Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 168 (2004) 36–45
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities

V. Jayalakshmi and N. Rama Krishna*

Department of Biochemistry and Molecular Genetics and the NMR Core Facility, Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL 35294-2041, USA

Received 24 November 2003; revised 20 January 2004

Abstract

We describe an intensity-restrained optimization procedure for refining approximate structures of ligands within the protein binding pockets using STD-NMR intensity data on reversibly forming weak complexes. In this approach, the global minimum for the bound-ligand conformation is obtained by a hybrid structure refinement method involving CORCEMA calculation of intensities and simulated annealing optimization of torsion angles of the bound ligand using STD-NMR intensities as experimental constraints and the NOE R-factor as the pseudo-energy function to be minimized. This method is illustrated using simulated STD data sets for typical carbohydrate and peptide ligands. Our procedure also allows for the optimization of side chain torsion angles of protein residues within the binding pocket. This procedure is useful in refining and improving initial models based on crystallography or computer docking or other algorithms to generate models for the bound ligand (e.g., a lead compound) within the protein binding pocket compatible with solution STD-NMR data. This method may facilitate structure-based drug design efforts.

Keywords: CORCEMA; CORCEMA-ST; STD-NMR; Saturation transfer; Torsion angle optimization; Bound ligand conformation; Weak complexes; Structure based drug design

1. Introduction

The discovery and design of novel pharmaceutical drugs is often facilitated by a screening of suitably designed large compound libraries to identify promising lead compounds that are specific to the target protein of interest. A knowledge of the bound structure of the lead compound within the active site of the target protein will then significantly facilitate a structure-based design of a drug with desirable properties. Many of these lead compounds often exhibit only weak affinity to the target protein [1]. In this regard, NMR spectroscopy is now playing an increasingly important role as a tool in drug discovery programs. A large number of NMR-based screening methods have been developed to identify bioactive ligands. The advantages and limitations of each of these methods have been reviewed [1–3]. Additional new methods are also being proposed [4,5]. Among the NMR methods, the recently introduced STD-NMR technique [6–8] as a screening method is gaining popularity because of its applicability even when only a small quantity of a target protein is available (e.g., with ligand/protein ratios in excess of 1000), the ease of implementation, and its ability to locate the apparent group epitopes of the bound ligand in a rather qualitative manner. This method is best suited when the binding is relatively weak, with dissociation constants (K_d) in the range $\sim 10^{-3}$ to $\sim 10^{-7}$ M. We have previously presented the complete relaxation and conformational exchange matrix [CORCEMA] theory [9] of the STD-NMR experiment and described the properties of the STD-NMR spectra. We have demonstrated [9,10] that: (1) the *apparent* epitope map based on STDs can exhibit rather significant dependence on which particular protein proton(s) is being saturated: For example,

* Corresponding author. Fax: 1-205-934-6475.
E-mail address: nrk@uab.edu (N. Rama Krishna).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.01.017

saturation of resolved methyl proton resonances from two separate residues may sometimes result in different STD spectra. Similarly, saturation of a specific tyrosine ring protons vs a specific methyl group can result in different STD spectra. (2) Saturation of protein residues within the binding pocket are generally likely to produce larger STDs (because of direct effects, i.e., $E2' \rightarrow L'$ in Scheme II of [9]) than saturation of residues far away from the binding pocket. (3) The STDs for small saturation times are independent of the free ligand conformation and its correlation time, and are more reflective of the bound-ligand proton distances with respect to saturated protein residue. (4) At longer saturation times the STDs exhibit a rather complex dependence on a number of factors including free ligand conformation and correlation time, and protein indirect effects (i.e., $E2' \rightarrow E1' \rightarrow L'$ in Scheme II of [9]). (5) The STDs and hence the *apparent* epitope map can exhibit significant dependence on intra-ligand relaxation rates, thus necessitating caution in simple qualitative interpretations. (6) For dissociation constants (K_d) in the range 10^{-4}–10^{-7} M, the STDs are essentially independent of the binding constant for large ligand/protein ratios (where uncomplexed protein concentration is negligible). (7) In this range, exchange-mediated leakage from the rapidly exchanging bound-ligand (or bound water molecules [11]) can significantly reduce the efficiency of spin diffusion in transferring r.f. saturation to the residues within the binding pocket.

The applications of this method in screening compound libraries [6,7], and in the qualitative identification of apparent group epitopes of the bound ligands to proteins [12–14], nucleic acids [11] and to virus coat proteins [15] have been reported. However, the potential of the STD-NMR experiment as a tool in quantitatively refining the bound-ligand conformation in weak ligand–protein complexes has not been addressed before. Such knowledge is essential in structure-based drug design. Often times crystallographic structures and models generated by computational docking programs are used to generate bound-ligand structures within the binding pocket. However, the crystallographic structure may not be preserved in solution in its entirety (which affects the STDs significantly). Similarly, the docking programs can only generate approximate structures. Thus, it is essential to have a refinement tool that can take these structures and refine them quantitatively to generate models that are compatible with experimental STD-NMR data. The present work demonstrates an approach by which such refinement can be obtained using STD-NMR data.

In this paper, we present a STD-NMR Intensity-restrained CORCEMA Optimization (SICO) procedure that utilizes a hybrid structure refinement protocol involving a complete relaxation and conformational exchange matrix (CORCEMA) calculation [9,10,16] and simulated annealing to refine the bound-conformation of a weakly binding ligand positioned within the binding pocket of a target protein. This method requires a knowledge of the pdb coordinates for the bound and free forms of the protein. This procedure is not designed to dock or orient a ligand within the binding pocket, or optimize the protein backbone conformation; rather it is meant to find the global minimum conformation of the ligand giving the lowest NOE R-factor at a given docked position. In our approach, the global minimum for the bound-ligand conformation is obtained by a refinement of the torsion angles of an approximate (or starting) structure for the bound ligand using STD-NMR intensities as experimental constraints, and the NOE R-factor [17,18] as the energy function to be minimized. The approximate structure for the complex may be obtained either from crystallographic data on the complex of interest, or generated from the crystallographic/NMR structure for a closely related complex, or from computer-docking of a ligand in the protein's binding pocket, or NMR based methods for docking [19,20], or other approaches [21]. It also allows the optimization of protein residue side-chains in the binding pocket.

