Evaluation of Site-Directed Spin Labeling for Characterizing Protein-Ligand Complexes Using Simulated Restraints

Keith L. Constantine
Structural Biology and Modeling, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543 USA

ABSTRACT Simulation studies have been performed to evaluate the utility of site-directed spin labeling for determining the structures of protein-ligand complexes, given a known protein structure. Two protein-ligand complexes were used as model systems for these studies: a 1.9-Å-resolution x-ray structure of a dihydrofolate reductase mutant complexed with methotrexate, and a 1.5-Å-resolution x-ray structure of the V-Src tyrosine kinase SH2 domain complexed with a five-residue phosphopeptide. Nitroxide spin labels were modeled at five dihydrofolate reductase residue positions and at four SH2 domain residue positions. For both systems, after energy minimization, conformational ensembles of the spin-labeled residues were generated by simulated annealing while holding the remainder of the protein-ligand complex fixed. Effective distances, simulating those that could be obtained from ^{1}H-NMR relaxation measurements, were calculated between ligand protons and the spin labels. These were converted to restraints with several different levels of precision. Restrained simulated annealing calculations were then performed with the aim of reproducing target ligand-binding modes. The effects of incorporating a few supplementary short-range (≤5.0 Å) distance restraints were also examined. For the dihydrofolate reductase-methotrexate complex, the ligand-binding mode was reproduced reasonably well using relatively tight spin-label restraints, but methotrexate was poorly localized using loose spin-label restraints. Short-range and spin-label restraints proved to be complementary. For the SH2 domain-phosphopeptide complex without the short-range restraints, the peptide did not localize to the correct depth in the binding groove; nevertheless, the orientation and internal conformation of the peptide was reproduced moderately well. Use of the spin-label restraints in conjunction with the short-range restraints resulted in relatively well defined structural ensembles. These results indicate that restraints derived from site-directed spin labeling can contribute significantly to defining the orientations and conformations of bound ligands. Accurate ligand localization appears to require either a few supplementary short-range distance restraints, or relatively tight spin-label restraints, with at least one spin label positioned so that some of the restraints draw the ligand into the binding pocket in the latter case.

INTRODUCTION

In recent years, methodological advances have significantly increased the upper size limit of proteins for which detailed structural information can be obtained using NMR spectroscopy (Arrowsmith and Wu, 1998; Farmer and Venters, 1998; Gardner and Kay, 1998; Wider and Wüthrich, 1999). Global folds for monomeric proteins as large as 42 kDa have been determined (Mueller et al., 2000). In general, for proteins larger than ~30 kDa, NMR assignments are obtainable only for a subset of the proton resonances, i.e., the backbone amide protons and the protons of specifically labeled residues (Metzler et al., 1996; Goto et al., 1999). This limits the number of nuclear Overhauser effect (NOE) restraints available for structure determination. To obtain accurately converged structures of larger proteins, it may often be necessary to include additional information, such as dipolar coupling restraints (Mueller et al., 2000) or restraints derived from paramagnetic relaxation rate enhancements (Battiste and Wagner, 2000; Gaponenko et al., 2000).

Data on the locations, orientations, and conformations of bound ligands also become more difficult to obtain with increasing protein size. Isotope filtering and editing methods (Fesik, 1988, 1991; Otting and Wüthrich, 1990; Breeze, 2000) can be applied to acquire intra-ligand and protein-ligand NOE restraints. Due to limited assignments for the ^{1}H resonances of the protein, the number of protein-ligand NOE restraints that can be obtained will generally be quite low for large proteins. Transferred NOE experiments (Clore and Gronenborn, 1982, 1983; Campbell and Sykes, 1993; Ni and Scheraga, 1994) can be applied to determine the internal conformations of bound ligands in large systems, but these experiments are applicable only to ligands of relatively low affinity, and they do not provide any information on the location and orientation of the bound ligand. Thus, there is a clear need for NMR approaches aimed at determining ligand-binding modes for large proteins.

An approach that has been used for many years in biomolecular NMR to obtain structural information is the measurement of nuclear relaxation rate enhancements caused by proximity to a paramagnetic center (McConnell, 1967; Swift, 1974; Berliner, 1976; Berliner, 1979; Holtzman, 1984). The measured rate enhancements afford estimates of proton-electron distances. A number of studies have exploited naturally occurring paramagnetic metal ion binding sites within proteins (Lee and Sykes, 1980; Frederick et al., 1988; Ray and Rao, 1988), whereas others have utilized paramagnetic ligands (Anglister et al., 1984; de Jong et al., 1989; Johnson et al., 1999; Mahoney et al., 2000).

Received for publication 2 October 2000 and in final form 31 May 2001.
Address reprint requests to Dr. Keith L. Constantine, Bristol-Myers Squibb PRI, PO Box 4000, Princeton, NJ 08543-4000. Tel.: 609-252-6926; Fax: 609-252-6030; E-mail: keith.constantine@bms.com.

0006-3495/01/09/1275/10 $2.00

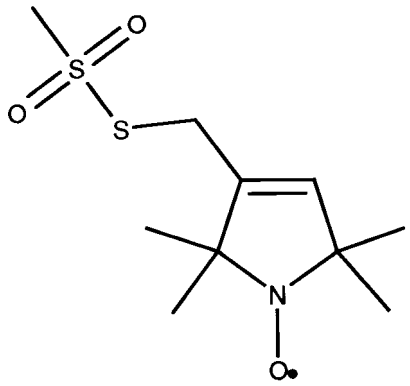

FIGURE 1 A commonly used cysteine-specific spin label ((1-oxy-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methanethiosulfonate). The simulations reported in this article are based on using this compound to produce spin-labeled proteins.

A general method for incorporating paramagnetic centers into proteins is site-directed spin labeling (SDSL) (Hubbell and Altenbach, 1994; Hubbell et al., 1998). In this approach, a nitroxide-containing compound, such as a methanethiosulfonate spin label (Fig. 1), is covalently attached to a cysteine residue in a protein containing only one such residue. Site-directed mutagenesis is used to introduce a single cysteine residue at a suitable surface location and, if necessary, to remove native cysteines. Dipolar interactions with the unpaired electron of the spin label will enhance the relaxation rates of protons within a range of ~25–30 Å, allowing distance restraints to be obtained. SDSL has recently been applied to characterize the denatured states of staphylococcal nuclease (Gillespie and Shortle, 1997a,b) and protein L (Yi et al., 2000) and for determining the global folds of proteins in their native state (Battiste and Wagner, 2000; Gaponenko et al., 2000).

In addition to the applications mentioned above, SDSL may also provide structural information for bound ligands, given a known protein structure. This method should be applicable in cases where intermolecular NOE-based methods encounter difficulties, such as with weakly binding ligands, or with large proteins that are unassigned or only partially assigned. Several protein mutants, each with a spin label attached at a different strategic location near the ligand-binding site, could provide detailed structural data.

