Supporting Information

Refinement of the Conformation of UDP-Galactose Bound to Galactosyltransferase Using the STD NMR Intensity-Restrained CORCEMA Optimization

V. Jayalakshmi,[†] Thorsten Biet,[‡] Thomas Peters,[*,‡] and N. Rama Krishna[*,†]

(I) STD-NMR Measurements :

The experiments were performed on a Bruker Avance DRX-500 system with a 5 mm TXI probe using the procedures described before.[1] The measurements were performed at 293 K. The NMR sample was prepared in D_2O with the ligand/protein ratio of 55:1 (protein concentration = 14.5 μM). Saturation transfer was achieved by using 40 selective gaussian pulses (duration 50 ms, spacing 1 ms). The protein envelope was irradiated at δ=0 ppm (on-resonance) and δ=40 ppm (off-resonance). Saturation time was 2 seconds. The time delay between each experiment for signal averaging was 2.3 s. A T1ρ-filter was used to suppress broad protein signals in the difference spectrum. The STD signal was obtained by recording the free-induction decays in an interleaved manner, with 2K scans. For the CORCEMA analysis, only STD intensities from ligand signals without overlap were used (see Table 1 in the text).

(II) Details of STD-NMR Intensity-restrained CORCEMA Optimization:

The STD-NMR Intensity-restrained CORCEMA Optimization (SICO) protocol has been previously described.[2] The intensities of the NMR signals in the STD-NMR experiment are given by[3]:

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

where **I** is a column matrix of magnetizations for the protons in the ligand and the enzyme in their free and complexed states. **D** is the sum of the relaxation rate matrix **R** and the kinetic matrix **K.** **Q** is the cross-relaxation rate matrix for the saturated protons in the protein. The above equation is valid for infinitely long delays between experiments. The CORCEMA-ST program has a provision to take into account finite delays between experiments and this provision was used in the current work.

To minimize the dimensions of the matrices in CORCEMA calculations, only those protein residues within a 10 Å distance from the ligand were included in the calculations. Thus, in the current calculations twenty-six amino acid residues within the binding pocket (P187, F188, R189, N190, R191, Q192, H194, F226, R228, S251, D252, V253, D254, L255, K279, Y289, G292, G313, W314, G315, E317, D318, M344, H347, D350 and N353) were included. To speed up the calculations further, spectral densities were calculated for only those proton pairs having distance of 10 Å or less; for all others they were set to zero. Since STD-NMR measurements are usually performed in D_2O, we have excluded all exchangeable hydrogens (OH and NH) in our calculations. The STD intensities were calculated as percentage fractional intensity changes ($[(\mathbf{I}_{0k}-\mathbf{I}(t)_k)*100]/\mathbf{I}_{0k}$, where k is a particular proton in the complex, and $\mathbf{I}_{0k}$ its thermal equilibrium value) from the intensity matrix **I**(t) and compared to the experimental STD values using an NOE R-factor defined as,[4,5]

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

In these equations $S_{exp,k}$ and $S_{cal,k}$ refer to 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. We have also included an empirical van der Waals repulsion term[6] in the target function to minimize heavy atom conflicts during the refinement[2]:

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

where the value 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.[7] However, in the program one can introduce radii for specific atom types if necessary. The value of van der Waals radii is assumed as 0.8Å for hydrogen and 1.5Å for heavy atoms. The lower limit of the interproton distance restraints is set to 1.75Å.

