed upon W-7 binding, although lower affinity non-specific interactions were detected. The W-7 ligand binds calmodulin at two similar sites with an IC_{50} of 31 μM for Ca^{2+}-CaM dependent activation of phosphodiesterase (Tanaka et al., 1982; MacNeil et al., 1988). Side chains of Phe92, Ile100, Leu105, Met109, Met124, Ile125, Ala128, Val136, Phe141, Met144 and Met145 define the C-terminal binding site. The side chains of Phe19, Ile27, Leu32, Met36, Met51, Ile52, Val55, Ile63, Phe68, Met71 and Met72 define the N-terminal binding site.

The experimental ^{1}HA-CS perturbations due to W-7 binding to the C-terminal pocket are shown in an atom representation in Figure 3A. The N-terminal pocket shows a different pattern of shifts since W-7 binds in a different conformation. No calculations were performed on this domain. The largest perturbations in the C-terminal domain were observed for HA-Met145 (-0.33 ppm) and HA-Phe141 (-0.24 ppm); the radii of the remaining atoms are linearly scaled. Only perturbations greater than ± 0.14 ppm are shown. To simulate the experimental CS perturbations from the naphthalene group of W-7, we replaced it with a Trp probe (Figure 2) and performed ^{1}HA-CS calculations using SHIFTS as described above. The quality of the fit of the calculated CS perturbations to the experimental values can be quantified by

$$Q = \sum_{i=1} \left[\frac{\Delta CS_i(obs)}{|\Delta CS_{max}(obs)|} - \frac{\Delta CS_i(calc)}{|\Delta CS_{max}(calc)|} \right]^2$$

The Q-score was calculated for a collection of 2160 Trp probe orientations using experimental ^{1}HA-CS perturbations (ΔCS(obs)) and calculated ^{1}HA-CS perturbations (ΔCS(calc)). A portion of this collection is shown in Figure 3B. The Trp probe orientations (60 rotations $\times$ 36 translations) were produced in GRASP (Nichols et al., 1993) and written to a single pdb-file. Each of the 2160 Trp probes was extracted and concatenated with the protein pdb-file to produce 2160 protein+probe pdb-files. Shell and awk scripts were written to calculate differences in HA shifts between the protein+probe pdb-files and a single reference pdb-file where the probe has been moved 500 Å so that it would not produce any ^{1}HA-CS perturbations for the protein. Awk scripts are used to scale and format a GRASP history file that is used to display both the calculated and experimental data in Figures 3D and 3A, respectively. The size of the spheres is scaled relative to the largest CS perturbation for the HA atoms, while the sign of the perturbation determines the sphere color. Finally, a shell script is used to calculate the Q-score which is used to quantify the difference between the experimental and calculated data (for example, Figures 3A and 3D) for all Trp probe orientations.

In the expression for Q, normalized ΔCS values were used to emphasize the fit of a pattern of calculated shifts to the pattern of observed shifts and to reduce possible errors from, for example, having less than 100% of the ligand bound. When summed over all residues, Q-scores ranged from 2.1 (best fit) to 19.0, with a mean and average score of 7.7. Most Trp orientations matched the experimental data poorly; only 48 conformations ($\sim$2%) produced scores less than 3.0. The lowest scoring conformations are shown in Figure 3C relative to the W-7 conformation; these best Q-scores are $\sim$10–15% lower than the scores of the 'next best' conformations.

The pattern of calculated ^{1}HA-CS perturbations from the 'best fit' Trp probe, Figure 3D, agrees well with the experimentally determined shift values, Figure 3A. The largest perturbations in the simulated data are for HA-Met145 (-0.24 ppm) and HA-Phe141 (-0.15 ppm); the radii of the remaining atoms are linearly scaled. Only perturbations greater than ± 0.07 ppm are shown. The final position of the Trp probe was about 0.5 Å from the position of the naphthalene group of W-7 that was determined in the solution structure from 37 intermolecular NOE restraints.