In this article, the potential utility of SDSL for characterizing protein-ligand complexes is addressed by simulations, using a 1.9-Å-resolution x-ray structure of a mutant dihydrofolate reductase complexed with methotrexate (Brown et al., 1993) (Protein Data Bank (PDB) entry 1DHI) and a 1.5-Å-resolution x-ray structure of the V-Src tyrosine kinase SH2 domain complexed with a five-residue phosphopeptide (Waksman et al., 1992) (PDB entry 1SHA) as model systems. Given the potential flexibility of the spin-labeled side chain (Mchaourab et al., 1996, 1999) and the long-range nature of the electron-proton distance restraints, it is important to investigate the utility of such restraints for defining ligand-binding modes by simulation studies before undertaking experimental work. Herein, the ability of SDSL restraints to reproduce two known protein-ligand complexes is evaluated. Approximate distance restraints have been derived from ensembles that widely sample the conformational space accessible to the spin-labeled side chains. These restraints are used in simulated annealing calculations, and their ability to reproduce the overall target ligand-binding modes and internal conformations is assessed. The effects of incorporating a minimal number of short-range (NOE-type) distance restraints on the accuracy of the computed binding modes are also evaluated. The results indicate that SDSL restraints can contribute significantly to defining the orientations and conformations of bound ligands. Accurate ligand localization is more difficult to achieve using only SDSL restraints. Potential applications and improvements to the methodology are discussed.

Background

To obtain SDSL distance restraints to a bound ligand, the contributions to the bound-state proton longitudinal and/or transverse relaxation rates due to the paramagnetic center must be determined. For small peptides and drug-like organic compounds, the ligand's proton spectrum will generally be well resolved, allowing the use of one-dimensional (1D) pulse sequences for measurement of the longitudinal (R_1) and transverse (R_2) relaxation rates. For more complex ligands, two-dimensional (2D) homonuclear methods for measuring proton relaxation rates have been developed (Arseniev et al., 1986; Kay and Prestegard, 1988). Measurements should be feasible for both the slow and fast ligand exchange regimes. For the latter, relaxation rates could be determined for the averaged (free and bound) ligand resonance positions. In the case of slow exchange, protein ^{13}C, ^{15}N, and ^{2}H labeling combined with isotopic filtering methods (Fesik, 1988, 1991; Otting and Wüthrich, 1990; Breeze, 2000) should allow ligand proton relaxation rates to be measured. In the fast exchange case, the fraction of ligand bound to the protein (f_b) must be known.

For estimating absolute electron-proton distances, the correlation time τ_c for each electron-proton dipolar interaction must be determined or approximated. For individual electron-proton dipolar interactions, τ_c can be determined from the ratio of the bound-state paramagnetic contributions to R_1 ($R_{1,para}$) measured at two different field strengths, from the ratio of the bound-state paramagnetic contributions to R_2 ($R_{2,para}$) measured at two different field strengths, or from the ratio $R_{1,para}/R_{2,para}$ at one field strength. Alternatively, all of the τ_c values can be approximated by τ_r, the overall rotational correlation time for the protein-ligand complex. This latter approach is feasible when τ_r is short relative to both the electronic relaxation time and the bound-state lifetime. However, this approach will introduce some

systematic errors into the derived distances depending on how much τ_r differs from the true individual τ_c values, e.g., due to internal dynamics or rotational anisotropy. $R_{1,para}$ and $R_{2,para}$ are functions of τ_c and the electron-proton distance r (Kosen, 1989).

METHODS

Derivation of simulated distance restraints

All calculations were performed with the X-PLOR program (Brünger, 1992), unless noted otherwise. The coordinates of the mutant dihydrofolate reductase-methotrexate complex were obtained from the PDB (entry 1DHI), as were the coordinates of the V-Src tyrosine kinase SH2 domain complexed with a five-residue phosphopeptide (pY-V-P-M-L; entry 1SHA). All of the hydrogen atoms were built on to the proteins and ligands using the HBUILD routine of X-PLOR. Parameters for methotrexate were generated using QUANTA (MSI, Burlington, MA). Bond lengths and angles for the phosphate group of the phosphopeptide were derived directly from the 1SHA coordinates. Side chains corresponding to those obtained by reacting the methanethiosulfonate spin label shown in Fig. 1 with a cysteine residue were built on to positions 18, 33, 71, 109, and 146 of dihydrofolate reductase and on to 11, 52, 65, and 74 of the SH2 domain, using the BUILD routines of X-PLOR. These positions were chosen by visual inspection. Parameters defining the bond lengths and bond angles for this residue type were derived by calculations on the nitroxide radical performed with the SAM1 (Dewar et al., 1993) semi-empirical SCF-MO methodology. The restricted open shell Hartree-Fock approach was used to treat the radical's doublet electronic state, and the molecular geometry was fully optimized.

After building on the hydrogens, the dihydrofolate reductase-methotrexate complex was subjected to unrestrained energy minimization using the Powell method. The resulting conformation of the complex has an overall protein backbone atom root mean square difference (RMSD) to the original x-ray structure of 0.65 Å. Likewise, the SH2 domain-phosphopeptide complex was energy minimized, and the resulting conformation of the complex has an overall backbone atom RMSD (protein and phosphopeptide) to the original x-ray structure of 0.78 Å. These energy-minimized structures served as the target binding modes for all subsequent calculations.

Fifty conformations of each side chain were generated by unrestrained simulated annealing (Nilges et al., 1988) by varying the initial velocity random number seed. All atoms of the spin-labeled residues were allowed to move during these calculations. For both systems, the ligand and the remainder of the protein were held fixed throughout. The spin label positions are well separated in both complexes, allowing all spin-labeled side chains to be present simultaneously during the calculations. The high-temperature phase of annealing protocol consisted of 10 ps at 1000 K. The form of the restraint potential and repulsive non-bonded interactions were adjusted to their final settings, and the temperature was gradually lowered to 100 K, followed by a final 2000 steps of unrestrained Powell minimization. Electrostatic, attractive van der Waals, and empirical dihedral terms were excluded from the potential function.

In the case of the dihydrofolate reductase-methotrexate complex, 2 of the 50 structures with varied spin-labeled side-chain conformations were deleted from the ensemble due to unrealistic, high-energy burial of a spin-labeled side chain. All 50 structures were retained in the case of the SH2 domain-phosphopeptide complex. Using the ensembles of structures with varied spin-labeled side-chain conformations, effective distances were computed from the nitrogens of the spin-labeled side chains to ligand protons. Effective distances were computed as $\langle r^{-3}\rangle^{-1/3}$ averaged over all structures. This assumes that any internal motions that modulate r (e.g., motions of the spin-labeled side chain) are fast relative to τ_r. Averaging was also performed over all methylene pairs, methyl protons, and equivalent aromatic protons. Prochiral protons and prochiral methyl groups were assumed not to be stereospecifically assigned. For both the dihydrofolate reductase-methotrexate and the SH2 domain-phosphopeptide complexes, three sets of SDSL restraints were derived by adding (for upper bounds) and subtracting (for lower bounds) 5%, 10% or 20% of the calculated effective distance from each restraint. These bounds were selected to span the likely error range of experimental restraints. In practice, distance errors can be obtained from errors of the relaxation rates and standard error propagation methods (e.g., Jacob et al., 1999; Johnson et al., 1999). For effective distances >25 Å, lower bounds only of 20 Å were used. The use of lower bound restraints assumes that protons more than 25 Å from the spin label would experience a negligible relaxation enhancement. For both complexes, there were no effective distances <8 Å. The effects of eliminating a significant fraction of the restraints were examined in the case of the SH2 domain-phosphopeptide complex.