The simulated annealing refinement, based on the version by Alotto *et al.*[8] was used for optimizing the torsion angles to get best fit between the experimental and predicted intensities. It utilizes the standard Metropolis criterion for accepting or rejecting incremental random changes in torsion angles, one at a time. The target function to be minimized is the sum of NOE R-factor (eq. (2)) and an empirical van der Waals repulsion term E_{repel} (eq. (3)). In our calculations, in addition to the 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.[9,10] We measured the dissociation constant value of UDP-Gal in complex with β4 GalT1 as 1.6μM at 298K. We have assumed the diffusion limited on rate as 10^8 $s^{-1}M^{-1}$. The NOE R-factor is not very sensitive to variations in S^2 in the range 0.6 to 0.9;[2] thus a reasonable value of 0.8 was assumed for S^2. For τ_L here we chose a typical value of 0.5 ns for the ligand based on its low molecular weight. Thus, the only remaining parameter needed is τ_P, this was determined by simultaneous optimization together with the torsion angles. 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. The methyl group internal correlation time (τ_m) was set at a reasonable value of 10 ps. For the free ligand, the torsion angles were taken from ref 11 (α = 60°, β = -180°, γ = 59° , ϕ = 74° , ψ = -90°, χ = -178°, μ = -179°, ν = 90°). A simple optimization of the protein correlation time (τ_P) for crystal structure and the energy minimized structure resulted in abnormally short values of 5 ns and 10 ns respectively. This result suggests the need for torsion angle refinement. During the optimization we fixed the torsion angles between the two phosphate groups since these groups are known to be coordinated with the metal ion in the crystal structure.[12] We optimized galactose (ϕ, ψ, and ω) and ribose (α, β and γ) torsion angles in the ligand (Figure S1). The galactose orientation is defined by the torsion angles ϕ (P2-O1"-C1"-O5"), ψ (OP4-P2-O1"-C1") and ω (O5"-C5"-C6"-O6"), and the ribose orientation is defined by α (Op4-P1-OP1-C5'), β (P1-OP1-C5'-C4') and γ (OP1- C5'-C4'-C3'), The numbering of the atoms is same as scheme 1 in ref.11. The sampling range for the interglycosidic angles is from –180° to +180° and for the protein correlation time is from 0.1 ns to 100 ns. Table 1 of the text shows a comparison of the experimental and calculated STDs for the various models. The discrepancy between experimental and SICO predicted STDs for the H5U may be the result of a systematic error from the overlap of H5U and H1R resonances. The UDP-Gal and V253 torsion angles are shown in Table S1 for the crystal structure, the energy-minimized crystal structure, and the CORCEMA-ST refined structure.

Since the protein signals at zero ppm were saturated in the STD experiment, for the computation of the Q-matrix in Eq(1) we made the assumption that the V253 and L255 methyl protons are instantaneously saturated together (they are the only methyl-group containing residues within a 10 Å distance from the ligand that are most likely to have methyl proton resonances near 0 ppm. The M344 CH_3 resonance is somewhat less likely to be near zero ppm range). The validity of this assumption (and assignment) was further tested by performing separate SICO optimizations in which either only L255 or V253 methyl groups were assumed to be saturated and CORCEMA-ST refinement of seven torsion angles and the protein correlation time (τ_P) were performed. These resulted in optimized R-factors greater than 0.46 (i.e., 0.467 with τ_P = 33.16 ns for V253 saturation and 0.48 with τ_P = 49.93ns

for L255 saturation), whereas a simultaneous saturation of both L255 and V253 methyl resonances resulted in a significantly lower optimized R-factor of 0.332 with τ_P = 24.12 ns, a value which is reasonable for the protein of MW of 38 kDa[13]. Fig S2 shows the stereo view of the energy-minimized and CORCEMA-ST structure. Even though the R-factor based assignment procedure has worked in this particular case, one has to use caution in using this approach for the general case.

This CORCEMA-ST method does not require an explicit knowledge of the bound-ligand conformation from transferred-NOE but relies only on STD intensities as constraints. However, the calculations can be augmented, if necessary, by the inclusion of intra-ligand distance constraints from tr-NOE, ligand-saturated protein residue distances (r) from the initial slopes of STD-intensity build up (which are proportional to $1/r^6$) [3], torsion angle and intra-ligand relaxation rates as additional constraints.

In the crystal structure β4 GalT1 upon substrate binding undergoes a significant conformational change, brought about by a loop movement of residues 345-365, and a change of the side chain orientation of Trp314. Only three residues in this structure (H347, D350 and N353) in the loop are within the 10 Å distance from the ligand. To see if the loop movement has any effect on the optimizations, we performed CORCEMA-ST calculations with and without these three loop residues to simulate loop movement. To see the effect of TRP314 sidechain orientation we performed CORCEMA-ST refinement for only TRP314 sidechain torsion angles (χ_1 and χ_2). Neither of these calculations (i.e., simulation of loop movement and change in the Trp314 sidechain orientation) produced any significant changes in STDs generated from a saturation of the methyls at 0 ppm (V253 and L255). This is not surprising because the STDs and CORCEMA-ST calculations are sensitive to only those conformational changes in the ligand or protein that affect significantly the saturation transfer from the saturated protein protons (E2') to the bound ligand (L') by either direct (E2'→L') or indirect (E2' →E1'→L') pathways[3]. That is not the case with the loop residues and Trp314 for saturation of the methyls at 0 ppm (V253 and L255). A separate set of STD measurements involving the selective saturation of all the aromatic ring CH proton resonances will be required to answer the question of loop movement and Trp314 sidechain reorientation upon substrate binding.