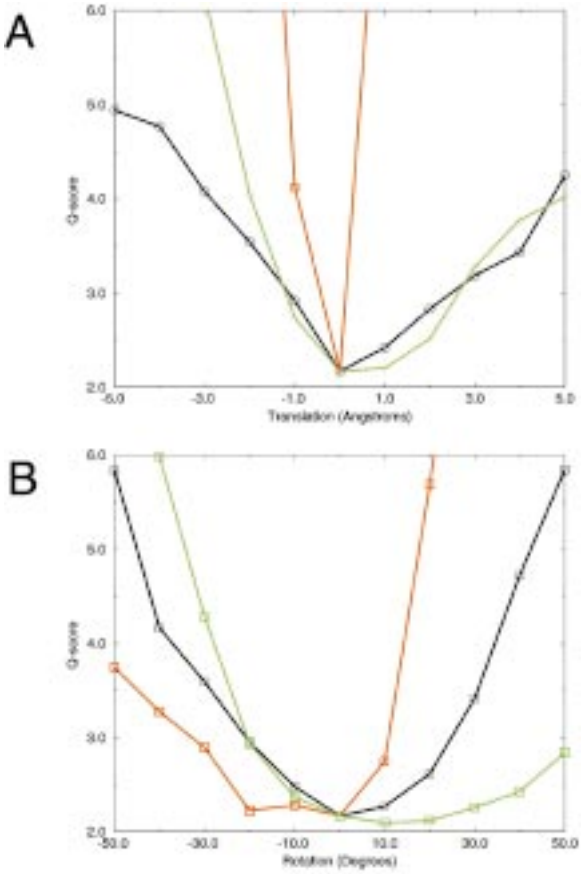

Figure 4. The sensitivity of the fit of calculated to experimental CS perturbations (Q-score) is evaluated as a function of distance and rotation angle for the 'best fit' Trp probe conformation. Scores > 3 are very bad fits to the experimental data; only 2% conformations in Figure 3B had Q scores 2–3. (A) The Trp probe is translated ± 5 Å in 1 Å increments along the x (black), y (red) and z (green) axes and Q is calculated for each step; the protein position is kept constant. Q is sensitive to x- and z-translations of $\sim\pm 1$ Å, and y-translations of $\sim\pm 0.25$ Å. (B) The Q-score is calculated at 10° increments for Trp probe rotations about the x (black), y (red) and z (green) axes. The Q-score is sensitive to 15–20° rotations about x,y and, as expected, is not very sensitive to rotations that are perpendicular to the Trp ring (z-rotations). As a consequence of the sensitivity of Q to xyz-translations and xy-rotations and its insensitivity to z-rotations, most low scoring Trp conformations have the same location and differ mostly by z-rotations.

The experimental pattern of positive and negative CS perturbations and the extent of perturbations limit the number of probe orientations that can fit the observed data. The sensitivity of the Trp probe to translation and rotation is demonstrated in Figure 4. In Figure 4A, the Q-score is plotted for x,y,z translations of the Trp probe in 1 Å increments. The starting point (x,y,z $= 0$) is the lowest scoring Trp probe orientation from Figures 3C and 3D. Figure 4B shows the depen-

dence of Q-score on rotations of the lowest scoring Trp probe orientation. Figure 3 shows that CS calculations using a Trp probe can simulate W-7 binding. Figures 3 and 4 show that a small collection of calculated conformations fit the experimental data. Figure 4 shows Q-scores can easily distinguish Trp probe conformations that differ from the 'best' conformations by translations of more than 1–1.5 Å or by x,y-rotations of greater than ∼15°, but are relatively insensitive to z-rotations. The Q-score is very sensitive to x- and y-rotations due to the sign change of the chemical shift as a function of these rotations. Small rotation angles are used in Figure 4B to emphasize the sensitivity of the Q-score to rotations. Also, the ring current is symmetric about the x-axis so that the resulting Q-score at 50° is the same as at −130°. The ring current is not symmetric with respect to y- or z-rotations so that ±180° rotations will produce local minima. Note that the naphthalene ring of W-7 is symmetric about all 3 axes and, consequently, 180° rotations will result in conformations that are indistinguishable based solely on ring current calculations. Some of these ambiguities will be resolved when the probe is replaced by the ligand. For example, when the naphthalene ring of W-7 is aligned with the Trp probe, conformations of the naphthalene ring that would place the aminohexylsulfonamide into the protein surface can be excluded.

