Entropy−Entropy Compensation between the Protein, Ligand, and Solvent Degrees of Freedom Fine-Tunes Affinity in Ligand Binding to Galectin-3C

Johan Wallerstein, Vilhelm Ekberg,[‡] Majda Misini Ignjatović,[‡] Rohit Kumar,[‡] Octav Caldararu, Kristoffer Peterson, Sven Wernersson, Ulrika Brath, Hakon Leffler, Esko Oksanen, Derek T. Logan, Ulf J. Nilsson, Ulf Ryde,* and Mikael Akke*

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆂�🅸 Supporting Information

ABSTRACT: Molecular recognition is fundamental to biological signaling. A central question is how individual interactions between molecular moieties affect the thermodynamics of ligand binding to proteins and how these effects might propagate beyond the immediate neighborhood of the binding site. Here, we investigate this question by introducing minor changes in ligand structure and characterizing the effects of these on ligand affinity to the carbohydrate recognition domain of galectin-3, using a combination of isothermal titration calorimetry, X-ray crystallography, NMR relaxation, and computational approaches including molecular dynamics (MD) simulations and grid inhomogeneous solvation theory (GIST). We studied a congeneric series of ligands

with a fluorophenyl-triazole moiety, where the fluorine substituent varies between the *ortho*, *meta*, and *para* positions (denoted O, M, and P). The M and P ligands have similar affinities, whereas the O ligand has 3-fold lower affinity, reflecting differences in binding enthalpy and entropy. The results reveal surprising differences in conformational and solvation entropy among the three complexes. NMR backbone order parameters show that the O-bound protein has reduced conformational entropy compared to the M and P complexes. By contrast, the bound ligand is more flexible in the O complex, as determined by ^{19}F NMR relaxation, ensemble-refined X-ray diffraction data, and MD simulations. Furthermore, GIST calculations indicate that the O-bound complex has less unfavorable solvation entropy compared to the other two complexes. Thus, the results indicate compensatory effects from ligand conformational entropy and water entropy, on the one hand, and protein conformational entropy, on the other hand. Taken together, these different contributions amount to entropy−entropy compensation among the system components involved in ligand binding to a target protein.

KEYWORDS: Drug design, Molecular recognition, Ligand binding specificity, Conformational entropy, Solvation entropy, Thermodynamics
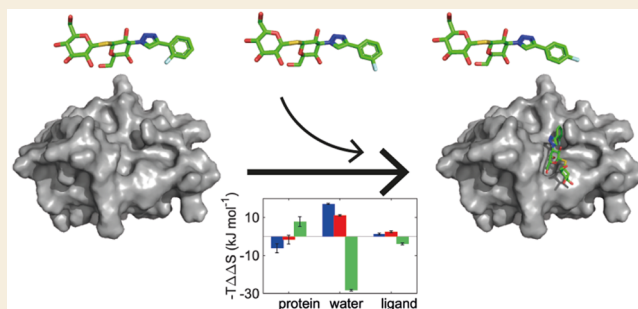
INTRODUCTION

Binding of low-molecular-weight ligands to proteins is fundamental to a large number of signaling pathways in biology and also of central interest in drug design aiming to interfere with such pathways for medicinal purposes. Rational structure-based design of effectors by computational approaches has advanced immensely in recent decades yet remains an extremely challenging endeavor. The main challenge involves the fact that the free energy of binding is manifested as a small difference between large numbers originating from a vast number of interactions between the protein, ligand, other solutes, and solvent molecules. The interaction energies depend sharply on interatomic distances and orientations, which vary in a time-dependent manner with the conformational dynamics. Furthermore, changes in

conformational fluctuations in proteins upon ligand binding can give rise to significant entropic contributions to affinity.[1−8] Ligand binding may also affect the water network in and around binding sites, causing appreciable effects on the thermodynamics of solvation.[8−13]

We have studied the carbohydrate recognition domain of galectin-3 (residues 113−250; denoted galectin-3C) to address specifically the roles of conformational entropy[4,5,8] and
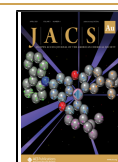
solvation in ligand binding.[8,13] Galectin-3 is an interesting model system in this regard, because its binding site is relatively solvent-accessible, being located in a shallow groove across one of the two β-sheets. Several water molecules form an integral part of the ligand environment by forming bridging hydrogen bonds between the ligand and protein.[13,14] Galectin-3 is a member of the galectin family of mammalian lectins, which is defined by a conserved sequence motif that confers affinity for β-galactoside-containing glycans.[15,16] Galectins play important roles in cell growth, cell differentiation, cell cycle regulation, signaling, and apoptosis, making them targets for medicinal chemistry programs to treat inflammation and cancer,[17,18] with several cases reported for galectin-3.[19−21] Recently, it has also been suggested that galectin-3 plays a role in the pathology of Alzheimer's disease.[22]

Here, we investigate the effects of minute changes in ligand structure on the binding enthalpy and entropy, broken down into separate contributions from various interactions or degrees of freedom. We report a comparative analysis of galectin-3C in complex with three congeneric fluorophenyl-triazolyl-thiogalactosides that differ by the position of the fluorine substituent, which is placed in the *ortho*, *meta*, or *para* position; henceforth, the ligands are denoted O, M, and P. The advantage of using ligands with minimal structural differences is that differences in binding thermodynamics are dominated by the properties of the ligand−protein complexes, while the chemical potential of the free ligands contributes relatively less in the comparative analysis. On the other hand, the minor structural differences between the ligands in this series lead us to expect limited differences in the structures of the complexes and in their binding thermodynamics. Thus, this study focuses on the thermodynamic effects of minute changes in protein−ligand structure and aims at characterizing the response of the entire system, including the solvent. In particular, we address how the various entropic contributions to binding might vary between slightly different protein−ligand complexes.

We used a combination of experimental and computational approaches, including isothermal titration calorimetry (ITC), competitive fluorescence polarization, X-ray crystallography, NMR relaxation, and molecular dynamics (MD) simulations followed by conformational entropy and grid inhomogeneous solvation theory (GIST) calculations. Our results show that minor differences between protein−ligand complexes in their overall binding thermodynamics might encompass greater differences among individual contributions, including a case of entropy−entropy compensation between the protein, ligand, and solvent degrees of freedom.

■ MATERIALS AND METHODS

Ligand Synthesis

The synthesis of ligands M and P was performed as described for O.[23] 1,3-Dipolar cycloaddition of 3-fluoro- and 4-fluorophenylacetylene to 3′-azido-3′-deoxy-β-D-galactopyranosyl-1-thio-β-D-glucopyranoside under Cu(I) catalysis afforded ligands M and P, respectively. Procedures and physical data for M and P are described in the Supporting Information.