2. Methods

In the STD-NMR experiment, one generally works with a solution containing an excess of a single ligand or a library of compounds in the presence of a small amount of a target protein (typically with ligand/receptor ratios ranging from 15:1 to 1000:1). The NMR signal(s) of the target molecule is saturated by r.f. irradiation without directly affecting the NMR signals from the low molecular weight ligand. Under weak binding conditions, the saturation spreads to the remaining protein protons as well as to the ligand protons in the free and bound states through a network of dipolar and conformational exchange processes as shown in Scheme II in [9]. The STD-NMR spectra can be obtained either by recording 1D-NMR spectra of the ligand with (I) and without (I_0) saturation of the protein protons and taking a difference ($I - I_0$), or by collecting free-induction decays (FID) in an interleaved manner to directly obtain the difference FID which is then Fourier transformed to obtain the difference spectrum ($I - I_0$).

The expression for the observable magnetization in a STD experiment by assuming infinite delay between each scan is given by [9]

$$\mathbf{I}(t) = \mathbf{I_0} + [1 - \exp\{-\mathbf{D}t\}]\mathbf{D}^{-1}\mathbf{Q}, \quad (1)$$

where $\mathbf{I}(t)$ is a column matrix containing the magnetizations for the ligand and for those protein protons that do not experience a direct r.f. saturation, in their free and bound states. $\mathbf{I_0}$ is the corresponding thermal equilibrium magnetization matrix. $\mathbf{Q}$ is a column matrix

containing cross-relaxation terms between the protein protons that experience a direct r.f. saturation and the rest of the protein protons and the bound ligand protons. The dynamic matrix **D** is a square matrix and is the sum of the relaxation rate matrix **R** and a kinetic matrix **K** of reduced dimensions. These matrices have been defined earlier [9]. 't' is the time period for which the protein proton(s) experience r.f. irradiation. The program we have written in MatLab also has a provision for taking into account the effect of finite delays (t_d) between scans in calculating the STD effects [9]. The flowchart for the CORCEMA-ST protocol (ST: saturation transfer) with the implementation of simulated annealing (SA) [22–24] refinement for torsion angles and/or experimental parameters such as the bound ligand correlation time is shown in Fig. 1.

To unequivocally demonstrate the ability of our procedure in successfully identifying the global minimum conformation for the bound ligand, the most convincing way to demonstrate this method is using a STD data set where the global minimum (i.e., target) structure is known a priori so that one can objectively compare the results from CORCEMA-ST optimization of various starting structures with the target structure. The use of experimental STD data on a complex and its known crystallographic structure as the target structure will not meet this requirement since if there are differences in the CORCEMA-ST optimized structure and the crystallographic structure, it will be difficult to say if this is due to the crystallographic structures not being preserved in solution in their entirety (unpublished experimental work, 2003), or if it is due to a shortcoming of our methodology. Thus, the best way to accomplish this objective is by utilizing a known crystallographic structure for the ligand–protein complex representing the target structure (global minimum), generate simulated experimental STD-NMR data for this target structure under typical experimental conditions (including noise in the 1D-NMR spectra represented by the elements of I_0), and show that the CORCEMA-ST procedure is able to *consistently and successfully* optimize any arbitrary starting conformation for the bound ligand to its known global minimum conformation. For this purpose, we generated a hypothetical ligand–protein complex consisting of a trisaccharide (Gal(1,4)-[Fuc(1,3)]GlcNAc, Fig. 2A) using the PDB coordinates (ID# 2kmb) from the crystal structure of the complex sLe^x/MBP (sialyl Lewis^x tetrasaccharide/mannose binding protein) [25]. This complex is shown in Fig. 3 of [9]. We retained eight residues (D184, E185, N187, H189, G190, N205, D206, and I207) in the MBP binding pocket, all within a 10 Å distance from the ligand protons. Since STD-NMR measurements are usually performed in D_2O, we have excluded all exchangeable hydrogens (OH and NH) in our calculations. For simulating the STD-NMR data, we have assumed that the ring CH protons of His189 were instantaneously saturated, and the intensities were

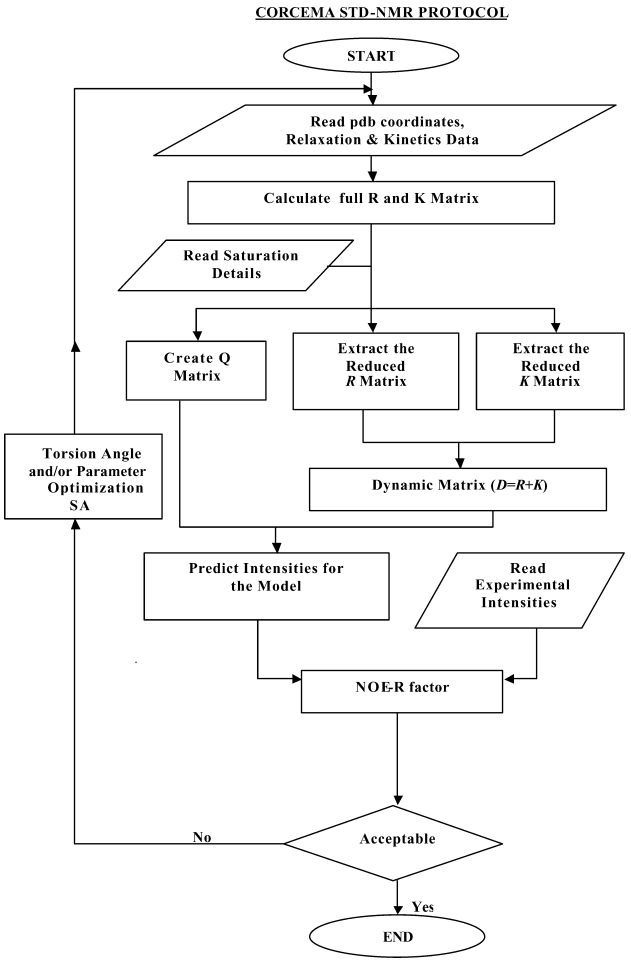

Fig. 1. CORCEMA–ST protocol for optimizing the bound-ligand conformation.