Methionine HE shifts detect structural rearrangement

If extensive side chain assignments are available, chemical shift perturbations can be used to identify subtle conformational changes that accompany ligand binding. A comparison of an X-ray structure of Ca^{2+}-CaM to the solution NMR structure of the Ca^{2+}-CaM-W-7 complex suggests that ligand binding to Ca^{2+}-CaM is accomplished, in part, by conformational changes in the methionine side chains (Osawa et al., 1998). To confirm this model, we have performed calculations of ^{1}HE Met-CS perturbations with and without W-7. We assume that, if the methionine side chains rearrange, then we will not be able to reproduce the observed pattern of CS perturbations, since only the ligand position moves in our simulation. Figure 5 summarizes the comparison between the observed and simulated ^{1}HE shifts. Unexpectedly, we found that ^{1}HE-Met CS perturbations are consistent with little, if any, side chain rearrangement upon W-7 binding to the C-terminal ligand binding site; the pattern of CS perturbations can be explained completely by ring current effects from W-7. Considering the possible inaccuracies due to using a Trp probe or from insufficient NOE restraints in the Met side chains, the simulations are in excellent agreement with the experimental data. The experimental CS data for W-7 binding to the N-terminal site do not fit well. This result is consistent with the suggestion that ligand binding is accompanied by conformational change in the side chains of the N-terminal methionine residues. This result would be in agreement with the work of Osawa and co-workers who explain the diverse ligand recognition of CaM to be a result of side chain rearrangements (Osawa et al., 1998).

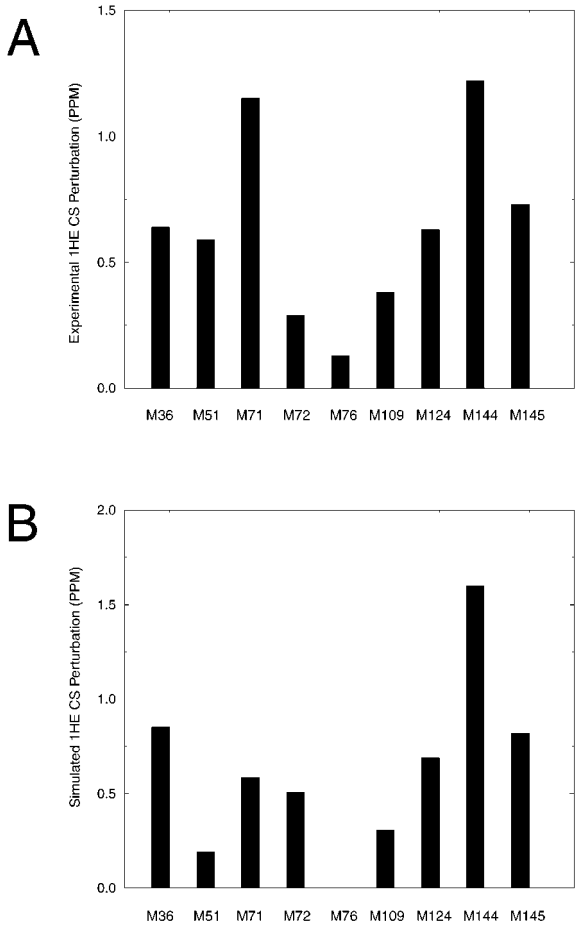

Figure 5. (A) Experimental ^{1}HE chemical shift perturbations caused by W-7 binding to Ca^{2+}-CaM are plotted as a function of Ca^{2+}-CaM residue number for nine methionine residues. (B) Calculated ^{1}HE chemical shift perturbations from the Trp probe interacting with Ca^{2+}-CaM are plotted as a function of residue number. The correspondence between the experimental and simulated CS data suggests that no change in methionine side chain conformation is present in the W-7 binding to the C-terminal binding site.