Protein Expression and Purification

Galectin-3C was expressed and purified following published protocols[4,5] to produce ^{15}N-, ^{15}N/^{13}C-, and ^{15}N/^{13}C/^{2}H-labeled protein as well as unlabeled protein. The protein stock solution was 16 mg/mL in buffer, consisting of 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, 2.7 mM KCl, pH 7.3, 2 mM ethylenediaminetetraacetic acid (EDTA), 4 mM tris(2-carboxyethyl)-

phosphine hydrochloride (TCEP), and 150 mM lactose. The protein stock solution was stored at 278 K.

Isothermal Titration Calorimetry

^{15}N-labeled galectin-3C samples were prepared by extensive dialysis (Slide-A-Lyzer MINI Dialysis ThermoScientific) against 5 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer to remove all lactose, followed by centrifugation at 14 000 rpm for 5 min to remove any aggregates. The protein concentration was determined by UV absorption at 280 nm, as described before.[4] The ligands are water-soluble and do not require addition of DMSO. The three ligands were dissolved into 7−10 mM stock solutions using the same HEPES buffer as used in the protein samples and sonicated immediately prior to experiments. ITC experiments were performed on a MicroCal PEAQ−ITC instrument (Malvern Panalytical) at pH 7.3 and a temperature of 301.1 K by titrating the protein at a concentration of 100 ± 4 μM into the cell containing the ligand at a concentration of 1000 ± 10 μM. Two replicate experiments were performed for each complex. The time between injections was set to 4 min. Peak integration was done using NITPIC,[24] with individual error bars assigned to each injection. A single-site binding model was fitted simultaneously to the two titration curves to yield the binding enthalpy ($\Delta H°$), fraction of binding-competent protein (n), and dissociation constant (K_d), using ITCsy.[25] The heat released or absorbed during the ith injection is given by[26]

$$\Delta Q_i = Q_i - Q_{i-1} + (V_i/V_0)[Q_i - Q_{i-1}]/2 + Q_{off}$$

where V_i is the volume of the ith injection, V_0 is the cell volume, Q_{off} is an offset parameter accounting for the heat of mixing, and Q_i is the heat following the ith injection

$$Q_i = (\Delta H V_0/2)[\alpha - \sqrt{\alpha^2 - 4nM_iX_i}]$$

where $\alpha = nM_i + X_i + K_d$ and M_i and X_i are the total concentrations of protein and ligand, respectively, in the cell at any given point of the titration.

The free energy and entropy of binding were determined via the equations $\Delta G° = RT \ln(K_d)$ and $-T\Delta S° = \Delta G° - \Delta H°$. The statistical uncertainty is greater for $-T\Delta S°$ than for $\Delta G°$ or $\Delta H°$, since it is derived from two parameters. ITCsy/SEDPHAT produces asymmetric error estimates, but in the present case, the errors are nearly symmetric, and we report the errors as the average of upper and lower errors.

X-ray Crystallography

Small crystals of lactose-bound galectin-3C were grown with the hanging drop method in NeXtal plates (Qiagen) and with the following reservoir condition: 20% (w/v) PEG 4000, Tris-HCl pH 7.5, 0.4 M NaSCN, 10 mM β-mercaptoethanol. The drop volume was 2 μL, and the protein solution/reservoir ratio was 1:1. Small crystals were then moved to drops containing the same reservoir with the addition of 10 mM of the ligand (O, M, or P), from a 100 mM stock solution in PBS buffer. Soaking lasted for 12−15 h. Before data collection, crystals were placed for a couple of seconds in a drop containing 1 volume of 100% PEG 400 and 3 volumes of crystallization solution to prevent ice formation during cryo-cooling. Data were collected on EMBL beamline P13 at PETRA-III, DESY, Hamburg.[27] For each complex, 3600 diffraction images were collected with 0.1° rotation and 0.01 s exposure time. All data were integrated using XDS.[28]

Scaling and merging were done with Aimless.[29] Cross-validation was based on 10% of the reflections. Molecular replacement was done using Phaser in the Phenix suite version 1.14,[30,31] using as template the lactose−galectin-3C structure[13] with lactose and water molecules removed. Ligands and their crystallographic restraints were generated through phenix.eLBOW.[32] Restrained refinement was then performed using phenix.refine[33] using all remaining reflections. Manual rebuilding, including addition of water molecules, was done using Coot.[34] Data collection and refinement statistics for analysis of X-ray crystallography data are found in Table S1.

Ensemble Refinement of Crystal Structures

Ensemble refinement of the X-ray diffraction data was performed using the module phenix.ensemble_refinement[35] in the Phenix 1.14 software suite. The X-ray crystal structures of the M, P, or O complexes were used as starting structures. The crystallographic water molecules were kept, and hydrogen atoms and missing atoms in the protein were added using the phenix.ready_set module. Ligand restraints and coordinates were the same as those used in the original refinement.

The collective dynamics of the protein was described using a TLS (translation−libration−screw) model with a single group, which included both the protein and the ligand atoms. The percentage of atoms included in the TLS fitting (p_{TLS}) was optimized by testing five different values (0.5, 0.6, 0.7, 0.8, and 0.9) and choosing the one that yielded the lowest R_{free}, which was $p_{TLS} = 0.6$ for all three protein−ligand complexes. An ensemble of structures was then generated by running MD simulations, in which the model was restrained by a time-averaged X-ray maximum-likelihood target function. Three X-ray weight-coupled temperature bath offsets were tested (2.0, 5.0, and 10.0 K), with the default value of 5 K giving the lowest R_{free}. A 1.25 ps relaxation time of the time-averaged restraints was used, resulting in 25 ps long MD simulations, with structures stored every 0.05 ps. All structures generated by ensemble refinement were kept, resulting in 500 different structures in each ensemble. Atomic fluctuations were calculated using the *cpptraj* module of Amber after removal of the water molecules.[36]

NMR Sample Preparation

Samples of ^{15}N-labeled, ^{15}N/^{13}C-labeled, or ^{15}N/^{13}C/^{2}H-labeled galectin-3C in complex with each of the M, P, and O ligands were prepared in the same buffer as used for ITC. Samples were prepared by titrating 5 μL of ligand stock solution in 6−9 steps to the NMR tubes containing a protein concentration of 0.8−0.9 mM and recording a ^{15}N-heteronuclear single quantum correlation (HSQC) spectrum at each titration point. The resulting saturation levels were 98.0−98.5%, as calculated from the value of K_d determined by ITC measurements. In addition, ^{15}N-labeled samples of each complex were obtained by pooling the cell contents from the two replicate ITC experiments, resulting in a galectin-3C concentration of 0.08 mM. ^{15}N-HSQC spectra were acquired on these samples to confirm that the samples resulting from the ITC titrations are essentially identical to those used in the NMR relaxation experiments.