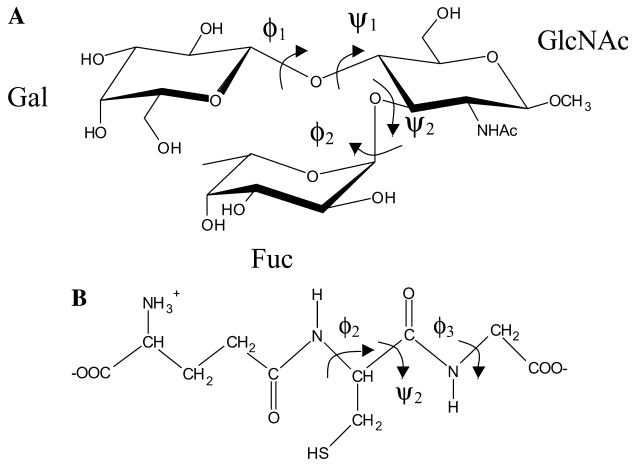

Fig. 2. (A) Structure of the trisaccharide: Gal(1,4)-[Fuc(1,3)]GlcNAc. (B) Structure of the tripeptide: γ-glutamyl-cysteinyl-glycine.

computed using Eq. (1) for a saturation time of 2 s, which is typical in many experiments. The carbohydrate ligand concentration was fixed at 0.5 mM, and calculations were performed at two separate ligand/protein ratios of 12.5/1 (protein ~40 μM) and 300/1 (protein ~1.67 μM). A uniform leakage relaxation of $0.2\,s^{-1}$ was assumed for all the protons in their free and bound states, to mimic non-specific leakage relaxation with paramagnetic oxygen in the solution [16]. The *simulated experimental* intensities ($S_{exp,k}$) were first calculated as percentage fractional intensity changes ($[(\mathbf{I}_{0k} - \mathbf{I}(t)_k) \times 100]/\mathbf{I}_{0k}$), where k is a particular proton in the complex, and $\mathbf{I}_{0k}$ its thermal equilibrium value) from the intensity matrix $\mathbf{I}(t)$ and using the target structure of the complex for the trisaccharide. Next, to demonstrate the ability of CORCEMA-ST program to successfully optimize the bound-ligand torsion angles to identify global minimum conformation (i.e., target conformation), we generated five different starting conformations for the bound ligand by *arbitrarily* changing the glycosidic torsion angles by significant amounts from those of the target structure.

To illustrate that the method is not limited to carbohydrate ligands, we have also performed a typical calculation using a peptide ligand, viz., a complex consisting of tripeptide (γ-glutamyl-cysteinyl-glycine, Fig. 2B) bound to glutathione transferase [26]. The pdb coordinates (ID# 2GSR) for the peptide complex were utilized. The residues within the binding pocket included in the calculation (within a 10 Å distance from the ligand) were: Chain A: Y7, F8, G12, R13, W38, K42, C45, F47, R48, Q49, L50, P51, Y61, Q62, S63, N64, A65, I66, Q95, R98, and C99. Chain B: D92, G93, E95, and D96. For simulating STD data, the methyl groups of alanine, leucine, and isoleucine were assumed to be saturated together. Other details such as the saturation time, spectrometer frequency were the same as for the carbohydrate ligand. The simulated experimental STD-NMR data without any random noise were calculated for different protons in tripeptide from the crystal structure coordinates [26] using the CORCEMA-ST program for a ligand/protein ratio of 12.5:1. A starting structure for the bound peptide was generated by changing the backbone torsion angles for the crystallographic structure by significant amounts (with $\phi 2$, $\psi 2$, and $\phi 3$ torsion angles set to −45°, −45°, and −45°) and was subjected to CORCEMA-ST optimization. We optimized only these backbone torsion angles $\phi 2$, $\psi 2$, and $\phi 3$, shown in Fig. 2B. Other parameters are the same as for the trisaccharide.

For each starting structure and all subsequent intermediate structures at each cycle of SA refinement, the STD intensities of different protons in the ligand were calculated using Eq. (1) and compared with the *simulated experimental* STD intensities. This comparison utilized an NOE R-Factor [17,18].

$$\text{NOE R-Factor} = \sqrt{\frac{\sum W_k (S_{exp,k} - S_{cal,k})^2}{\sum W_k (S_{exp,k})^2}}. \quad (2)$$

In these equations $S_{exp,k}$ and $S_{cal,k}$ refer to simulated experimental and calculated STD values for proton k. The use of a weighting (W_k) proportional to $1/S_{exp,k}$ for each individual STD intensity has the effect of making the R-factor sensitive to significant deviations in small STD values as well as to deviations in the large STD values. The NOE R-Factor (Eq. (2)) is taken as the pseudo-energy function to be minimized in a combined CORCEMA/SA refinement. Since occasionally there could be steric clashes during optimization, the CORCEMA-ST program has a provision to add, if necessary, an empirical van der Waals repulsion term [27] to Eq. (2) to minimize heavy atom conflicts during the refinement (results not shown, 2003):

$$E_{repel} = \begin{cases} 0 & \text{if } r \geqslant s r_{min}, \\ k_{vdw}(s^2 r_{min}^2 - r^2)^2 & \text{if } r < s r_{min}, \end{cases} \quad (3)$$

where the values of r_{min} is the sum of the van der Waals radii, the scale factor 's' is set to 0.8, and the force constant k_{vdw} is set to 0.1. For simplicity we reduced the number of atom types, since some of the atoms have very similar van der Waals radii [28]. The value of van der Waals radii is assumed as 0.8 Å for hydrogen and 1.5 Å for heavy atoms. In the calculations described in the current work, it was not necessary to include this term. The program also has a provision to automatically reject any conformation where the intramolecular distance between two protons is closer than 1.75 Å.