Discussion

Our alignment protocol mimics experimental proton CS perturbations using a simulation of the ligand binding to the protein. The conformation of the simulated ligand that most closely reproduces the experimental values is used as a model for the protein-ligand complex. From our experience, fitting patterns of CS changes is important to obtaining a good alignment. Matching changes in the sign of shifts for nearby residues (when possible) provides a sensitive restraint for ring orientation. Once an alignment is made using a probe residue, a model of the ligand is made and the aromatic group(s) of the ligand can be superimposed upon the probe.

We have used the Ca^{2+}-CaM-W-7 pdb-file to demonstrate the ligand alignment protocol. In this structure file, calmodulin amino acid side chains accommodate the W-7 ligand. If the results of the CS simulations place the Trp probe in the same location and orientation as the W-7 ligand (which they have) then there will be no van der Waals violations. In terms of building models of ligand binding to protein surfaces, it will likely be easier to fit ligands using a ligand bound pdb-file.

Note, however, that in our ligand alignment protocol the probe position is determined only by the fit of ^{1}HA CS perturbation data (calculated vs. experimental). Side chain coordinates cannot influence the fit of the probe. Once the probe is replaced by the ligand it can be energy minimized relative to the protein, using the position determined from the alignment as a restraint. No energy minimization was used in the Ca^{2+}-CaM-W-7 alignment; only CS data was used to determine the probe/ligand position. In our experience (with ligand-bound X-ray and NMR structures) our ligand alignment protocol does not produce structures with van der Waals violations. We anticipate that the ± 1 Å, $\pm 20°$ accuracy of our method would allow for ligand placement that would simultaneously satisfy CS perturbation data, NOE restraints and van der Waals forces. If the probe orientation that best matches the experimental data results in large vdW violations then one would suspect that the data is not sufficient to define the probe location or that the assumption that no protein conformational change is induced by ligand binding is a bad assumption.

In fitting our data, we emphasize the importance of fitting the observed pattern of CS perturbations; we do not try to match the intensities of the calculated and the experimental shifts. This reduces difficulties such as imperfect parameterization in the calculations. If we try to match the intensities, we would need to account for the difference in the occupancies between experimental and simulated CS perturbation data. In the calculated data there is 100% occupancy while experimentally there is a dependence of the observed shift on the occupancy of the binding site in the fast exchange limit. In our simulations with a single Phe or Trp probe, ^{1}H perturbations are typically ± 0.1–0.3 ppm (100% occupancy) which agrees well with our experimental data. Such shifts are expected for aromatic rings that bind on the surface of proteins through hydrophobic interaction with protein side chains. Shifts from hydrogen bond formation can be much larger and are typically, although not necessarily, downfield.

In principle, any proton (or carbon) chemical shift can be used with the protocol outlined above. We use data from backbone atoms to avoid complications that may arise due to side chain conformational changes upon ligand binding. We use backbone proton CS perturbations (^{1}HN, ^{1}HA) because they can be easily acquired and simulated with public domain software. Obviously, CS data from ^{1}HN protons is the most desirable to use, but care must be exercised. For example, the correspondence of calculated with observed ^{1}HN shifts is low. ^{1}HN shifts are more difficult to simulate since they make electrostatic contacts and form hydrogen bonds whereas aliphatic protons do not. While they may be more difficult to obtain, the correspondence between calculated and observed ^{1}HA shifts is good. Furthermore, it has been suggested that the chemical shift dispersion of aliphatic protons in proteins is 'nearly entirely' due to ring current effects (Wüthrich, 1986). We can conclude that the primary source of CS perturbation of aliphatic protons will also be due to ring current effects. Fitting CS perturbation data for aliphatic protons should therefore be more accurate since many of the assumptions that we have made in our protocol are violated less often.

Limitations in the experimental data

Perturbation data may contain too few or too many peaks to be fit. We typically try to fit three or more experimental perturbations in each model. We do not recommend the use of this procedure to fit 1 or 2 data points – there are too many conformations that can fit the data. At the other extreme, peaks all throughout the structure may be perturbed. These may be a result of large scale conformational changes in the protein or due to non-specific interactions. If non-

specific interactions are suspected, a lower ligand-to-protein ratio can be used to collect the experimental CS perturbation data.