NMR Resonance Assignments and Chemical Shift Perturbation Analysis

Backbone chemical shift assignments were based on HNCACB[37] and CBCA(CO)NH[38] spectra and previous assignments for various galectin-3C complexes.[5] The ligands were titrated into the NMR tubes in five to seven steps so as to follow the movement of individual resonances from the apo state to the ligand-bound state. Methyl groups were assigned using ^{13}C-HSQC, CCH-TOCSY, and HCCH-TOCSY experiments[39,40] and previous assignments.[5] All spectra were processed using NMRPipe,[41] employing a solvent filter, squared cosine apodization, linear prediction, and zero filling to twice the number of points in all dimensions. Resonance assignments were done using the CcpNmr program suite.[42] Chemical shift perturbations were calculated as $([\Delta\delta(^{1}H)]^2 + [0.16\Delta\delta(^{15}N)]^2)^{1/2}$ for backbone amides and $([\Delta\delta(^{1}H)]^2 + [0.25\Delta\delta(^{13}C)]^2)^{1/2}$ for methyl-bearing side chains.

NMR Relaxation Experiments and Data Analysis

Backbone amide ^{15}N R_1, R_2, and {^{1}H}−^{15}N nuclear Overhauser effect (NOE) experiments were performed at static magnetic field strengths of 11.7, 14.1, and 18.8 T and a temperature of 301 K. Temperature calibration was done prior to each relaxation series experiment using a type-T copper-constantan thermocouple element with one electrode in an ice−water bath and the other in an NMR tube in water, positioned at the sample location inside the magnet. Experiments at 11.7 and 14.1 T were performed on Varian/Agilent VNMRS DirectDrive spectrometers, and experiments at 18.8 T were performed

on a Bruker Avance III HD spectrometer. The spectral widths were 14−16 and 28−30 ppm for ^{1}H and ^{15}N, respectively, covered by 1024 and 128 points. Relaxation decays were recorded with 10 relaxation delays ranging between 0−1 s for R_1 and 0−0.2 s for R_2 with a 1.2 ms delay between refocusing pulses. The NOE experiments at 11.7, 14.1, and 18.8 T all used the same ^{1}H saturation time of 7 s, and the recycle delay between experiments was 3, 7, and 3 s, respectively. The reference experiment (without ^{1}H saturation) was acquired using a recycle delay of 10, 14, and 10 s at 11.7, 14.1, and 18.8 T, respectively. All spectra were processed as described above. Relaxation rates were extracted using PINT, which employs line-shape fitting to resolve overlapped peaks.[43] Peak intensities were evaluated using a weighted sum of Lorentzian and Gaussian line shapes. Monoexponential functions were fitted to the R_1 and R_2 relaxation decays using a jackknife error estimation.[44] NOEs were calculated as the ratio of the peak intensities in the saturated and reference experiments, and the standard deviations were determined by propagating the errors of intensities estimated from the baseplane noise. Error estimates are reported as 1 standard deviation (SD).

Methyl ^{2}H relaxation experiments[45] measuring $R_1(D_Z)$, $R(3D_Z^2 − 2)$, $R_2(D_+)$, and $R(D_+D_Z + D_ZD_+)$ were recorded at 11.7 and 14.1 T. Spectral widths were 16 and 20 ppm for ^{1}H and ^{13}C, respectively, covered by 1024 points in the ^{1}H dimension at both field strengths and by 70 and 84 points for ^{13}C at 11.7 and 14.1 T, respectively. Relaxation decays were sampled by 9 points covering 0−0.1 s for $R_1(D_Z)$ and $R(3D_Z^2−2)$ and 0−20 ms for $R_2(D_+)$ and $R(D_+D_Z+D_ZD_+)$. The recycle delay was 1.8−2 s. Peak intensities were evaluated as described above.

^{19}F relaxation R_2 experiments were recorded at 470 and 659 MHz on Agilent/Varian VNMRS and Bruker spectrometers, respectively. The temperature was calibrated to 301 ± 0.1 K using a methanol reference sample. ^{1}H decoupling during the relaxation period was achieved using a 180° pulse or WALTZ train. The experiment at 470 MHz utilized a spectral width of 4596 Hz, 3584 transients, relaxation delays of 0, 4, 8 (×2), 12, 16, 20, 28, and 40 ms, and a recycle delay of 4 s. The experiment at 659 MHz utilized a spectral width of 13 158 Hz, 512 transients, relaxation delays of 0, 2, 4 (×2), 6, 8 (×2), 12, 16, 20 (×2), 28, 32, and 40 ms, and a recycle delay of 5 s. Both experiments were run in an interleaved manner. All ^{19}F spectra were processed using NMRPipe[41] with zero filling and matched filter apodization. Peak intensities were evaluated using in-house MATLAB scripts. Errors in peak intensities were estimated from the baseline noise. Monoexponential functions were fitted to the R_2 decays using least-squares optimization functions available in MATLAB, with parameter uncertainties estimated from 1000 Monte Carlo simulations.[46]

Model-Free Analysis of NMR Relaxation Data

Backbone NH model-free parameters[47,48] were fitted using the *relax* software package (v 4.0.2),[49−51] with an NH bond length of 1.02 Å and a ^{15}N chemical shift anisotropy of −172 ppm. Hydrogen atoms were added to the PDB structures using UCSF Chimera.[52] The backbone optimization in *relax* was restricted to the four most basic models defined by the parameter sets: {O^2}, {O^2, τ_e}, {O^2, R_{ex}}, or {O^2, τ_e, R_{ex}}, where O^2 denotes the generalized order parameter, τ_e denotes the effective correlation time for the internal motion, and R_{ex} denotes exchange contributions to R_2; in addition, the correlation time for overall rotational diffusion τ_c was fitted as well as the diffusion tensor.[53] The definitions of anisotropy (ζ) and rhombicity (η) of the diffusion tensor[54] are $\zeta = 2D_{ZZ}/(D_{XX} + D_{YY})$ and $\eta = 1.5(D_{YY} − D_{XX})/[D_{ZZ} − 0.5(D_{XX} + D_{YY})]$, where {$D_{XX}$, D_{YY}, D_{ZZ}} are the principal components of the molecular rotational diffusion tensor.

Side-chain methyl-axis model-free optimization was performed using in-house MATLAB routines employing the fmincon function to find the minimum of a constrained nonlinear multivariable function. Numerically stabilized model-free equations suggested previously were used.[51] Three models were fitted using two {O^2, τ_f}, three {O^2, τ_f, τ_{eff}}, or four {O_f^2, O_s^2, τ_f, τ_{eff}} parameters, where τ_f is associated with fast motions, $\tau_{eff} = (1/\tau_c + 1/\tau_s)^{-1}$, and τ_s denotes the correlation time for slow internal motions on par with τ_c.[55] The global correlation

time and diffusion tensor were fixed to values obtained from the ^{15}N backbone model-free optimization in *relax*. Model selection was performed using an *F* test at the level $\alpha = 0.95$ ($p < 0.05$).[56] Error estimates, reported as 1 SD, were based on Monte Carlo simulations using 500 samples.[46]

Comparison of Three Complexes

We report the comparisons of the three complexes as the difference between the parameter value of a certain property of complex A minus the average value for the other two complexes, B and C, i.e., $\Delta P(A) = P(A) - (P(B) + P(C))/2$. The values of ΔP resulting from this intercomplex comparison are a factor of 1.5 greater than differences calculated relative to the average value of all three complexes. The sum of all three $\Delta P(i)$ is zero.