The SA refinement we used is based on the version by Alotto et al. [24]. It utilizes the standard Metropolis criterion for accepting or rejecting incremental random changes in torsion angles, one at a time. A random move is made in the "ith" dihedral angle "X_i" according to $X_{i-new} = X_{i-old} + (R \times v_i)$ where R is a random number between zero and one and v_i is the random walk step size. The v_i value is determined by a scaling factor, number of accepted moves for the ith variable and the number of conformations. The cooling procedure used here is the constant temperature reduction by a factor (α) between zero and one. The new temperature $T_{new} = T_{old} \times \alpha$. The closer the value of 'α' is to one the better is the chance of locating the global minimum. The initial annealing temperature T_0 is best determined by trial and error. In our calculation T_0 is set at 1 and α is set at 0.89. Rotations across specified bonds were performed using Eulerian transformation matrices. The resulting structures from CORCEMA-ST optimization are those that are compatible with the experimental STD values. If necessary, these structures can be subjected to additional energy minimization separately (outside CORCEMA-ST) using standard force field parameters in programs such as AMBER or CHARMM.

In the ideal case of a successful global energy minimization using STDs without any experimental errors, the NOE R-Factor is zero. In practice however the STD intensities can have significant random errors depending upon the S/N of the 1D-NMR spectra with (I) and without (I_0) r.f. irradiation from which the difference NMR spectrum $(I - I_0)$ is obtained. Thus, in practice the NOE R-factor will always be greater than zero even when the global minimum is achieved during optimization. To test the ability of the CORCEMA-ST program in optimizing structures using more realistic STD data, we generated simulated data sets including random noise for the trisaccharide example. For this, we first generated **I** and $\mathbf{I}_0$ matrices for the crystallographic structure of the target structure, and then added 1%, 2.5 or 5% of the total ligand concentration (0.5 mM) multiplied by a random number between −1 and +1 to the elements representing ligand proton intensities in the column matrices **I** and $\mathbf{I}_0$, separately. Using this procedure, random noise levels of 1, 2.5, and 5% in the 1D-NMR spectrum correspond respectively to S/N ratios of 125:1, 50:1, and 25:1 in the 1D-NMR spectra (i.e., $\mathbf{I_0}$ elements representing the ligand).

We have obtained good convergence for the bound ligand structure using any arbitrary conformation for the free ligand (data not shown). This is not surprising since the STDs are significantly more sensitive to variations in the bound ligand conformation than to variations in the free ligand conformation. Further, for small saturation times, the STDs are essentially independent of the free ligand conformation. For the results shown here, the free ligand conformation for the carbohydrate is maintained to be that in the crystal structure. For the peptide ligand, the free conformation was fixed to be an all-trans structure. We also assumed that the protein conformation is known (e.g., from crystallography or NMR) and is fixed in both free and bound states. Note that the side chain orientations of the protein residues within the binding pocket could also be optimized using our procedure. In particular this may be necessary if the optimization of ligand structure alone doesn't yield acceptable R-factors inspite of using high-quality STD data sets. Thus in the current work where we want to demonstrate our procedure, the only torsion angles to be optimized correspond to those of the bound ligand (an experimental example where we had to optimize one of the protein residue sidechain orientation will be reported elsewhere). The starting conformation for the ligand–protein complex can be either an existing model from the protein data bank (PDB), or a model generated by a computational docking algorithms, or any hypothetical model generated from structures of related complexes. The PDB coordinates of the free ligand, free protein and protein–ligand complex are read-in parameters. The remaining parameters such as correlation times and binding constants etc were also fixed as described below.

In order to obtain meaningful optimization, it is preferable for the number of parameters to be optimized to be less than the number of experimental STDs. In the current set of calculations, in addition to the four ligand torsion angles, four additional parameters are needed: the dissociation constant (K_d), the correlation times for the free ligand (τ_L), and the protein (τ_P), and the order parameter S^2 for methyl group-external proton interactions [9,29,30]. Since the STDs are independent of K_d in the range 10^{-4}–10^{-7} M for high ligand/protein ratios [9], we chose arbitrarily a typical value of 10^{-6} M for the weak complex to compute the exchange matrices using a diffusion limited on rate of $10^8\ s^{-1}\ M^{-1}$. The NOE R-factor is not very sensitive to variations in S^2 in the range 0.6–0.9 (changing only by about 0.05 in magnitude, results not shown); thus a reasonable value of 0.85 was assumed for S^2. The τ_L can always be experimentally determined; here we chose a typical value of 0.5 ns for the trisaccharide. Thus, the only remaining parameter needed is τ_P. This correlation time can be determined experimentally from independent methods (e.g., from fluorescence decay). Alternatively, this could be determined by simultaneous optimization together with the torsion angles. For the current calculation τ_p was fixed at 40 ns. A spectrometer frequency of 600 MHz was assumed. For the carbohydrate ligand, the sampling range for the interglycosidic torsion angles ϕ: H1–C1–O–Cn (−60° to +120°) and for ψ: C1–O–Cn–Hn (−60° to +60°) where Cn and Hn refer to the aglycone carbon and its attached proton (Fig. 2A), were restricted to the allowed regions as per the Ramachandran type population contour map of the glycosidic linkages [31] in sLe^x. Each ϕ (across C1–O) and ψ (across O–Cn) optimization was done in both directions of the pertinent bonds as the intermolecular dipolar interactions and hence the associated STDs are different for each choice. For the peptide ligand, the sampling range for the torsion angles is from −180° to +180°.