Our ligand alignment protocol can tolerate small local conformational changes but is inappropriate if there is significant protein reorganization (Foster et al., 1998). Our protocol was developed for use with small, weakly interacting ligands (mM-μM affinity) that, in our experience, do not cause protein conformational changes. We can try to distinguish specific from non-specific interactions by simultaneously fitting ^{1}HN and ^{1}HA shifts. While ^{1}HN shifts may be more plentiful and easier to detect, ^{1}HA shifts are less affected by hydrogen bond formation and their simulation is more accurate.

Even if we have a sufficient number of points to use in our alignment procedure, the models can still be ambiguous. This is especially true if only ring current calculations are performed; only the ring placement and orientation affect the shift. It is therefore important that, in addition to simulations, we try to correlate patterns of CS changes with structural differences in closely related analogs. This procedure can eliminate ambiguities such as identifying the interaction sites of specific rings in compounds with multiple rings or it can be used to identify specific atoms involved in hydrogen bond formation.

It is important to suspect that, due to imperfect assumptions and incomplete models, it may not be possible to simulate and fit all of the shift data. Some difficulties may arise if the chosen probe does not accurately mimic the ligand. Dynamics will also make model building less accurate. For example, the ligand might interact dynamically with the protein surface; perhaps there is movement of protein side chains in response to the binding of the ligand. Ligands that interact with surface exposed aromatic groups can be difficult to model since it may not be straightforward to distinguish between the movement of surface aromatic groups and the ring current of aromatic group(s) from the ligand. To characterize possible conformational changes in the protein we can simulate the expected shifts from several conformations. We observe that buried protons are mostly affected by conformational changes of aromatic side chains that belong to the protein – strong shifts for these residues (with the correct pattern of signs) would likely be correlated with motion of such side chains.

Other sources of CS perturbation

We typically assume that all perturbations are due to aromatic ring currents since these are the largest generators of CS perturbation and their simulation is straightforward. One must consider that hydrogen bond formation (or breaking) can occur in addition to the ring current shifts. Often, hydrogen bond formation is characterized by downfield ^{1}HN shifts. While downfield shifts can be due to increasing hydrogen bond character, they can also result from ring current or electrostatic effects. If a water, hydrogen bonded to a backbone amide of the protein, is displaced by a ligand that also forms a hydrogen bond to the same amide proton it is unclear how this would affect the chemical shift of the amide proton. We can attempt to distinguish CS perturbations caused by hydrogen bonding from those caused by ring currents by collecting and fitting CS data for other protein nuclei (e.g. buried HNs or HA protons) that are not involved in hydrogen bonding and are not susceptible to chemical exchange. Combinations of rings with hydrogen bond acceptors can lead to further complication in interpreting shift patterns. Improvements in SHIFTS and other programs will allow us to include interactions other than ring current in our calculations. Until the programs are improved, we use conservative models, fit as many points as possible and rely on SAR to verify perturbations that cannot be modeled using aromatic ring currents.

While some or all of the limitations discussed in this section may exist to some extent, we experimentally observe that ligands to calmodulin interact in a sequence specific manner that produces large chemical shift changes. To decode the structural content of these shifts, we fit as many of the perturbed HA protons as possible. Validation of our fitting procedure on the calmodulin-W-7 model system suggests that this procedure can produce an accurate model of ligand alignment.