Conformational Entropy Estimates from Order Parameters

The change in backbone conformational entropy, comparing two states, was estimated from the NMR order parameters using the relationship[1,57]

$$\Delta S = R \sum_k \ln[(1 - O^2_{A,k})/(1 - O^2_{B,k})] \qquad (1)$$

where R is the gas constant and the sum runs over all residues k. Using the intercomplex comparison described above, the equation becomes

$$\Delta S_A = R \sum_k \ln\{(1 - O^2_{A,k})/[(1 - O^2_{B,k})(1 - O^2_{C,k})]^{1/2}\} \qquad (2)$$

The conformational entropy change of the side-chain methyl axis was calculated as[57]

$$\Delta S = R \sum_m C_m \sum_n (O^2_{A,n} - O^2_{B,n}) \qquad (3)$$

where C_m is a parameter that depends on the residue type: $C_m = 1.32$ for Val and Thr, 3.1 for Ile and Leu, and 2.31 for Met.[57] The sums run over all residues n of type m. Since eq 3 is linear, the intercomplex difference is calculated in a straightforward manner. The entropy for Ala side chains was calculated using eq 1.

Radial Distribution of Conformational Entropy

The radial distribution of protein conformational entropy around the bound ligand was calculated by averaging the backbone entropy of residues located in 1−2 Å thick shells surrounding the ligand. The distance from the ligand was evaluated as the shortest distance between any ligand atom and any atom of a given residue in the corresponding X-ray crystal structure. The residues were sorted into shells with a thickness of 1 Å (2−9 Å) or 2 Å (9−27 Å); the thinner shells were used to improve the resolution of the data close to the ligand. The lowest entropy value for each complex was taken as the reference and set to 0. The cumulative average entropy values were also computed as a function of the distance from the ligand.

Molecular Dynamics Simulations and Analysis

All MD simulations were run with the Amber software suite.[58] The X-ray crystal structures of the three complexes were used as the starting points for MD simulations. All crystal-water molecules were kept in the simulations. Each galectin-3C complex was solvated in an octahedral box of water molecules extending at least 10 Å from the protein using the *tleap* module. The simulations were set up in the same way as in our previous studies of galectin-3C.[4,8,59,60] All Glu and Asp residues were assumed to be negatively charged, and all Lys and Arg residues positively charged, whereas the other residues were neutral. The binding-site residue His158 was protonated on the ND1 atom, whereas the other three His residues were protonated on the NE2 atom, in accordance with the neutron structure and NMR measurements[14] as well as previous extensive test calculations with MD.[61] This resulted in a net charge of +4 for the protein. No counterions were used in the simulations.

The protein was described by the Amber ff14SB force field,[62] water molecules with the TIP4P-Ewald model,[63] whereas the ligands were treated with the general Amber force field (GAFF).[64] Charges for the ligands were obtained with the restrained electrostatic potential

method.[65] The ligands were optimized with the semiempirical AM1 method, followed by a single-point calculation at the Hartree−Fock/ 6-31G* level to obtain the electrostatic potentials, sampled with the Merz−Kollman scheme.[66] These calculations were performed with the Gaussian 09 software.[67] The potentials were then used by antechamber to calculate the charges. A few missing parameters were taken from a previous study.[8]

For each complex, 1000 steps of minimization were used, followed by a 20 ps constant-volume equilibration and 20 ps constant-pressure equilibration, all performed with heavy nonwater atoms restrained toward the starting structure with a force constant of 4184 kJ/mol/Å^2. Finally, the system was equilibrated for 1 ns, followed by 100 ns of production simulation, both performed with constant pressure and without any restraints. For each protein−ligand complex, 10 independent simulations were run.[68] Consequently, the total simulation time for each complex was 1 μs. All bonds involving hydrogen atoms were constrained to the equilibrium value using the SHAKE algorithm,[69] allowing for a time step of 2 fs. The temperature was kept constant at 301 K using Langevin dynamics,[70] with a collision frequency of 2 ps^{-1}. The pressure was kept constant at 1 atm using a weak-coupling isotropic algorithm[71] with a relaxation time of 1 ps. Long-range electrostatics were handled by particle-mesh Ewald (PME) summation[72] with a fourth-order B spline interpolation and a tolerance of 10^{-5}. The cutoff radius for Lennard−Jones interactions between atoms of neighboring boxes was set to 8 Å. The snapshots were analyzed with the *cpptraj* module.[36] We also performed simulations of the three ligands free in solution, using the same setup.

Conformational Entropy Estimates from MD Simulations

We calculated NH order parameters from the MD trajectories using isotropic reorientational eigenmode dynamic analysis.[73] The covariance matrix of the NH bond vectors was obtained from the trajectories by the *cpptraj* module[36] in the Amber software.[58]

Conformational entropies were calculated from the ensemble of configurations of the protein and ligands by analyzing the dihedral angle fluctuations.[4,60,74,75] The Cartesian coordinates from the trajectories were transformed to internal coordinates, and the entropies were then calculated from probability distributions over all possible states of these coordinates using a bin size of 5° (i.e., 72 bins per dihedral). Entropies were normalized to that of a free rotor.[4] All entropies are reported as $-T\Delta S$ at 301 K. Protein-backbone entropies were calculated from two-dimensional distributions of the ϕ and ψ dihedral angles. In our previous study, we showed that the effect of correlated motions is minimal for entropy estimates of galectin-3C, less than 1 kJ/mol.[8]

A total of 100 000 snapshots with a 10 ps sampling frequency were used for entropy and order parameter estimates, employing separate simulations for the complexes, for free galectin-3C, and for the solvated ligands. Entropies and order parameters were calculated as averages over 200 simulations of 5 ns (with 500 snapshots in each, i.e., each of the 10 simulations was divided into 20 parts of equal length). The 5 ns time window is similar to the rotational correlation time of the protein (∼7 ns). This procedure yields more stable entropy estimates by restricting the dependence on rare events.[60] The reported uncertainties are standard errors over these 200 simulations.

Solvation Thermodynamics

We identified water sites and analyzed the thermodynamics of the solvent around the galectin-3C complexes, using GIST,[76,77] implemented in the *cpptraj* module of the Amber software. The method requires snapshots from MD simulations in which the solute is kept restrained. Therefore, we performed 10 independent 100 ns long MD simulations, in which the solute was kept restrained toward the starting crystal structure (the unrestrained MD simulations did not show any extensive movements of the ligands or the protein, except for individual side chains).