3. Results and discussions

Fig. 3A shows a comparison of the five starting structures for the trisaccharide and the target structure. Fig. 3B shows the results of optimization of the five starting structures using simulated STD data sets (with 23 intensities) without any noise. They are shown within the protein binding pocket which is included in the calculation. The torsion angles and the NOE R-factors for this case before and after CORCEMA-ST optimization are summarized in Table 1. It is seen that CORCEMA-ST is successfully able to optimize all the five different structures to the same global minimum structure (the target structure). We obtained equally

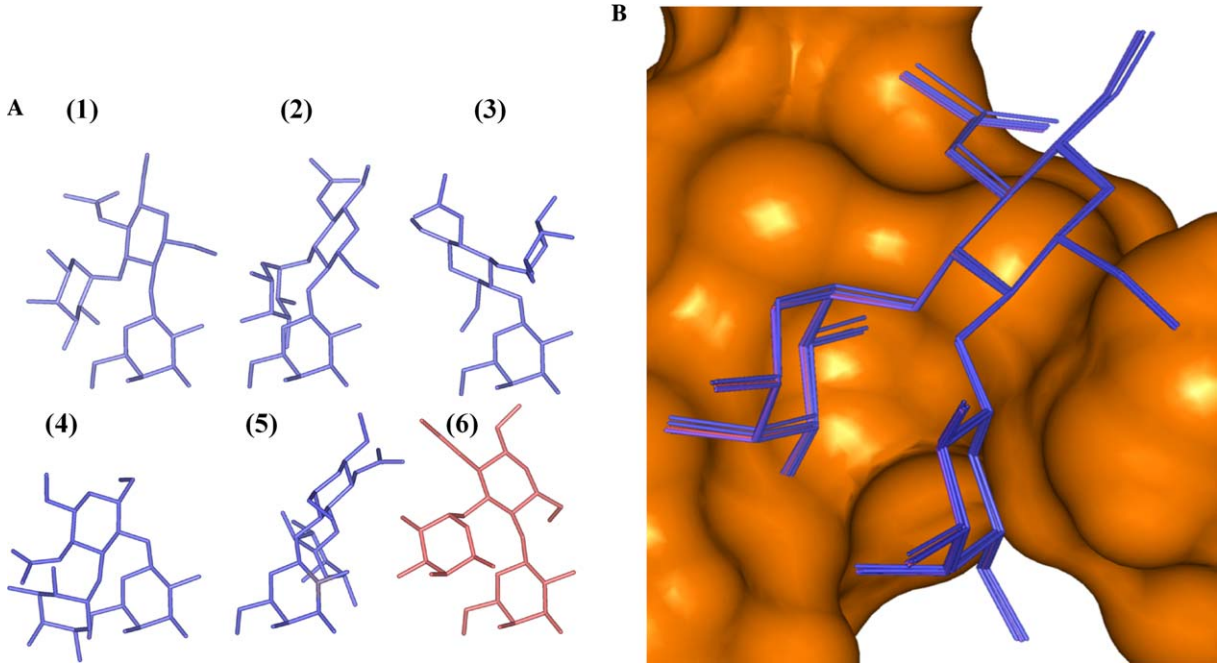

Fig. 3. (A) Comparison of the five arbitrary starting structures (1–5) and the target structure (6) for the branched trisaccharide Gal(1,4)-[Fuc(1,3)]GlcNAc. The hydrogens were omitted for clarity. For the benefit of the reader in assessing the wide variations in the torsion angles of the different starting structures, all structures are shown in the figure with the Gal residue at the bottom. However, none of the residues are fixed within the binding pocket (see the torsion angles for the starting structures in Table 1). (B) Best-fit superposition of the structures resulting from the CORCEMA-ST optimization of the five starting structures in (A) with the target structure (ligand/protein ratio of 12.5/1). Note that the target structure is shown at a slightly different orientation than the one in (A). The noise level was assumed to be zero. The smooth surface represents the protein residues within the binding pocket. Tables 2 and 3 shows the results of successful optimization when the 1D-NMR spectra (i.e., $\mathbf{I_0}$) are recorded with a finite S/N value prior to calculating STDs.

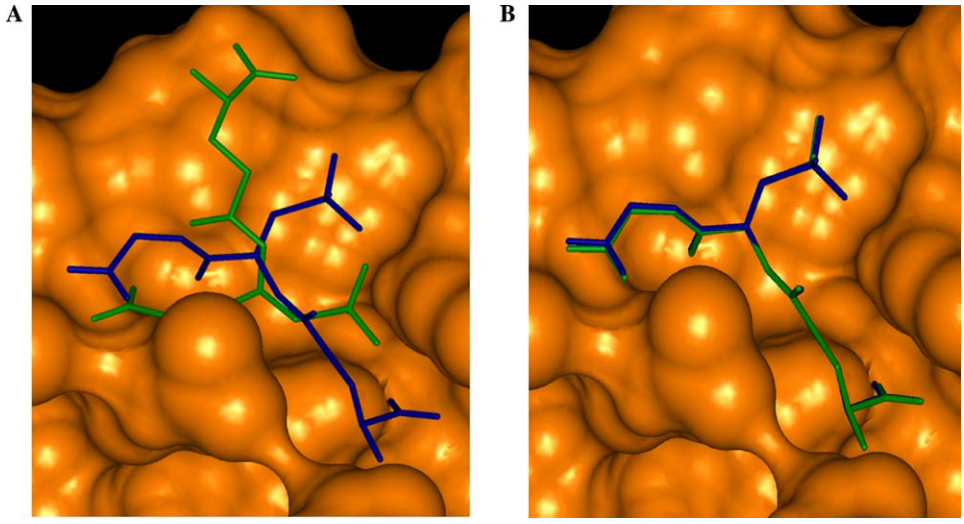

Fig. 4. (A) Comparison of the arbitrary starting structure (green) and the target structure (blue) for the tripeptide (γ-glutamyl-cysteinyl-glycine). The hydrogens were omitted for clarity. The smooth surface represents the protein residues within the binding pocket. (B) Best-fit superposition of the structure resulting from the CORCEMA-ST optimization of the starting structure in (A) with the target structure (ligand/protein ratio of 12.5/1). The noise level was assumed to be zero. The smooth surface represents the protein residues within the binding pocket.

good convergence when we reduced the number of STDs included in the calculation from 23 to 11. In Table 2, we have summarized the results of systematically varying the S/N ratio in the 1D-NMR spectra (*Note.* S/N here refers to the normal 1D-NMR spectrum of the ligand, i.e., the elements of $\mathbf{I_0}$ and *not* the STD-spectrum which is the difference $\mathbf{I} - \mathbf{I_0}$) on the ability of the CORCEMA-ST program in locating the global

Table 1
Results of optimization[⊥] of different starting structures for the trisaccharide (ligand/protein = 12. 5/1)