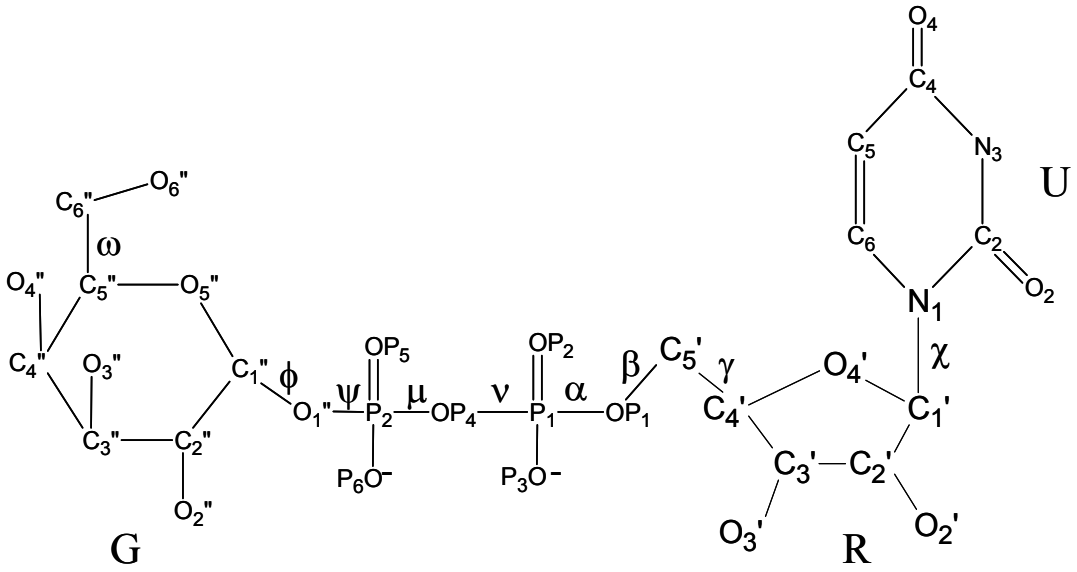

Figure S1. Structure of UDP-Gal with the labeling of the residues (G: Galactose; R: Ribose; U: Uracil) and torsion angles.