For each protein−ligand system, the water−water interaction energy, E_{ww}, and solute−water interaction energy, E_{sw}, as well as translational, S_{trans}, and rotational, S_{rot}, entropy contributions were calculated for a rectangular grid of dimensions 24 × 14 × 14.5 Å, centered on the ligand and extended at least 3 Å on each side of the

ligand; for the free ligands, the dimensions were 16.5 × 18 × 21 Å. The grid was divided into cubic boxes (0.5 × 0.5 × 0.5 Å), for which the thermodynamic properties were calculated. The sum of these properties over the entire region reveals the changes in the hydration thermodynamics of the region for each cluster, relative to bulk water.

Solvation free energies for the free ligands in water solution were calculated with the conductor-like solvent model (COSMO)[78,79] real-solvent (COSMO-RS) approach[80,81] using the COSMOTHERM software.[82] This is one of the most accurate methods to estimate solvation free energies of small molecules.[83,84] The calculations were based on two single-point density-functional theory calculations with the BP86 method[85,86] and the TZVP basis set,[87] one performed in a vacuum and the other in the COSMO solvent with an infinite dielectric constant, as required by method. The structures were taken from unrestrained MD simulations of the free ligands, 110 snapshots from each ligand.

RESULTS AND DISCUSSION

Ligand Design

We synthesized o-, m-, and p-fluoro-phenyltriazolyl-galactosylthioglucosides (O, M, and P; see Figure 1) following
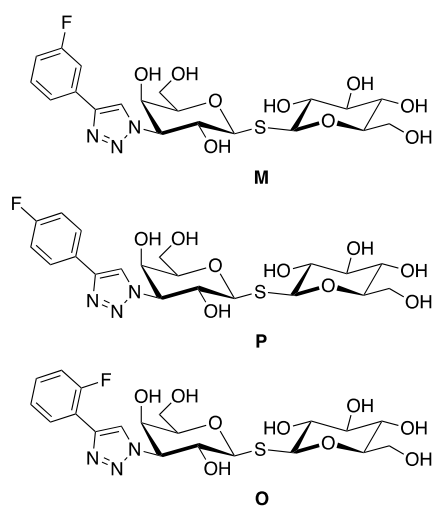
Figure 1. Chemical structures of the three fluorinated phenyltriazolyl-galactosylthioglucoside galectin-3C ligands M, P, and O.

protocols published recently.[23] The present series of ligands includes a glucose unit, instead of the S-toluene group used in previous studies,[23] in order to make the ligands water-soluble without the addition of DMSO.

Overall Binding Thermodynamics

We characterized the overall macroscopic thermodynamics for the binding of the three ligands to galectin-3C using ITC. We performed two replicate titrations with each of the ligands, followed by a combined fit of the duplicate data (Figure 2A–C). Table 1 lists the resulting thermodynamic parameters, which are also summarized in Figure 2D. The resulting dissociation constants fall in the low micromolar range, $K_d(M)$ = 2.0 ± 0.3 μM, $K_d(P)$ = 2.5 ± 0.2 μM, and $K_d(O)$ = 7.2 ± 1.3 μM. Similar to our previous observations for other ligand series,[5,8] the relative affinities correlate well with results from competitive fluorescence polarization experiments. The $\Delta H°$ values for M− and P−galectin-3C are −50.4 ± 1.2 and −49.1 ± 0.8 kJ/mol, hence equal within error, while O−galectin-3C has a smaller enthalpic contribution of −45.5 ± 1.9. The entropies ($-T\Delta S°$) are 17.5 ± 1.2, 16.7 ± 0.8, and 15.8 ± 1.9 kJ/mol for the complexes with M, P, and O, respectively. The

data follow a trend suggesting enthalpy−entropy compensation (Figure 2E), as often seen in ligand−protein interactions, but the association between $\Delta H°$ and $-T\Delta S°$ is not statistically significant (p = 0.15). Taken together, the ITC experiments reveal that the complexes have similar thermodynamic signatures overall, which might be expected from the similar structure of the ligands. The samples resulting from the ITC titrations were characterized by ^{1}H−^{15}N HSQC spectroscopy and compared to the separately prepared NMR samples, to make sure that the protein chemical shifts were identical in the two samples, thereby verifying that the sample conditions and protein characteristics were virtually identical in the two types of experiments. Based on the determined K_d values and reactant concentrations, the saturation levels of the complexes after the last ITC injection were 97.6, 97.0, and 92.4% for M, P, and O, respectively.

Crystal Structures Reveal Subtle Differences in Binding Modes

Very high-resolution crystal structures were obtained for galectin-3C in complex with the three ligands (resolution = 0.94−1.01 Å, cf. Table S1). The ligands are very well-defined by the electron density (Figure S1). The overall binding mode of the ligands (Figure 3A) is identical to that of previously reported analogous phenyltriazolyl-thiodigalactosides.[23] The galactose moiety forms a CH−π stacking interaction with Trp181 and the glucose HO2 forms hydrogen bond interactions with Arg162, Glu184, and Arg186 (Figure 3B).[13,88] Arg144 forms a cation−π (or π−π) interaction with the phenyl group.[89] The orientation of the phenyl ring differs subtly with the position of the fluorine (Figure 3B). The fluorophenyltriazole moiety has polar interactions with the backbone carbonyl moieties of Arg144, Ile145, Asn160, Ser237, and Gly238. Fluorine atoms tend to orient toward nearby backbone amides, side-chain amides, and carbonyl carbons in an orthogonal manner.[90−92] The geometry of these interactions depends on the distance to the peptidic NH and CO groups, such that the angle tends toward 90° at shorter distances. Figure 3C−E shows how the fluorines are oriented toward the neighboring carbonyl moieties. The fluorine in M forms an interaction with the carbonyl moieties of Arg144 and Ile145. The fluorine in P interacts with the carbonyl moiety of Ser237. For the fluorine in O, there is no peptide backbone nearby. Instead, it orients toward the carbonyl group of the Asn160 side chain. However, recent quantum-mechanical calculations have indicated that these fluorine−amide interactions have little influence on the binding affinities (the fluorine atom does not form stronger amide interactions than a phenyl hydrogen atom).[93] The electron density of the fluorophenyl group in O is slightly less well-defined (Figure S1), which likely is due to increased mobility. Possibly, the higher mobility of O could be related to the fluorine being coordinated by a side chain rather than a backbone carbonyl as in the case of M and P. The positions and occupancies of water molecules around the binding site are quite similar for all three ligands. Complete data collection and refinement statistics are listed in Table S1 of the Supporting Information.