Ligand conformation	$\phi1$	$\psi1$	$\phi2$	$\psi2$	R-Factor
Target (crystal) structure	44.32	14.88	49.77	20.69	0.00
Starting structure 1	14.28	−35.11	99.74	50.70	2.10
Optimized structure 1	45.13	14.91	48.57	19.77	0.02
Starting structure 2	−25.67	−20.20	109.81	−14.45	2.19
Optimized structure 2	43.22	14.57	47.39	22.74	0.02
Starting structure 3	107.11	58.23	−54.32	−46.09	0.82
Optimized structure 3	43.40	14.68	49.27	20.72	0.02
Starting structure 4	114.23	−55.16	−49.82	55.73	0.70
Optimized structure 4	44.09	14.55	49.79	20.18	0.02
Starting structure 5	−45.72	−55.04	−45.26	−49.16	6.62
Optimized structure 5	40.95	17.39	48.49	20.25	0.02

[⊥] without noise.

minimum at a ligand/protein ratio of 12.5/1. Similar calculations at a ligand/protein ratio of 300/1 were also performed (Table 3). It is seen that at 12.5/1 ratio, good convergence is obtained even when the S/N ratio is only 50/1 for the 1D-NMR spectrum. At 300/1 ratio the reasonable convergence is obtained if the 1D-NMR spectra (I_0) are recorded at S/N ratio of 1250/1 or better, which is very routinely accomplished these days from the use of perdeuterated solvents, signal averaging and high magnetic fields. As the number of optimizable parameters increases (e.g., the number of torsion angles), a need for recording 1D-NMR spectra with better S/N ratios is necessary to avoid getting trapped in local minima. Fig. 4 shows the results of successful optimization for the hypothetical peptide/protein complex. Fig. 4A shows a comparison of the starting structure for the bound ligand and the target structure, both positioned within the binding pocket. Fig. 4B shows the superposition after CORCEMA-ST optimization.

3.1. Calculations involving proteins without assignments

To undertake these calculations, it is necessary to set up the **Q**-matrix which requires a knowledge of the assignment of the protein resonance(s) being saturated. Thus, in those instances where the proteins have already been extensively characterized and the NMR assignments completed, this method readily enables one to refine the bound structures of weakly binding ligands using STD data. In practice, however, many STD-NMR measurements involve novel target proteins for which the assignments are not yet available for a variety of reasons. Such unassigned proteins may potentially pose a challenge for undertaking these calculations. This topic could be the subject of a separate study. For such proteins, we suggest the following practical approaches in generating usable STD data for undertaking CORCEMA-ST calculations: (1) selective saturation (using shaped pulses) of all the aromatic ring CH reso-

Table 2
Results for the optimization of the same starting structure (#2 in Table 1) for the trisaccharide using simulated STD-NMR data obtained from 1D-NMR spectra with different S/N values (ligand/protein = 12. 5/1)

Ligand conformation	S/N[a] of 1D-NMR spectra	$\phi1$	$\psi1$	$\phi2$	$\psi2$	R-Factor
Target structure	—	44.32	14.88	49.77	20.69	0.00
Initial structure 2	—	−25.67	−20.20	109.81	−14.45	2.19
Optimized structure 2	High	43.22	14.57	47.39	22.74	0.02
Optimized structure 2A	125:1	45.85	14.37	54.77	18.98	0.05
Optimized structure 2B	50:1	40.34	17.43	49.08	21.72	0.07
Optimized structure 2C	25:1	58.65	6.99	29.98	31.02	0.18

[a] *Note*. S/N here refers to the normal 1D-NMR spectrum of the ligand, i.e., the elements of $\mathbf{I_0}$, and *not* the STD-spectrum it self which is the difference $\mathbf{I} - \mathbf{I_0}$. The corresponding STD spectra ($I - I_0$) will have substantially lower S/N. See the text for details on adding noise to elements of **I** and $\mathbf{I_0}$.

Table 3
Results for the optimization of the same starting structure (#2 in Table 1) for the trisaccharide using simulated STD-NMR data with noise[a] (ligand/protein = 300/1)

Ligand conformation	S/N[a] of 1D-NMR spectra	$\phi1$	$\psi1$	$\phi2$	$\psi2$	R-Factor
Target structure	High	44.32	14.88	49.77	20.69	0.00
Initial structure 2	High	−25.67	−20.20	109.81	−14.45	5.94
Optimized structure 2	High	44.87	15.05	49.32	21.18	0.03
Optimized structure 2A	1250:1	43.55	15.97	48.44	22.50	0.17
Optimized structure 2B	500:1	20.99	35.48	113.17	−1.37	0.43

[a] *Note*. S/N here refers to the normal 1D-NMR spectrum of the ligand, i.e., the elements of $\mathbf{I_0}$, and *not* the STD-spectrum it self which is the difference $\mathbf{I} - \mathbf{I_0}$. The corresponding STD spectra ($I - I_0$) will have substantially lower S/N. See text for details on adding noise to elements of **I** and $\mathbf{I_0}$.