For the SH2 domain-phosphopeptide complex, nine short-range (<5.0 Å) distance restraints were derived. These include one restraint to enforce a salt bridge between the phosphate of the peptide and the Arg 32 side chain, which is a conserved interaction in all known SH2 domain-phosphopeptide structures. The remaining eight short-range restraints simulate NOEs that could be derived from an isoleucine-leucine-valine-labeled sample (Metzler et al., 1996), assuming a 5.0-Å cutoff for observing an intermolecular NOE. For the dihydrofolate reductase-methotrexate complex, eight short-range restraints were derived between ligand protons and backbone amide protons of the protein, again assuming a 5.0-Å cutoff. These simulate NOEs that could be obtained for a $^{2}H/^{13}C/^{15}N$-labeled protein for which only backbone atom assignments are available, and for which only NOEs to exchangeable protons of the protein are observable. The short-range restraints were classified into bins corresponding to strong, medium, and weak NOEs on the basis of the distance observed in the target complex. The strong, medium, and weak NOE bins correspond to upper bounds of 2.5 Å, 3.5 Å (SH2 domain-phosphopeptide) or 4.0 Å (dihydrofolate reductase-methotrexate), and 5.0 Å, respectively, with additions for equivalent and prochiral protons. Lower bounds for the short-range restraints were set uniformly to 1.8 Å.

Modeling based on simulated distance restraints

The following protocol was used for both systems utilizing the simulated distance restraints. Structures were calculated by restrained simulated annealing (Nilges et al., 1988), using the ensembles with varied spin-labeled side-chain conformations described above for the starting coordinates. The ligands were subjected to random translations (within ±20 Å along *x*, *y*, and *z*) and random rotations (0–360° around the *x*, *y*, and *z* axes) to randomize the starting locations and orientations. The ligands and all atoms of the spin-labeled residues were allowed to move during the simulated annealing calculations; all other protein atoms were held fixed. Distance restraints involving methylene, methyl, and equivalent aromatic protons were treated using the "R-3" averaging option in X-PLOR (Brünger, 1992). A force constant of 50 kcal/mol Å^2 was used for all distance restraints. Electrostatic, attractive van der Waals, and empirical dihedral terms were excluded from the potential function.

The simulated annealing protocol (Nilges et al., 1988) consisted of 15 ps at 1000 K, followed by restraint potential and non-bonded interaction adjustment, temperature lowering to 100 K, and a final 2000 steps of restrained Powell minimization. Structures were selected for the final ensembles on the basis of their total restraint energies. RMSD values for the ligands were calculated in the reference frame of the protein and after superposition of the calculated ligand conformations on to the target ligand conformations. The former reflects differences in location, orientation, and conformation, whereas the latter is determined by internal conformational differences only. Structures were visualized using InsightII, version 98.0 (MSI); this program was used to produce Figs. 2–7.

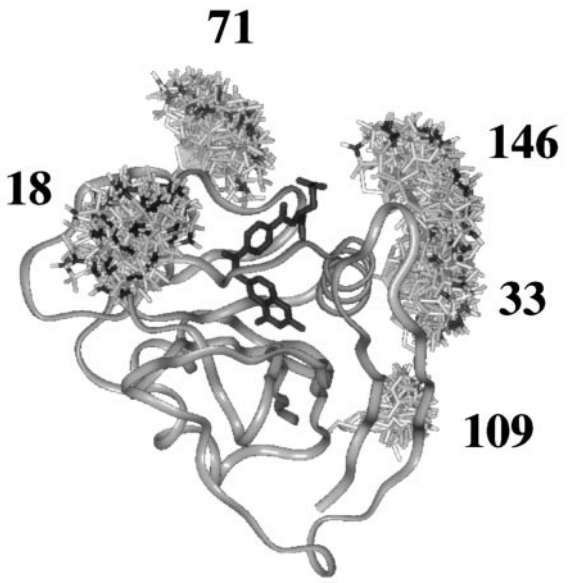

FIGURE 2 Ensemble of 48 structures of dihydrofolate reductase complexed with methotrexate, with randomized spin-labeled residue conformations. The protein backbone conformation is depicted by the medium gray ribbon, and all heavy atoms of the methotrexate (target binding mode) are shown in black. The spin-labeled side chains (with residue numbers) are shown in light gray, with the side-chain nitrogen atoms shown in black. Effective electron-proton distance restraints were derived from this ensemble.

TABLE 1 Methotrexate RMSD values (Å) for the five computed structural ensembles of the dihydrofolate reductase-methotrexate complex

	ENS1	ENS2	ENS3	ENS4	ENS5
N_{struct}*	48	48	46	23	48
RMSD1	1.17	1.65	0.95	1.57	1.13
RMSD2	1.50	2.52	1.06	2.34	1.04
RMSD3	2.81	4.94	1.95	4.11	1.86

ENS1, ensemble obtained with SDSL restraints with ±10% error bounds; ENS2, ensemble obtained with SDSL restraints with ±20% error bounds; ENS3, ensemble obtained with SDSL restraints with ±5% error bounds; ENS4, ensemble obtained with short-range distance restraints only; ENS5, ensemble obtained with SDSL restraints (±10%) and short-range distance restraints; RMSD1, average heavy atom RMSD to the target methotrexate conformation, after superposition on to the target methotrexate conformation; RMSD2, average heavy atom RMSD to the average methotrexate structure in the reference frame of the protein; RMSD3, average heavy atom RMSD to the target methotrexate structure in the reference frame of the protein.

*Number of converged structures in the ensemble.

RESULTS

Dihydrofolate reductase-methotrexate complex

Spin-labeled side chains corresponding to the product of reacting (1-oxy-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)-methanethiosulfonate (Fig. 1) with a cysteine residue were built on to five selected positions of dihydrofolate reductase. These side chains simulate those that would be obtained for the following spin-labeled single-site mutants: N18C, R33C, R71C, K109C, and Q146C. The residues chosen for introducing the spin label were selected based on solvent exposure, proximity to the binding pocket, lack of direct interactions with the ligand, and good separation between the locations of attachment. The K109C mutant (Fig. 2) is particularly noteworthy, because this spin label is in a position that can be characterized as behind or beneath the ligand-binding pocket.

Fig. 2 shows an ensemble of 48 structures in which the conformations of the five spin-labeled residues were randomized by unrestrained simulated annealing. Fig. 2 also shows the target binding mode of methotrexate. Effective distances (see Methods) were calculated from the side-chain nitrogen atom of each spin label to protons of methotrexate. After averaging over the methyl, methylene, and NH_2 group protons, this procedure yielded 50 effective distances. Of these, only one was greater than 25 Å; this restraint was incorporated with only a lower bound (20 Å). The remaining 49 restraints ranged between 13 and 25 Å and were treated with varying levels of precision, as described below. For all of the resulting structural ensembles discussed below, there are no SDSL restraint violations >0.1 Å.