Ensemble Refinement of Crystal Structures Highlights Differences in Flexibility

We carried out ensemble refinement of the three complexes in order to compare their flexibility in the crystal. The results are complementary to the order parameters measured by NMR relaxation and to the MD simulations (described below). The
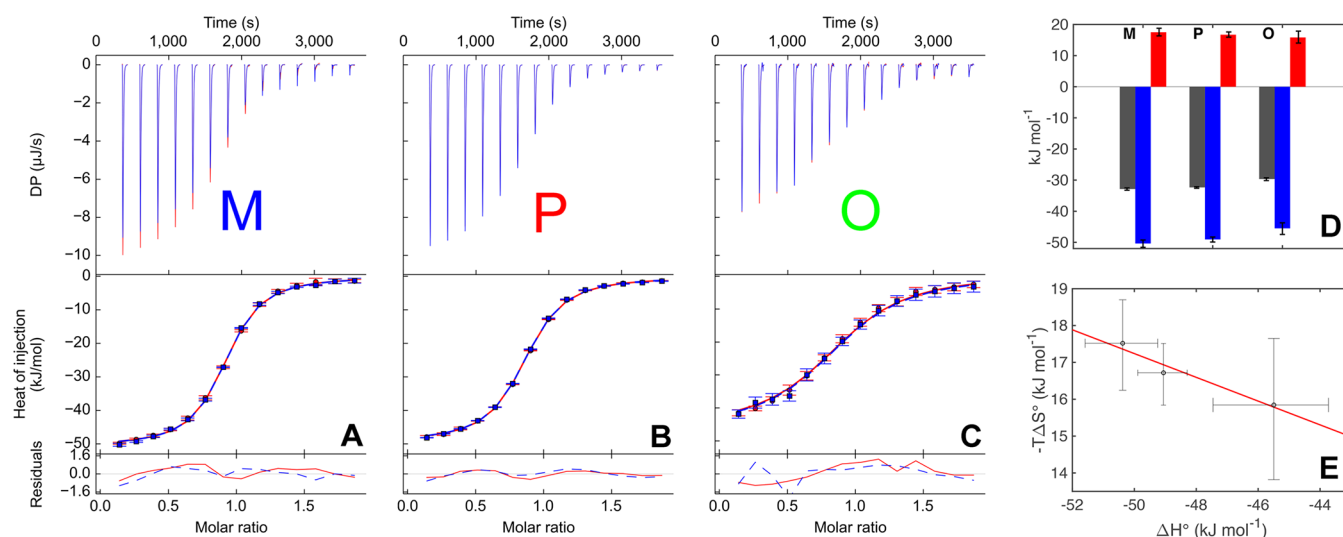
Figure 2. ITC experiments of ligand binding to galectin-3C. (A−C) Raw thermograms showing the differential power (upper panels) and isotherms with residuals of the fits (lower panel). Duplicate data sets are shown in red and blue. The ligands are ordered left to right according to their affinity. The concentrations of ligand and protein are identical in the different experiments, and the isotherms are therefore directly comparable. (D) Thermodynamic profiles of the three complexes with $\Delta G°$ (dark gray), $\Delta H°$ (blue), and $-T\Delta S°$ (red). (E) Graph of $-T\Delta S°$ versus $\Delta H°$ suggesting enthalpy−entropy compensation. The error bars indicate a very modest asymmetry, as determined by the ITCsy/ SEDPHAT software. The adjusted R^2 correlation coefficient is 0.89, and the p value is 0.15, showing that the statistical confidence for a linear relation between $\Delta H°$ and $-T\Delta S°$ is quite low.

Table 1. Overall Binding Thermodynamics from ITC

complex	K_d ITC (10^{-6} M)[a]	$\Delta G°_{tot}$ (kJ/mol)[a]	$\Delta H°_{tot}$ (kJ/mol)[a]	$-T\Delta S°_{tot}$ (kJ/mol)[a]
M	2.00 ± 0.29	−32.9 ± 0.4	−50.4 ± 1.2	17.5 ± 1.2
P	2.46 ± 0.24	−32.3 ± 0.3	−49.1 ± 0.8	16.7 ± 0.8
O	7.21 ± 1.28	−29.6 ± 0.4	−45.5 ± 1.9	15.8 ± 1.9

[a]Reported errors are averages of the asymmetric upper and lower error bounds (1 SD) given by ITCsy/SEDPHAT, which are shown Figure 2E.

resulting ensembles show what conformations are compatible with the crystallographic raw data and may point out groups that show larger movements than indicated by the crystallographic B factors. These coordinate distributions can be translated into conformational entropies, although not quantitatively.[94] Figure 4 shows that the M and P ligands exhibit only one conformation in the crystal structure, with M having slightly higher fluctuations. In contrast, O exhibits a relatively high degree of mobility of the fluorophenyl moiety. The glucose ring is also flexible in all three complexes, most so for O and least so for P. This indicates a higher conformational entropy of O in the complex. However, the protein atoms show the opposite behavior, with a slightly higher root-mean-square fluctuation (RMSF) for the M and P complexes (RMSF = 0.06 ± 0.01 Å for both complexes), whereas the O complex has a lower RMSF of 0.053 ± 0.01 Å. Individual protein residues show varying degrees of flexibility. Among those interacting with the ligands, Arg144 is the most flexible. This residue interacts through π−π stacking with the fluorophenyl ring in the ligands, but its flexibility is not significantly different between the three complexes. The side chain of Ser237, which is modeled in two alternative conformations in the crystal structures of all three complexes, also shows multiple conformations in the ensemble refinement simulations of all three complexes. The carbonyl groups that interact with the fluorine atoms of the ligands, viz., the Asn160 side chain for O,

the Ile145 backbone for M, and the Ser237 backbone for P, do not show any significant flexibility in any of the three complexes.

Chemical Shift Mapping of Ligand Binding

Differences in chemical shifts between protein−ligand complexes provide a sensitive indication of differences in structure and dynamics. By mapping chemical shift changes in the protein upon ligand binding, we verify that the ligands bind to the protein in solution in the same pose as that observed in the crystal structures. In addition, chemical shift differences between complexes serve to validate results of the ensemble-refined X-ray structures and can further identify subtle differences between the complexes in their structure and dynamics.

We assigned the chemical shifts of O−, M−, and P− galectin-3C based on HNCACB and CACB(CO)NH experiments and assignments reported previously.[4,5,8] Superimposed spectra of the three complexes are shown in the SI (Figures S2 and S3). In keeping with previous results,[5,8] ligand binding leads to chemical shift differences for the backbone amides throughout the protein. The chemical shifts of 115 backbone amides were used to monitor ligand binding (Figure 5; intercomplex comparisons are shown in Figure S4A). Significant chemical shift differences in the $^1H−^{15}N$ HSQC spectra between the three complexes are observed primarily in regions close to the ligand (identified by gray vertical bars in Figures 5 and S4B), including residues 144−147, 158−165, 173−175, 184, and 235−239 but also for residues further away, viz., residues 116, 139, and 243.