nances together since they show up generally in a short window of 6.5–8.5 ppm. The **Q**-matrix is easily set up for this case. (2) Selective saturation of well-resolved ring CH protons of an easily identifiable residue such as Tyr or His. If this residue happens to be unique (i.e., appearing only once in the sequence), its assignment is known. If there is more than one Tyr or one His in the sequence, it may be still possible to infer its assignment because if these residues are part of the protein binding pocket they tend to produce larger STDs (due to direct effects, $E2' \rightarrow L'$) at low to moderate ligand/protein ratios (15:1–50:1) whereas if they are far away (more than 7 Å) from the ligand, the STDs are much smaller (9). Further, a misassignment of the residue will not yield low R-factors even after extensive optimization. *These two approaches involving the saturation of all or some of the aromatic CH resonances should work very well for ligands that do not have signals in the aromatic resonance range (e.g., carbohydrate ligands)*. (3) Similarly, the methyl resonances which appear as a group (generally between −1 and 1.5 ppm) can be saturated together and incorporated in the **Q**-matrix calculation. If some non-methyl proton resonances (e.g., Lys γCH_2) happen to overlap with this methyl group, errors introduced by their saturation are small provided they are well outside a 7 Å distance from the ligand in the binding pocket. If these non-methyl resonances arise from residues within the pocket or in its vicinity, there will be errors, and large R-factors may be expected even after optimization. (4) On the other hand, if the selective saturation of a well-resolved methyl resonance produces large STDs at moderate ligand/protein ratios, it may be possible to infer its assignment by looking for methyl-group containing residues in the binding pocket. If there are two or more such methyl-group residues either within the protein binding pocket or within a 7 Å cut-off from the ligand, it may be possible to infer the assignments indirectly from the magnitudes of optimized NOE R-factors (i.e., correct assignments will yield low optimized R-factors whereas wrong assignments will result in large R-factors even after extensive optimization). For example, the simulated STD data for the peptide ligand in this work were generated by assuming a saturation of methyls of Leu, Ala, and Ile residues. Using this data, if the analysis was done assuming saturation of Ala65 methyl group only (which is far away from the ligand and hence produces only small STDs), the initial R-factor is 0.94 and the optimized R-factor is 0.74, a too high a value. Thus, when large STDs are observed, inclusion of residues within the binding pocket together with similar residues far away (>7 Å) doesn't introduce much error. When only small STDs are observed even at moderately low ligand/protein ratios, it may mean that residues located far away from the ligand are being saturated, and hence they may be difficult to assign when multiple residues are present. (5) Finally, a saturation-transfer screening (using short saturation times) of the protein resonance envelope to identify irradiation frequencies that yield large STDs is likely to aid in the identification of residues within the binding pocket, and in assignments for the calculations.

4. Conclusions

In conclusion, we have described a STD-NMR intensity-restrained CORCEMA optimization (SICO) procedure which permits the refinement of the bound-conformation of a ligand. This approach will be meaningful only if the ligand is docked in an optimal or near optimal manner since improper docking or a wrong orientation of the ligand may result in large R-factors even after optimization. Thus, a judicious use of crystallographic/NMR structures and/or docking algorithms is initially required to position the ligand within the binding pocket before the SICO refinement the bound ligand conformation using experimental STD data.

This method is primarily meant for weak complexes (with K_d in the range 10^{-3}–10^{-7} M), and utilizes STD-NMR intensities as experimental constraints even though the CORCEMA theory it self is applicable over a wide range of binding conditions, from weak to tight complexes (9,16). The CORCEMA-ST method belongs to the general class of intensity-restrained optimization procedures described in literature previously (e.g., see [17,18,23,32–36]) except it explicitly incorporates an exchange matrix also to treat reversibly forming weak complexes. The success of our procedure is in large part due to the fact that the ligand STDs are sensitive to the conformation of the bound-ligand itself due to the variations in the intermolecular dipolar relaxation with the protein [9,10]. These CORCEMA calculations can be further augmented by the incorporation of additional constraints for the bound ligand such as intra-ligand distance constraints (from transferred NOESY), torsion angle constraints, and intra-ligand relaxation rates [37]. Even though here for simplicity we fixed the orientations of the protein residue side chains, our procedure permits an optimization of their orientations also (unpublished work, 2003). Our method can be utilized to simulate some test examples to identify optimal experimental conditions (in particular ligand/protein ratios to be employed and the S/N ratios required for 1D-NMR spectra) in the design of STD-NMR experiments for a quantitative refinement of the bound-ligand conformation. The CORCEMA-ST procedure significantly extends the utility of the STD-NMR technique beyond its current applications such as compound library screening and qualitative mapping of apparent group epitopes. Since many lead compounds bind target proteins weakly with affinities in the mM to μM range [1], the CORCEMA-ST procedure is likely to be of value in

determining the bound conformations of these lead compounds positioned within the binding pocket of the target protein. Such knowledge is essential in structure-based design of novel drugs.

Acknowledgments

The authors would like to acknowledge partial support of this work by the NCI Grants CA-84177 and CA-13148.