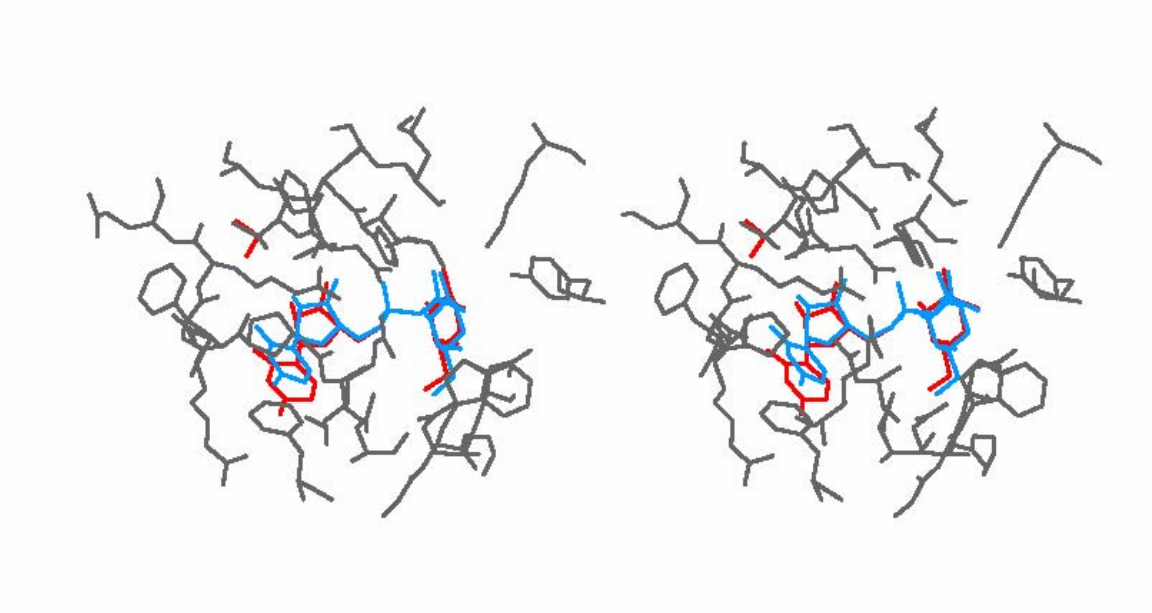

Figure S2. Stereo view of the best-fit superposition of CORCEMA-ST optimized structure (red colour) with the energy minimized starting structure (Protein residues with in the binding pocket: black colour and ligand residues: blue colour). The hydrogens were omitted for clarity.

Table S1: Comparision of torsion angles for different structures of UDP-Gal/β4 GalT1 complex

Torsion Angle	**Crystal Structure**	**Energy minimized Crystal structure**	**SICO Structure**
Valine (chi1)	70.75	73.54	177.99
Ribose (alpha)	-146.22	149.97	149.20
Ribose (beta)	-177.11	106.06	105.35
Ribose (gamma)	164.22	-77.94	-95.19
Galactose (phi)	73.55	75.84	83.60
Galactose (psi)	-133.04	-132.85	-131.88
Galactose (sidechain)	172.39	161.1	130.85

For the CORCEMA calculations, since the crystal structure (PDB entry 1O0R) of the UDP-Gal/β4 GalT1 complex showed a bad bond angle and a bad contact (see text), it was energy minimized by using the Discover module in InsightII (Molecular Simulations, Inc) for 6000 iterations using steepest descent algorithm and the force field CVFF. The residues involved in the metal ion (Mn^{2+}) coordination, such as phosphate group of the ligand, His347 and Asp254 of protein were fixed, and for other protein residues only the backbone atoms were fixed.

References:

1. Biet, T.; Peters, T. *Angew. Chem. Int. Ed*. **2001,** *40,* 4189.

2. Jayalakshmi, V.; Krishna, N. R. *J. Magn. Reson*. **2004,** *168*, 36.

3. Jayalakshmi, V.; Krishna, N. R. *J. Magn. Reson*. **2002,** *155*, 106.

4. Krishna, N. R.; Agresti, D. G.; Glickson, J. D.; Walter, R. *Biophys. J*. **1978,** *24*, 791.

5. Xu, Y.; Sugar, I. P.; Krishna, N. R. *J. Biomol. NMR.* **1995**, *5,* 37.

6. Driscoll, P. C.; Gronenborn, A. M.; Beress, L.; Clore, G. M. *Biochemistry*. **1989,** *28*, 2188.

7. Li, A. J.; Nussinov, R. *Proteins*. **1998,** *32,* 111.

8. Alotto, P. G.; Eranda, C.; Brandstatter, B.; Furntratt, G.; Magele, C.; Molinari, G.; Nervi, M.; Preis, K.; Repetto, M.; Richter, K. R. *IEEE Trans. Magn*. **1998**, *34*, 3674.

9. Lipari, G.; Szabo, A. *J. Am Chem. Soc*. **1982**, *104*, 4546.

10. Dellwo, M. J.; Wand, A. J. *J. Am Chem. Soc*. **1993**, *115*, 1886.

11. Petrova, P.; Koca, J.; Imberty, A. *Eur. J. Biochem.* **2001,** *268,* 5365.

12. Ramakrishnan, B.; Balaji, P. V.; Qasba, P. K.; *J. Mol. Biol*. **2002,** *318,* 491.

13. Deitrick, M. K.; Egile, C.; Hoyt, D. W.; Ford, J. J; Li, R.; Rosen, M. K. *Biochemistry*. **2003**, 42, 8579.