In several cases, the observed chemical shift differences among the three complexes can be rationalized in terms of structural differences observed in the crystal structures. Starting with residues to the left in the binding site as presented in Figure 3B, we notice that the backbone NH of Gly238 in the P complex shows the largest chemical shift perturbation of all residues. This effect is explained by the
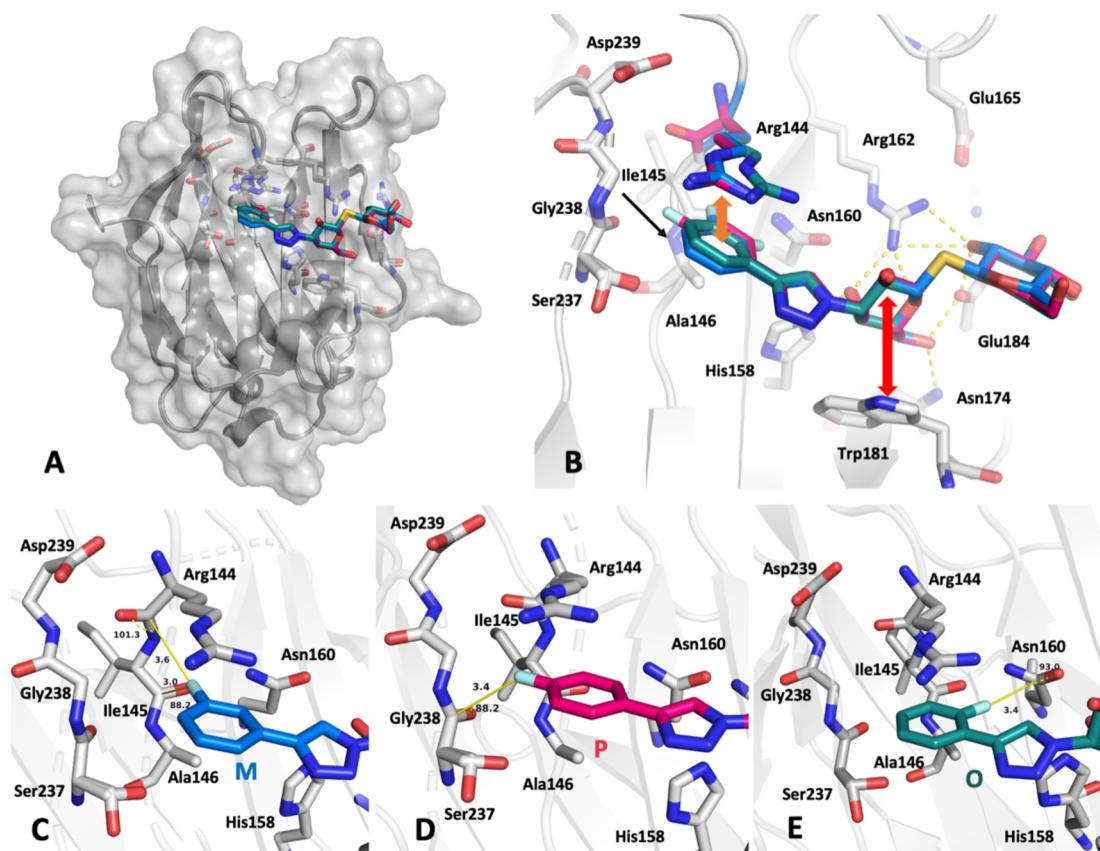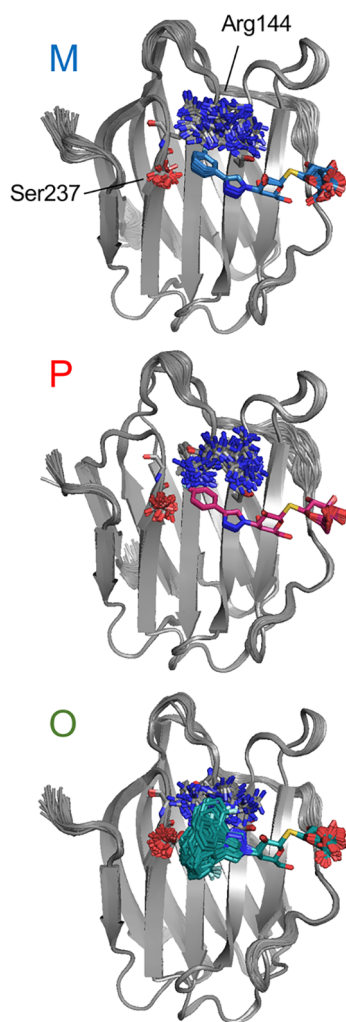
Figure 3. X-ray crystal structures of the ligand−galectin-3C complexes. (A) Surface view of the protein with all three ligands shown as sticks (O, M, and P colored green, blue, and red, respectively). (B) Overview of the binding site with the ligands superimposed. (C−E) Close-up view of the fluorophenyl moiety of the M (C), P (D), and O (E) ligands. The protein backbone is shown in ribbon representation (gray), whereas key ligand-coordinating backbone segments and side chains are shown as sticks. In panel B, key interactions are indicated: hydrogen bonds involving the ligand are shown as dashed yellow lines, the $\pi−\pi$ interaction involving Arg144 and the phenyl group of the ligand is shown as an orange double arrow, and the CH−π stacking interaction involving Trp181 and the galactose moiety is shown as a red double arrow. In panels C−E, distances and angles of the fluorine−amide interactions are indicated and depicted in yellow as solid lines and dashed semicircles, respectively. Note that the side chain of Ser237 is modeled in two conformations.

interaction between the carbonyl of Ser237 and the fluorine atom with an interatomic distance of 3.4 Å (Figure 3D). The NH shift of Gly238 is perturbed in the M and O complexes as well but not to the same extent. The backbone NH shift of Arg144 in the O complex differs considerably from the shift in the M and P complexes. This difference might reflect the average orientation of the Arg144 side chain, which is different in the O complex as compared to the M and P complexes (Figure 3B). The backbone NH chemical shifts of Ile145, Ala146, and Leu147 all show significant variation among the three complexes that likely can be explained in part by differences in the interactions between the fluorine and these amide groups. In the case of Ile145, the distance between the fluorine and the carbonyl group of the preceding residue varies from 3.6 Å in the M complex to 4.4 and 5.8 Å in the P and O complexes, respectively. As for Ala146, the corresponding distances are 3.0 Å (M), 4.1 Å (P), and 4.7 Å (O). Leu147 is further removed from the fluorine atoms with a distance greater than 5 Å in each complex, suggesting that variable fluorine interactions alone do not explain the observed variation in chemical shift. Rather, an additional contribution is expected from changes in the orientation of the fluorophenyl moiety (Figure 3B), which causes significant ring current shifts of the surrounding residues.

A number of other residues show NH chemical shift differences between complexes that cannot be explained by the static structures, which are virtually identical in these regions. Most likely, these shift changes report on differential dynamic averaging of the chemical shifts. Prominent examples include Arg162 and Glu165, the side chains of which pack together and interact with the glucose moiety, as seen in Figure 3B. These results reflect the differences in flexibility of the glucose ring observed in the ensemble-refined X-ray structures.