In the following, structural ensembles are characterized by three different RMSD values, referred to as RMSD1, RMSD2, and RMSD3, as described and summarized in Table 1. RMSD1 reflects the similarity of the internal conformations of the computed methotrexate structures to the internal conformation of the target, RMSD2 measures the precision with which the methotrexate binding mode is determined without reference to the target structure, and RMSD3 gauges how well the target binding mode (location, orientation, and internal conformation) is reproduced.

Fig. 3 shows an ensemble of 48 converged structures (ENS1) obtained with SDSL restraints. The 49 SDSL restraints ranging between 13 and 25 Å were incorporated using error bounds of ±10% of the effective distances. In Fig. 3, the superposition in the reference frame of the protein is shown. The overall binding mode, as reflected by RMSD3 (2.81 Å; Table 1), is reproduced with modest precision. Methotrexate tends to localize slightly above the target location. With reference to the average structure rather than the target binding mode, the localization is better defined (RMSD2 = 1.50 Å; Table 1). The internal conformation with respect to the target is reproduced moderately well (RMSD1 = 1.17 Å; Table 1).

The results described above indicate that SDSL restraints with ±10% error bounds do not precisely define the bound location of methotrexate. To ascertain the effects of increasing or decreasing the error bounds on the SDSL restraints, two additional ensembles were calculated. An ensemble of 48 converged structures (ENS2) was obtained with error bounds of ±20% on the SDSL restraints. The RMSD values for ENS2 (Table 1) demonstrate a decrease in precision

FIGURE 3 Ensemble of 48 converged structures of dihydrofolate reductase complexed with methotrexate obtained using SDSL restraints with ±10% error bounds, using no short-range restraints. Structures are shown in the reference frame of the protein coordinates. The protein backbone conformation is depicted by the medium gray ribbon, and all heavy atoms of the methotrexate in its target binding mode are depicted by the thick black stick diagram. The methotrexate binding modes derived by restrained simulated annealing are depicted by thin dark gray stick diagrams. The spin-labeled side chains are shown in light gray, with the side-chain nitrogen atoms shown in black.

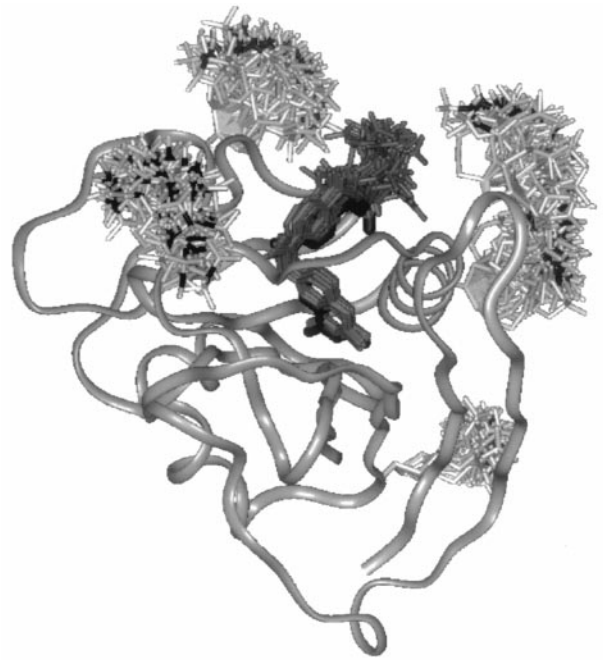

FIGURE 4 Ensemble of 48 converged structures of dihydrofolate reductase complexed with methotrexate obtained using SDSL restraints with ±10% error bounds, using also eight short-range restraints. Structures are shown in the reference frame of the protein coordinates. The gray-scale coding used is the same as that in Fig. 3.

relative to ENS1, both the overall binding mode and internal conformation. In particular, the RMSD relative to the target binding mode increases to 4.94 Å. Decreasing the error bounds on the SDSL restraints to ±5% produced an ensemble of 46 converged structures (ENS3). The RMSD values for ENS3 (Table 1) demonstrate an increase in precision relative to ENS1 and ENS2. The RMSD value for overall binding mode with respect to the target binding mode (RMSD3) is 1.95 Å, and the RMSD value of the internal conformation with respect to the target conformation (RMSD1) is 0.95 Å (Table 1).

To determine the effects of incorporating a few short-range restraints, two additional ensembles were calculated. An ensemble (ENS4) was derived using only eight short-range restraints between ligand protons and backbone amide protons of dihydrofolate reductase. This ensemble serves as a necessary benchmark for calculations that combine the short-range and SDSL restraints. The short-range restraints alone do not define the binding mode very well (Table 1). The eight short-range restraints were combined with SDSL restraints (±10% error bounds) to produce ENS5 (Fig. 4). This ensemble has considerably lower values for RMSD2 and RMSD3 (Table 1) than the ensemble produced using just the SDSL restraints with ±10% error bounds (ENS1). The accuracy of the binding mode displayed by ENS5 is similar to that obtained using tight (±5% error bounds) SDSL restraints only (ENS3). These results indicate that short-range and SDSL restraints can act synergistically to produce a relatively well-defined binding mode.

V-Src SH2 domain-phosphopeptide complex

Spin-labeled side chains were built on to four selected positions of the SH2 domain that simulate those that would be obtained for the following spin-labeled single-site mutants: T11C, K52C, D65C, and R74C. Unlike the dihydrofolate reductase-methotrexate case, no spin-labeled side chain can be characterized as behind or beneath the binding groove.

Fig. 5 shows the target phosphopeptide binding mode and an ensemble of 50 structures with randomized conformations of the four spin-labeled residues. Effective distances were calculated from each spin label to non-exchangeable protons of the phosphopeptide. After averaging over the methyl, methylene, and equivalent aromatic protons, this procedure yielded 76 effective distances. Of these, 14 were greater than 25 Å; these were treated as restraints with only lower bounds (20 Å). The remaining 62 restraints ranged between 8 and 25 Å. For all of the structural ensembles discussed below, there are no SDSL restraint violations

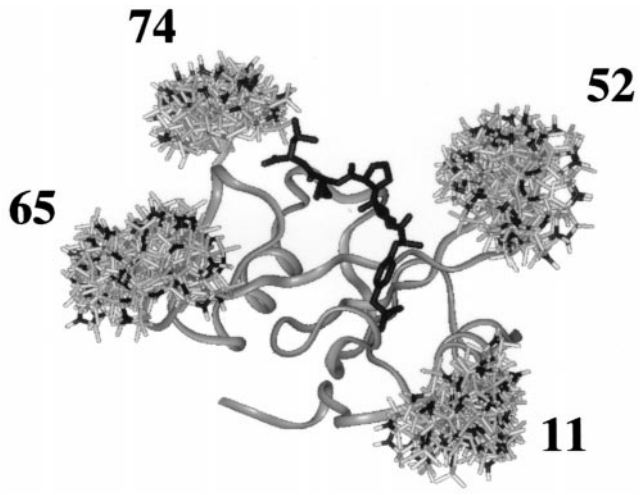

FIGURE 5 Ensemble of 50 structures of the V-Src tyrosine kinase SH2 domain complexed with a five-residue phosphopeptide, with randomized spin-labeled residue conformations. The protein backbone conformation is depicted by the medium gray ribbon, and all heavy atoms of the phosphopeptide (target binding mode) are shown in black. The spin-labeled side chains (with residue numbers) are shown in light gray, with the side-chain nitrogen atoms shown in black. Effective electron-proton distance restraints were derived from this ensemble.