References

[1] D.F. Wyss, M.A. McCoy, M.M. Senior, NMR-based approaches for lead discovery, Curr. Opin. Drug Discov. Dev. 5 (2002) 630–647.
[2] T. Peters, T. Biet, L. Herfurth, NMR Methods for screening the binding of ligands to proteins—identification and characterization of bioactive ligands, Biol. Magn. Reson. 20 (2003) 287–315.
[3] B. Meyer, T. Peters, NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors, Angew. Chem. Int. Ed. 42 (2003) 864–890.
[4] A.H. Siriwardena, F. Tian, S. Noble, J.H. Prestegard, A straightforward NMR-spectroscopy-based method for rapid library screening, Angew. Chem. Int. Ed. Engl. 41 (2002) 3454–3457.
[5] W. Jahnke, P. Floersheim, C. Ostermeier, X. Zhang, R. Hemming, K. Hurth, D.P. Uzunov, NMR Reporter screening for the detection of high-affinity ligands, Angew. Chem. Int. Ed. Engl. 41 (2002) 3420–3423.
[6] M. Mayer, B. Meyer, Characterization of ligand binding by saturation transfer difference NMR spectroscopy, Angew. Chem. Int. Ed. Engl. 35 (1999) 1784–1788.
[7] M. Vogtherr, T. Peters, Applications of NMR based binding assays to identify key hydroxy groups for intermolecular recognition, J. Am. Chem. Soc. 122 (2000) 6093–6099.
[8] T. Haselhorst, T. Weimar, T. Peters, Molecular recognition of sialyl lewis^x and related saccharides by two lectins, J. Am. Chem. Soc. 123 (2001) 10705–10714.
[9] V. Jayalakshmi, N.R. Krishna, Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand–receptor complexes, J. Magn. Reson. 155 (2002) 106–118.
[10] N.R. Krishna, V. Jayalakshmi, Complete relaxation and conformational exchange matrix (CORCEMA) analysis of saturation transfer difference (STD) NMR spectra of ligand–protein complexes, J. Korean Magn. Reson. Soc. 6 (2002) 94–102.
[11] M. Mayer, T.L. James, Detecting ligand binding to a small RNA target via saturation transfer difference NMR experiments in D_2O and H_2O, J. Am. Chem. Soc. 124 (2002) 13376–13377.
[12] M.A. Johnson, B.M. Pinto, Saturation transfer difference 1D TOCSY experiments to map the topography of oligosaccharides recognized by a monoclonal antibody directed against the cell-wall polysaccharide of group A *Streptococcus*, J. Am. Chem. Soc. 124 (2002) 15368–15374.
[13] H. Maaheimo, P. Kosma, L. Brade, H. Brade, T. Peters, Mapping the binding of synthetic disaccharides representing epitopes of chlamydial lipopolysaccharide to antibodies with NMR, Biochemistry 39 (2000) 12778–12788.
[14] M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with the protein receptor, J. Am. Chem. Soc. 123 (2001) 6108–6117.
[15] A.J. Benie, R. Moser, E. Bauml, D. Blass, T. Peters, Virus–ligand interactions: identification and characterization of ligand binding by NMR spectroscopy, J. Am. Chem. Soc. 125 (2003) 14–15.
[16] (a) H.N.B. Moseley, E.V. Curto, N.R. Krishna, Complete relaxation and conformational exchange matrix (CORCEMA) analysis of NOESY spectra of interacting systems; two-dimensional transferred NOESY, J. Magn. Reson. Ser. B 108 (1995) 243–261;
(b) N.R. Krishna, H.N.B. Moseley, Complete relaxation and conformational exchange matrix (CORCEMA) analysis of NOESY spectra of reversibly forming ligand–receptor complexes, in: Biological Magnetic Resonance, vol. 17, Structure Computation and Dynamics in Protein NMR, Kluwer Academic/Plenum Press, New York, 1999, pp. 223–310.
[17] N.R. Krishna, D.G. Agresti, J.D. Glickson, R. Walter, Solution conformation of peptides by the intramolecular nuclear overhauser effect experiment. Study of valinomycin-K^+, Biophys. J. 24 (1978) 791–814.
[18] Y. Xu, I.P. Sugar, N.R. Krishna, A variable target intensity-restrained global optimization (VARTIGO) procedure for determining three-dimensional structures of polypeptide from NOESY data: application to gramicidin-S, J. Biomol. NMR 5 (1995) 37–48.
[19] A.P. Zabell, C.B. Post, Docking multiple conformations of a flexible ligand in to a protein binding site using NMR restraints, Proteins 46 (2002) 295–307.
[20] R.P. Meadows, P.J. Hajduk, A genetic algorithm-based protocol for docking ensembles of small ligand using experimental restraints, J. Biomol. NMR 5 (1995) 41–47.
[21] A. Fahmy, G. Wagner, Treedock: a tool for protein docking based on minimizing van der Waals energies, J. Am. Chem. Soc. 124 (2002) 1241–1250.
[22] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[23] Y. Xu, N.R. Krishna, Structure determination from NOESY intensities using metropolis simulated annealing (MSA) refinement of dihedral angles, J. Magn. Reson. Ser. B 108 (1995) 192–196.
[24] P.G. Alotto, C. Eranda, B. Brandstatter, G. Furntratt, C. Magele, G. Molinari, M. Nervi, K. Preis, M. Repetto, K.R. Richter, Stochastic algorithms in electro magnetic optimization, IEEE Trans. Magn. 34 (1998) 3674–3684.
[25] K.K. Ng, W.I. Weis, Structure of selectin like mutant of mannose-binding protein complexed with sialylated and sulfated lewis(x) oligosaccharides, Biochemistry 36 (1997) 979–988.
[26] H. Dirr, P. Reinemer, R. Huber, The crystal structure of class pi glutothione S-transferase from porcine lung (pGST P1-1) in complex with glutothione, J. Mol. Biol. 243 (1994) 72–92.
[27] P.C. Driscoll, A.M. Gronenborn, L. Beress, G.M. Clore, Determination of the three-dimensional solution structure of the antihypertensive and antiviral Protein BDS-I from the sea anemone *Anemonia sulcata*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing, Biochemistry 28 (1989) 2188–2198.
[28] A.J. Li, R. Nussinov, A set of Van der Waals and Coloumbic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking, Proteins: Structure, Function, Genetics 32 (1998) 111–127.
[29] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation macromolecules. 1. Theory and the range of validity, J. Am. Chem. Soc. 104 (1982) 4546–4559.
[30] M.J. Dellwo, A.J. Wand, The influence of methyl rotor dynamics on hydrogen relaxation networks: derivations of spectral densities in model free form, J. Am. Chem. Soc. 115 (1993) 1886–1893.

[31] M. Frank, A.B. Lang, T. Wetter, C.W. Lieth, Rapid generation of a representative ensemble of N-glycans conformations, In Silico Biol. 2 (2002) 38–51.
[32] B.A. Borgias, T.L. James, Two-dimensional nuclear overhauser effect: complete relaxation matrix analysis, Methods Enzymol. 176 (1989) 169–183.
[33] J.E. Mertz, P. Guntert, W. Wuthrich, W. Braun, Complete relaxation matrix refinement of NMR structures of proteins using analytically calculated dihedral angle derivatives of NOE intensities, J. Biomol. NMR 1 (1991) 257–269.
[34] A.M. Bonvin, R. Boelens, R. Kaptein, Direct NOE refinement of biomolecular structures using 2D NMR data, J. Biomol. NMR 1 (1991) 305–309.
[35] B.A. Borgias, T.L. James, COMATOSE, A method for constrained refinement of macromolecular structure based on two-dimensional nuclear overhauser effect spectra, J. Magn. Reson. 79 (1988) 493–512.
[36] H.N. Moseley, W. Lee, C.H. Arrowsmith, N.R. Krishna, Quantitative determination of conformational, dynamic and kinetic parameters of a ligand–protein/DNA complex from a complete relaxation and conformational exchange matrix analysis of intermolecular transferred NOESY, Biochemistry 36 (1997) 5293–5299.
[37] J. Yan, A.D. Kline, H. Mo, M.J. Shapiro, E.R. Zartler, The effect of relaxation on the epitope mapping by saturation transfer difference NMR, J. Magn. Reson. 163 (2003) 270–276.