Conclusions

The work presented here establishes a protocol for the use of proton chemical shifts for alignment of molecules where the interaction interface does not show many intermolecular NOEs. Our applications have been to model the weak ligand binding in the Ca^{2+}-CaM-W-7 complex. The current protocol calculates expected proton shifts and matches these to

observed shifts. Only the contribution of ring current to the chemical shift changes upon ligand binding is used. We also assume that there is little conformational change upon ligand binding. The protocol we describe is most appropriate for the study of weakly interacting complexes since we expect that electrostatic and anisotropic contributions to the chemical shift changes will be more significant in tighter binding ligands and/or for ligands that form many hydrogen bonds. Future extensions of this work will include the use of chemical shifts from ^{1}HN, ^{15}N and ^{13}C nuclei, inclusion of electrostatic and anisotropic contributions to our CS simulations. We anticipate that ligand-induced chemical shift perturbations can be used as restraints in structure calculations and can be energy minimized with NOEs, van der Waals and electrostatics to give more accurate protein/ligand structures. Differential chemical shifts of close analogs (Medek et al., 2000) can also be used together with our alignment protocol to give accurate models of protein/ligand complexes.

Acknowledgements

We would like to thank Prof. Mitsu Ikura and Dr. Masanori Osawa for providing us with the proton chemical shifts of Ca^{2+}-CaM, Dr. Dingjiang Liu for discussions on CaM-W-7 and Prof. David Case for discussions concerning calculations of chemical shifts.

References

Case, D.A. (1998) *Curr. Opin. Struct. Biol.*, **8**, 624–630.
Case, D.A. (1995) *J. Biomol. NMR*, **6**, 341–346.
de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993) *Science*, **260**, 1491–1496.
Farmer II, B.T., Constantine, K.L., Goldfarb, V., Friedrichs, M.S., Wittekind, M., Yanchunas, J. Jr., Robertson, J.G. and Mueller, L. (1996) *Nat. Struct. Biol.*, **3**, 995–997.
Foster, M.P., Wuttke, D.S., Clemens, K.R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J. and Wright, P.E. (1998) *J. Biomol. NMR*, **12**, 51–71.
Freund, C., Dötsch, V., Nishizawa, K., Reinherz, E.L. and Wagner, G. (1999) *Nat. Struct. Biol.*, **6**, 656–660.
Laws, D.D., de Dios, A.C. and Oldfield, E. (1993) *J. Biomol. NMR*, **3**, 607–612.
Le, H. and Oldfield, E. (1994) *J. Biomol. NMR*, **4**, 341–348.
MacNeil, S., Griffin, M., Cooke, A.M., Pettett, N.J., Dawson, R.A., Owen, R. and Blackburn, G.M. (1988) *Biochem. Pharmacol.*, **37**, 1717–1723.
Medek, A., Hajduk, P.J., Mack, J. and Fesik, S.W. (2000) *J. Am. Chem. Soc.*, **122**, 1241–1242.
Nichols, A., Bharadwaj, R. and Honig, B. (1993) GRASP: graphical representation and analysis of surface properties. XXXVII Annual Meeting of the Biophysical Society, Washington, DC, USA, *Biophys. J.*, **64**, A166.
Oldfield, E. (1995) *J. Biomol. NMR*, **5**, 217–225.
Ösapay, K. and Case, D.A. (1994) *J. Biomol. NMR*, **4**, 215–230.
Osawa, M., Swindells, M.B., Tanikawa, J., Tanaka, T., Mase, T., Furuya, T. and Ikura, M. (1998) *J. Mol. Biol.*, **276**, 165–176.
Rajagopal, P., Waygood, E.B., Reizer, J., Saier, M.H. Jr. and Klevit, R.E. (1997) *Protein Sci.*, **12**, 2624–2627.
Schmiedeskamp, M., Rajagopal, P. and Klevit, R.E. (1997) *Protein Sci.*, **9**, 1835–1848.
Shuker, S.B., Hajduk, P.J., Meadows, R.P. and Fesik, S.W. (1996) *Science*, **274**, 1531–1534.
Tanaka, T., Ohmura, T. and Hidaka, H. (1982) *Mol. Pharmacol.*, **22**, 403–407.
Williamson, M.P., Kikuchi, J. and Asakura, T. (1995) *J. Mol. Biol.*, **247**, 541–546.
Wishart, D.S., Watson, M.S., Boyko, R.F. and Sykes, B.D. (1997) *J. Biomol. NMR*, **4**, 329–336.
Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **2**, 171–180.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York, NY, p. 31.