TABLE 2 Phosphopeptide RMSD values (Å) for the six computed structural ensembles of the V-Src tyrosine kinase SH2 domain/phosphopeptide complex

	ENS1	ENS2	ENS3	ENS4	ENS5	ENS6
N_{struct}*	46	50	47	48	22	50
RMSD1	0.88	1.46	0.55	0.76	0.35	0.99
RMSD2	1.87	2.48	1.41	1.63	1.20	1.99
RMSD3	2.95	1.46	0.71	0.94	0.46	1.09
RMSD4	3.39	1.80	0.93	1.25	0.70	1.42
RMSD5	5.01	2.50	1.35	1.51	1.09	2.13
RMSD6	5.81	2.99	1.63	1.91	1.41	2.45

ENS1, ensemble obtained with SDSL restraints (±10%), using no short-range distance restraints; ENS2, ensemble obtained with short-range distance restraints only; ENS3, ensemble obtained with SDSL restraints (±10%) and short-range distance restraints; ENS4, ensemble obtained with SDSL restraints (±20%) and short-range distance restraints; ENS5, ensemble obtained with SDSL restraints (±5%) and short-range distance restraints; ENS6, ensemble obtained with the reduced number of SDSL restraints (±10%) and short range restraints. RMSD1, average backbone atom (N, C, and C^{α}) RMSD to the target phosphopeptide conformation, after superposition on to the target phosphopeptide backbone atoms; RMSD2, average all heavy atom RMSD to the target phosphopeptide conformation, after superposition on to all heavy atoms of the target phosphopeptide; RMSD3, average backbone atom (N, C, and C^{α}) RMSD to the average phosphopeptide structure in the reference frame of the protein; RMSD4, average all heavy atom RMSD to the average phosphopeptide structure in the reference frame of the protein; RMSD5, average backbone atom (N, C, and C^{α}) RMSD to the target phosphopeptide structure in the reference frame of the protein; RMSD6, average all heavy atom RMSD to the target phosphopeptide structure in the reference frame of the protein.

*Number of converged structures in the ensemble.

>0.1 Å. The SH2 domain-phosphopeptide structural ensembles are characterized by six different RMSD values (RMSD1 through RMSD6) for both backbone atoms and all heavy atoms, as described in Table 2 .

Fig. 6 shows an ensemble of 46 converged structures (ENS1) obtained with the 14 lower bound SDSL restraints and with the 62 SDSL restraints ranging between 8 and 25 Å; the latter were incorporated using error bounds of ±10% of the effective distances. In Fig. 6 *a*, the superposition in the reference frame of the protein is shown. The overall binding mode, as reflected by RMSD5 and RMSD6 (Table 2), is not reproduced very well. Both of these values exceed 5.0 Å. The phosphopeptide tends to localize well above the binding groove, because there are no restraints to direct the phosphopeptide deep into the binding groove. With reference to the average structure rather than the target binding mode, the localization is still not very well defined, with RMSD3 = 2.95 Å and RMSD4 = 3.39 Å (Table 2). The internal conformation with respect to the target (Fig. 6 *b*) is reproduced moderately well, with RMSD1 = 0.88 Å and RMSD2 = 1.87 Å (Table 2).

The SDSL restraints alone do not accurately define the location of the bound phosphopeptide. As in the dihydrofolate reductase-methotrexate case, the effects of adding several short-range distance restraints were examined. In total, nine short-range restraints were incorporated into the subsequent calculations. One of these enforces a hydrogen bond between the phosphate group of the phosphopeptide and the strictly conserved R32 side chain (PDB entry 1SHA residue numbering). The remaining eight restraints are between M4 of the phosphopeptide and I71 and L94 of the protein. These represent NOE restraints that could be derived using an isoleucine-leucine-valine-labeled protein sample (Metzler et al., 1996).

An ensemble of 50 converged structures (ENS2) was computed using only the short-range restraints. The short-range restraints alone result in somewhat better localization than that obtained using only the SDSL restraints, with both RMSD5 and RMSD6 less than 3.0 Å (Table 2) in the former case. However, the internal conformation with respect to the target is reproduced less well, with RMSD1 = 1.46 Å and RMSD2 = 2.48 Å for ENS2 (Table 2).

Using the short-range restraints in conjunction with the SDSL restraints (±10% error bounds), an ensemble of 47 converged structures (ENS3) was obtained (Fig. 7). This set of structures is well defined, both with respect to overall binding mode and internal conformation (Table 2). Focusing on RMSD5 (1.35 Å) and RMSD6 (1.63 Å), it is clear that the overall binding mode (Fig. 7 *a*) is much better defined than with SDSL restraints alone (ENS1) or with the short-range restraints alone (ENS2). The internal conformation (Fig. 7 *b*) is also well defined, with an average backbone atom RMSD to the target (RMSD1) of 0.55 Å.

Three additional ensembles of the SH2 domain-phosphopeptide system were calculated to ascertain the effects

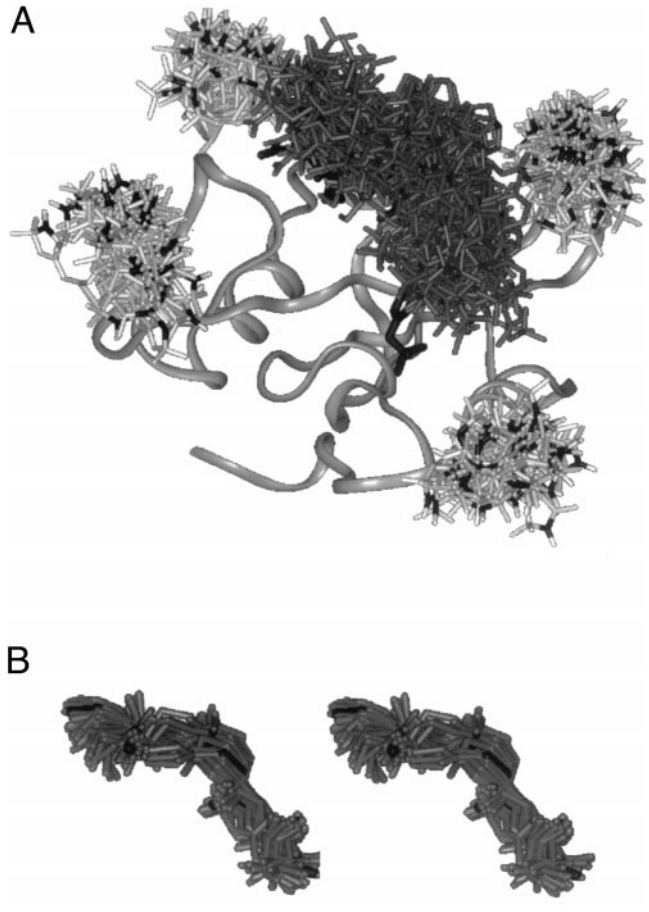

FIGURE 6 Ensemble of 46 converged structures of the V-Src tyrosine kinase SH2 domain complexed with a five-residue phosphopeptide obtained using SDSL restraints with ±10% error bounds, using no short-range distance restraints. (*a*) Structures in the reference frame of the protein coordinates. The protein backbone conformation is depicted by the medium gray ribbon, and all heavy atoms of the phosphopeptide in its target binding mode are depicted by the thick black stick diagram. The phosphopeptide binding modes derived by restrained simulated annealing are depicted by thin dark gray stick diagrams. The spin-labeled side chains are shown in light gray, with the side-chain nitrogen atoms shown in black. (*b*) Stereoview (relaxed eye) of the phosphopeptide backbone N, C, C^{α}, and O atoms, with the conformations derived by restrained simulated annealing (dark gray) superimposed on to the target peptide backbone conformation (black).

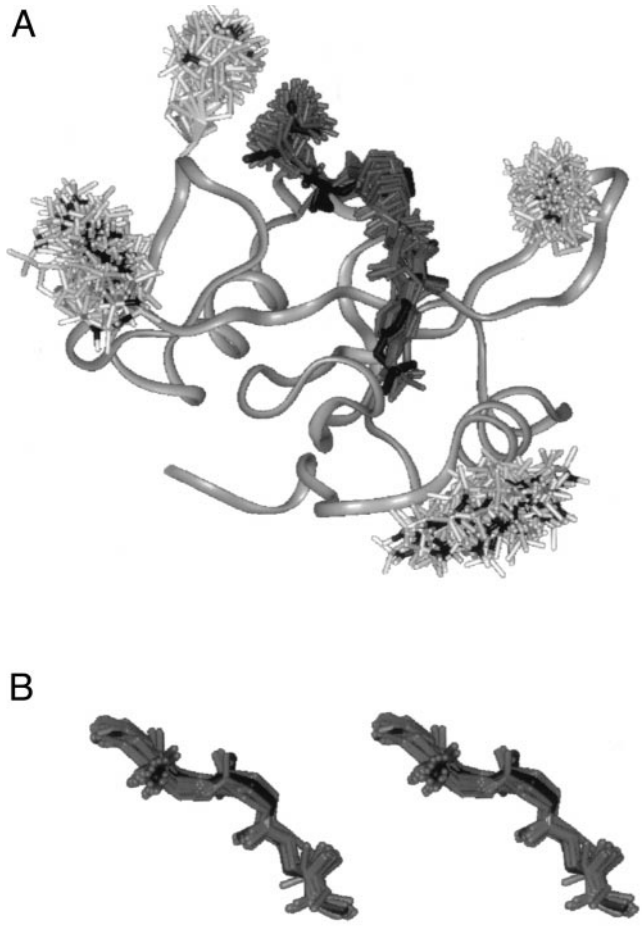

FIGURE 7 Ensemble of 47 converged structures of the V-Src tyrosine kinase SH2 domain complexed with a five-residue phosphopeptide obtained using SDSL restraints with ±10% error bounds, using also nine short-range distance restraints. (*a*) Structures in the reference frame of the protein coordinates. The gray-scale coding used is the same as that in Fig. 6 *a*. (*b*) Stereoview (relaxed eye) of the phosphopeptide backbone N, C, C^{α}, and O atoms, with the conformations derived by restrained simulated annealing (dark gray) superimposed on to the target peptide backbone conformation (black).

of increasing or decreasing the error bounds on the SDSL restraints while retaining the short-range restraints, and to ascertain the effects of eliminating a significant fraction of the restraints. An ensemble of 48 converged structures (ENS4) was obtained with error bounds of ±20% on the full set of SDSL restraints. Whereas the RMSD values for ENS4 (Table 2) demonstrate a decrease in precision relative to ENS3, both the overall binding mode and internal conformation are still better defined than in ENS1 or ENS2. Decreasing the error bounds on the full set of SDSL restraints to ±5% produced an ensemble of 22 converged structures (ENS5). (Convergence proved to be significantly more difficult to obtain with the tight error bounds in this system.) The RMSD values for ENS5 (Table 2) demonstrate an increase in precision relative to ENS3. The overall binding mode and internal conformation are well defined in ENS5. Finally, an ensemble (ENS6) was computed in which all restraints involving the following ligand ^{1}H resonances were eliminated: V2 β and γ-methyl; P3 β and γ; M4 β and L5 β; and γ and δ-methyl. These resonances are most likely to be involved in overlaps in a 1D ^{1}H-NMR spectrum. For ENS6, 8 short-range and 44 SDSL restraints (±10% error bounds) were retained. The RMSD values for ENS6 (Table 2) are between those of ENS2 and ENS3, indicating that a reduced number of SDSL restraints still produces an improvement over the results obtained with short-range restraints only (ENS2).

DISCUSSION

The simulations described in this article are based on the following considerations. 1) It is assumed that a majority of ligand proton resonances can be assigned and that relaxation rates can be measured. Stereospecific assignments are not assumed for prochiral methylene protons or prochiral methyl groups when deriving restraints. 2) It is assumed that paramagnetic contributions to the relaxation rates can be measured for ligand protons with effective distances between 8 and 25 Å of the nitrogen of the spin-labeled side chain. This range is consistent with those reported in the literature (Gillespie and Shortle, 1997b; Dunham et al., 1998; Jacob et al., 1999; Johnson et al., 1999; Battiste et al., 2000; Gaponenko et al., 2000). 3) It is assumed that a suitable structure of the protein in question is available and that the boundaries of the binding pocket are known reasonably well. 4) It is assumed that the mode of ligand binding is unperturbed by the spin label. This could be supported experimentally by measuring the affinity of the ligand for each spin-labeled protein. For proteins with a resolved ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum, highly similar protein chemical shift changes should be induced by ligand binding if the mode is unperturbed. 5) It is assumed that the spin-labeled side chains must be treated as flexible. Even if a particular spin-labeled residue is relatively rigid due to its location (Mchaourab et al., 1996), it is assumed that its conformation is unknown. 6) Several short-range distance restraints may be required for accurate ligand localization. Alternatively, moderately accurate ligand localization can be achieved with tight SDSL restraints when at least some of these restraints direct the ligand into the binding site. However, high levels of both precision and accuracy on the SDSL restraints may be difficult to achieve, and a suitable spin label placement beneath or behind the binding site will not be feasible in general, especially for larger proteins.

An advantage of the approach is that the number of restraints obtainable is approximately given by the number of spin-labeled sites multiplied by the number of resolvable ligand proton resonances. Protons that are too close or too far from the paramagnetic center for measurement of the paramagnetic contributions to their relaxation rates can still be restrained with upper or lower distance bounds, respectively. This consideration also holds for protein global fold determination, where restraints can be obtained for all peaks that are resolved and assigned in the 2D ^{1}H-^{15}N HSQC spectrum (Battiste and Wagner, 2000; Gaponenko et al., 2000). The approach should be applicable to all proteins with known structures for which suitable single cysteine mutants can be obtained. For proteins containing disulfides or cysteines critical to their function, site-directed spin labeling of residues other than cysteines could be utilized (Berliner, 1976; Schmidt and Kuntz, 1984; Hankovszky et al., 1987; Musci et al., 1988).

There are several limitations to the approach as well. It will not be applicable to ligands with severely exchange-broadened or highly overlapped resonances. The distribution of the spin-labeled side-chain conformations obtained by simulated annealing may differ from the true distribution. The precision and accuracy of the restraints will be affected by random errors associated with measurements of the paramagnetic contributions and systematic errors arising from inaccurate estimates of the correlation time and/or fraction of bound ligand. Nonlinear averaging of the effective distances due to internal dynamics will also affect the precision and accuracy of experimental restraints. Given these considerations, error bounds on the distance restraints of between $\pm$10% and $\pm$20% are likely to be required. Nevertheless, even with $\pm$20% error bounds, a moderately well defined binding mode was obtained when a few supplementary short-range restraints were included (Table 2).

Supplementary short-range distance restraints could be obtained in several different ways. For proteins that can be partially assigned, residue-type-specific labeling can be applied to obtain protein-ligand NOEs. For proteins that cannot be assigned, SDSL restraints could be used to check binding modes produced by docking programs (Lambert, 1997) for consistency. Distance restraints reflecting key interactions could be derived from alternate binding modes suggested by docking and used in conjunction with SDSL restraints in structure calculations.

The simulated annealing protocol used to randomize the spin-labeled side chains produced a wide sampling of conformational space (Figs. 2 and 5). This probably represents a worst-case scenario regarding the flexibility of these side chains, because it is known from electron paramagnetic resonance (EPR) studies (Mchaourab et al., 1996, 1999) and crystallographic studies (Langen et al., 2000) that the actual conformational mobility may be restricted. For proteins with backbone assignments, distance restraints from the spin-labeled side chains to backbone amide protons (Battiste and Wagner, 2000; Gaponenko et al., 2000) could be used in conjunction with restraints to the ligand to further restrict the conformational freedom of the system. With a larger database of crystallographic information, with the acquisition of EPR information, and/or with the use of more conformationally restricted spin labels, it may become possible to place a priori restrictions on the spin-labeled side-chain conformations. The currently available crystallographic data (Langen et al., 2000) indicate that g^{-} χ_1 rotamer of the spin-labeled side chain used here is disfavored.

Finally, the flexibility of the protein will need to be addressed if a given ligand produces changes in the protein's conformation relative to the available structures. For relatively small changes, it should be sufficient to allow a more extensive region of the complex to relax during a final restrained energy minimization. More extensive changes will likely require the incorporation of additional informa-

tion, such as intra-protein NOEs and SDSL restraints to the protein, to accurately define the structure of the complex.

CONCLUSIONS

The results reported in this article indicate that protein-ligand SDSL restraints have the potential of providing valuable information on ligand binding modes. When combined with several short-range restraints, the overall ligand binding mode can be defined relatively well. The combined restraints yield a significant improvement over results obtained using either short-range or SDSL restraints alone.

I thank Dr. Andrew Pudzianowski for performing the quantum mechanical calculations on (1-oxy-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methanethiosulfonate. Careful readings of the manuscript by Drs. Mark Friedrichs, Luciano Mueller, Patricia McDonnell, and William Metzler are gratefully acknowledged. The distance restraints used to calculate the structural ensembles are available upon request.

REFERENCES

Anglister, J., T. Frey, and H. M. McConnell. 1984. Distances of tyrosine residues from a spin-labeled hapten in the combining site of a specific monoclonal antibody. *Biochemistry.* 23:5372–5375.

Arrowsmith, C. H., and Y.-S. Wu. 1998. NMR of large (>25 kDa) proteins and protein complexes. *Prog. NMR Spectrosc.* 32:277–286.

Arseniev, A. S., A. G. Sobol, and V. F. Bystrov. 1986. T_1 relaxation measurement by two-dimensional NMR spectroscopy. *J. Magn. Reson.* 70:427–435.

Battiste, J. L., and G. Wagner. 2000. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry.* 39: 5355–5365.

Berliner, L. J. (ed.) 1976. Spin Labeling Theory and Applications. Academic Press, New York.

Berliner, L. J. (ed.) 1979. Spin Labeling II Theory and Applications. Academic Press, New York.

Breeze, A. L. 2000. Isotope-filtering NMR methods for the study of biomolecular structure and interactions. *Prog. NMR. Spectrosc.* 36: 323–372.

Brown, K. A., E. E. Howell, and J. Kraut. 1993. Long-range structural effects in a second site revertant of a mutant dihydrofolate reductase. *Proc. Natl. Acad. Sci. U.S.A.* 90:11753–11756.

Brünger, A. T. 1992. X-PLOR. A System for X-ray Crystallography and NMR. Yale University Press, New Haven, CT.

Campbell, A. P., and B. D. Sykes. 1993. The two-dimensional transferred nuclear Overhauser effect: theory and practice. *Annu. Rev. Biophys. Biomol. Struct.* 22:99–122.

Clore, G. M., and A. M. Gronenborn. 1982. Theory and applications of the transferred nuclear Overhauser effect to the study of the conformations of small ligands bound to proteins. *J. Magn. Reson.* 48:402–417.

Clore, G. M., and A. M. Gronenborn. 1983. Theory of the time dependent transferred nuclear Overhauser effect: applications to structural analysis of ligand-protein complexes in solution. *J. Magn. Reson.* 53:423–442.

de Jong, E. A. M., J. P. M. van Duynhoven, B. J. M. Harmsen, G. I. Tesser, R. N. H. Konings, and C. W. Hilbers. 1989. Two-dimensional ^{1}H nuclear magnetic resonance studies on the gene V-encoded single-stranded DNA-binding protein of the filamentous bacteriophage Ike. II. Characterization of the DNA-binding wing with the aid of spin-labelled oligonucleotides. *J. Mol. Biol.* 206:133–152.

Dewar, M. J. S., C. Jie, and J. Yu. 1993. SAM1: the first of a new series of general purpose quantum mechanical molecular models. *Tetrahedron.* 49:5003–5038.

Dunham, S. U., S. U. Dunham, C. J. Turner, and S. J. Lippard. 1998. Solution structure of a DNA duplex containing a nitroxide spin-labeled platinum d(GpG) intra- strand cross-link refined with NMR-derived long-range electron-proton distance restraints. *J. Am. Chem. Soc.* 120: 5395–5406.

Farmer, B. T., II, and R. A. Venters. 1998. NMR of perdeuterated large proteins. *In* Biological Magnetic Resonance, Vol. 16. N. R. Krishna, and L. J. Berliner, editors. Kluwer Academic/Plenum Publishers, New York. 75–120

Fesik, S. W. 1988. Isotope-edited NMR spectroscopy. *Nature.* 332: 865–866.

Fesik, S. W. 1991. NMR studies of molecular complexes as a tool in drug design. *J. Med. Chem.* 34:2937–2945.

Frederick, A. F., L. E. Kay, and J. H. Prestegard. 1988. Location of divalent ion sites in acyl carrier protein using relaxation perturbed 2D NMR. *FEBS Lett.* 238:43–48.

Gaponenko, V., J. W. Howarth, L. Columbus, G. Gasmi-Seabrook, J. Yuan, W. J. Hubbell, and P. R. Rosevear. 2000. Protein global fold determination using site-directed spin and isotope labeling. *Protein Sci.* 9:302–309.

Gardner, K. H., and L. E. Kay. 1998. The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27:357–406.

Gillespie, J. R., and D. Shortle. 1997a. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation by nitroxide spin labels. *J. Mol. Biol.* 268:158–169.

Gillespie, J. R., and D. Shortle. 1997b. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* 268:170–184.

Goto, N. K., K. H. Gardner, G. A. Mueller, R. C. Willis, and L. E. Kay. 1999. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl- protonated ^{15}N,^{13}C,^{2}H-labeled proteins. *J. Biomol. NMR.* 13:369–374.

Hankovszky, O. H., K. Hideg, E. V. Goldammer, E. Matuszak, H. Kolkenbrock, H. Tschesche, and H. R. Wenzel. 1987. New nitroxide reagents for the selective spin-labelling at the guanidino moiety of arginine residues in peptides and proteins. *Biochim. Biophys. Acta.* 916:152–155.

Holtzman, J. L. (ed.) 1984. Spin Labeling in Pharmacology, Academic Press, New York.

Hubbell, W. L., and C. Altenbach. 1994. Investigation of the structure and dynamics of membrane proteins using site-directed spin labeling. *Curr. Opin. Struct. Biol.* 4:566–573.

Hubbell, W. L., A. Gross, R. Langden, and M. A. Lietzow. 1998. Recent advances in site-directed spin labeling of proteins. *Curr. Opin. Struct. Biol.* 8:649–656.

Jacob, J., B. Baker, R. G. Bryant, and D. S. Cafiso. 1999. Distance estimates from para-magnetic enhancements of nuclear relaxation in linear and flexible model peptides. *Biophys. J.* 77:1086–1092.

Johnson, P. E., E. Brun, L. F. MacKenzie, S. G. Withers, and L. P. McIntosh. 1999. The cellulose-binding domains from *Cellulomonas fimi* β-1,4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations. *J. Mol. Biol.* 287:609–625.

Kay, L. E., and J. H. Prestegard. 1988. Spin-lattice relaxation rates of coupled spins from 2D accordion spectroscopy. *J. Magn. Reson.* 77: 599–605.

Kosen, P. A. 1989. Spin labeling of proteins. *Methods Enzymol.* 177: 86–121.

Lambert, M. H. 1997. Docking conformationally flexible molecules into protein binding sites. *In* Practical Application of Computer-aided Drug Design. P. S. Charifson, editor. Marcel Dekker, New York. 243–303.

Langen, R., K. J. Oh, D. Cascio, and W. L. Hubbell. 2000. Crystal structures of spin labeled T4 lysozyme mutants: implications for the interpretation of EPR spectra in terms of structure. *Biochemistry.* 39: 8396–8405.

Lee, L., and B. D. Sykes. 1980. Nuclear magnetic resonance determination of metal- proton distances in the EF site of carp parvalbumin using susceptibility contributions to the line broadening of lanthanide-shifted resonances. *Biochemistry.* 19:3208–3214.

Mahoney, N. M., V. K. Rastogi, S. M. Cahill, M. E. Girvin, and S. C. Almo. 2000. Binding orientation of proline-rich peptides in solution: polarity of the profilin- ligand interaction. *J. Am. Chem. Soc.* 122: 7851–7852.

McConnell, H. M. 1967. Magnetic Resonance in Biological Systems. Pergamon, Oxford.

Mchaourab, H. S., T. Kalai, K. Hideg, and W. L. Hubbell. 1999. Motion of spin-labeled side chains in T4 lysozyme: effect of side chain structure. *Biochemistry.* 38:2947–2955.

Mchaourab, H. S., M. A. Lietzow, K. Hideg, and W. L. Hubbell. 1996. Motion of spin- labeled side chains in T4. lysozyme: correlation with protein structure and dynamics. *Biochemistry.* 35:7692–7704.

Metzler, W. J., M. Wittekind, V. Goldfarb, L. Mueller, and B. T. Farmer II. 1996. Incorporation of $^{1}H/^{13}C/^{15}N$-{Ile, Leu, Val} into a perdeuterated, ^{15}N-labeled protein: potential in structure determination of large proteins by NMR. *J. Am. Chem. Soc.* 118:6800–6801.

Mueller, G. A., W. Y. Choy, D. Yang, J. D. Forman-Kay, R. A. Venters, and L. E. Kay. 2000. Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J. Mol. Biol.* 300:197–212.

Musci, G., K. Koga, and L. J. Berliner. 1988. Methionine-90-spin-labeled bovine α- lactalbumin: electron spin resonance and NMR distance measurements. *Biochemistry.* 27:1260–1265.

Ni, F., and H. A. Scheraga. 1994. Use of transferred nuclear Overhauser effect to determine the conformations of ligands bound to proteins. *Acc. Chem. Res.* 27:257–264.

Nilges, M., A. M. Gronenborn, A. T. Brünger, and G. M. Clore. 1988. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine protease inhibitor 2. *Protein Eng.* 2:27–38.

Otting, G., and K. Wüthrich. 1990. Heteronuclear filters in two-dimensional [$^{1}H,^{1}H$]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q. Rev. Biophys.* 23:39–96.

Ray, B. D., and B. D. N. Rao. 1988. ^{31}P NMR studies of enzyme-bound substrate complexes of yeast 3-phosphoglycerate kinase. II. Structure measurements using paramagnetic relaxation effects of Mn(II) and Co(II). *Biochemistry.* 27:5579–5585.

Schmidt, P. G., and I. D. Kuntz. 1984. Distance measurement in spin-labeled lysozyme. *Biochemistry.* 23:4261–4266.

Swift, T. J. 1974. NMR of Paramagnetic Molecules: Principles and Applications. Academic Press, New York.

Waksman, G., D. Kominos, S. C. Robertson, N. Pant, D. Baltimore, R. B. Birge, D. Cowburn, H. Hanafusa, B. J. Mayer, M. Overduin, M. D. Resh, C. B. Rios, L. Silverman, and J. Kuriyan. 1992. Crystal structure of the phosphotyrosine recognition domain SH2 of v-*src* complexed with tyrosine-phosphorylated peptides. *Nature.* 358:646–653.

Wider, G., and K. Wüthrich. 1999. NMR spectroscopy of large molecules and multi- molecular assemblies in solution. *Curr. Opin. Struct. Biol.* 9:594–601.

Yi, Q., M. L. Scalley-Kim, E. J. Alm, and D. Baker. 2000. NMR Characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* 299:1341–1351.