Chemical shift differences are also observed for methyl groups (Figure S4B), but there are few methyl-bearing side chains close to binding site. Notably, Ala146 Cβ shows the largest perturbation by far. The methyl group of Ala146 is located directly beneath the fluorophenyl ring (Figure 3C−E). Thus, the chemical shift of this methyl group is exquisitely sensitive to the ring orientation (cf., Figure 4) and the fluorine position. Val116 has the second-most perturbed methyl group, which is directed toward the fluorophenyl ring, located some 7 Å away. There is no difference in structure in this region among the complexes. However, we note that the backbone amide of Gly238 is positioned between Cγ1 of Val116 and the fluorophenyl group. Given that the NH of Gly238 has the most perturbed chemical shift, it might also be expected to find large perturbations for the Val116 side chain.

Figure 4. Ensemble-refined X-ray crystal structures. Overlay of the best 100 structures (in terms of the *R* factor) generated by ensemble refinement for M (blue), P (red), and O-bound (green) galectin-3C. Two residues that show large movement in all three complexes are shown in stick representation: Ser237 with its side-chain oxygen colored red and Arg144 with its side-chain nitrogens colored blue.
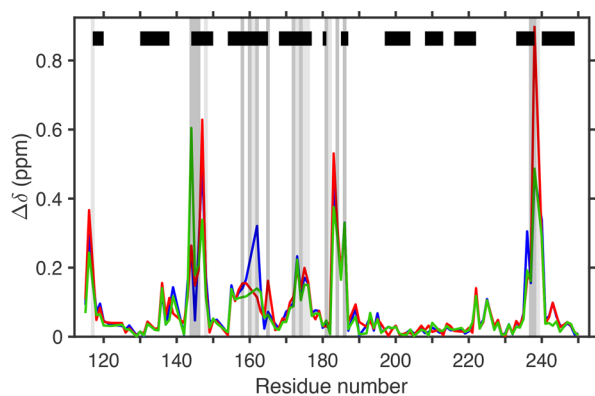
Figure 5. Chemical shift perturbations upon ligand binding. Backbone amide chemical shift differences relative to apo galectin-3C are shown for M−, P−, and O−galectin-3C in blue, red, and green, respectively. The black horizontal bars represent the *β*-sheet secondary structure of galectin-3C, and the gray vertical bars indicate residues with any atom closer than 4 Å (dark gray) or 6 Å (light gray) from any ligand atom.

Conformational Fluctuations Measured by NMR

We performed a series of protein NMR relaxation experiments to characterize the conformational dynamics on the pico- to nanosecond time scale. In the following analysis, we include only residues for which data are available for all three complexes, comprising 116 out of the 138 backbone amides and 73 out of the 85 methyl groups. The missing residues had cross-peaks that were overlapped or too broadened to allow for quantitative analysis.

We interpreted the relaxation data using the model-free formalism.[47,48] The best-fit rotational diffusion tensor determined from the backbone relaxation data is anisotropic for all three complexes, with ζ values of 1.11, 1.08, and 1.08 for the M, P, and O complexes, respectively. The P complex has an axially symmetric diffusion tensor, whereas the M and O complexes have rhombic tensors with η values of 0.54 and 0.75, respectively. The correlation times (τ_c) are 7.12, 6.97, and 6.95 ns for M, P, and O. ^{15}N relaxation data back-calculated from the optimized model-free parameters show good agreement with the experimental values for all three complexes (Figures S5−7).

In addition, we characterized the dynamics of the fluorinated phenyl rings of the bound ligands using ^{19}F R_2 relaxation experiments. The results reveal that the ^{19}F R_2 is significantly enhanced for the O ligand compared to the other two ligands, indicating that the O ligand undergoes conformational exchange on a fast time scale (Figure S8). In contrast, the M and P ligands both appear to populate a single conformation in the complex. Notably, these results are in perfect agreement with the ensemble-refined X-ray crystal structures (cf., Figure 4).

Differences in Backbone Order Parameters

The backbone order parameters are very similar among the three complexes (Figure 6A). Backbone O^2 values are listed in Table S2, and their distributions are represented as histograms in Figure S9. The mean values and standard error of the means are $<O^2> = 0.837 \pm 0.004$, 0.837 ± 0.004, and 0.843 ± 0.004 for M−, P−, and O−galectin-3C, respectively. There is no significant difference between the mean values, as gauged by one-way ANOVA ($p = 0.41$). The subtle differences in O^2 revealed by the intercomplex comparisons (Figure 6B; pairwise comparisons are shown in Figure S10) might be expected based on the small structural differences between the complexes. It is clear that M− and P−galectin-3C have more flexible residues close to the binding site than does O−galectin-3C, which is also reflected in the slightly higher $<O^2>$ for the O complex. These results are in qualitative agreement with RMSF values determined from the ensemble-refined X-ray structures (cf., Figure 4).

The high degree of similarity among the three complexes in O^2 for residues outside of the binding region attests to the high reproducibility of the relaxation rate measurements and indicates that statistically significant differences in O^2 can be meaningfully interpreted in terms of conformational entropy differences. The main differences in the backbone O^2 are seen for residues at the binding site (Figures 6B and S10). Among these, subtle differences in O^2 are observed in the segment 159−165, which includes several residues that are either in direct contact with the ligand (Asn160 and Arg162) or participating in the hydrogen-bonding network around the ligand (Glu165). Most likely, variation in side-chain interactions with the ligand among the different complexes affects
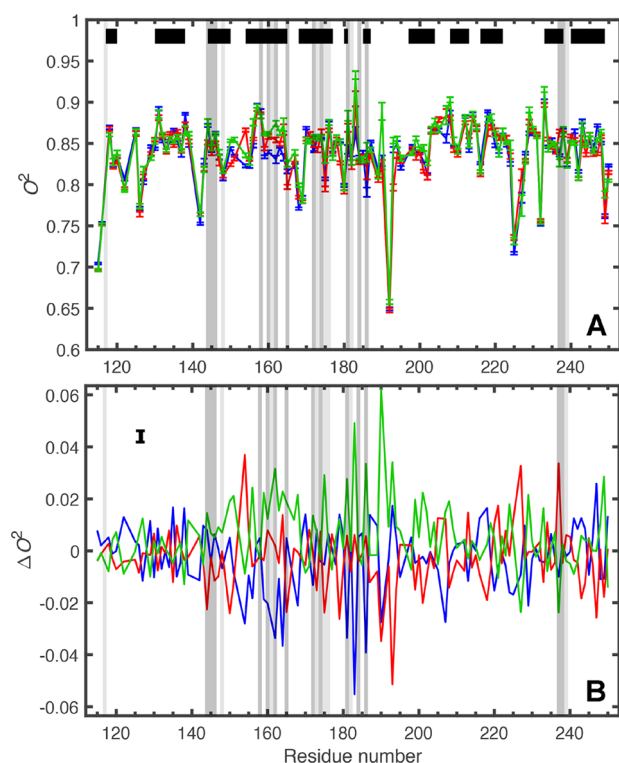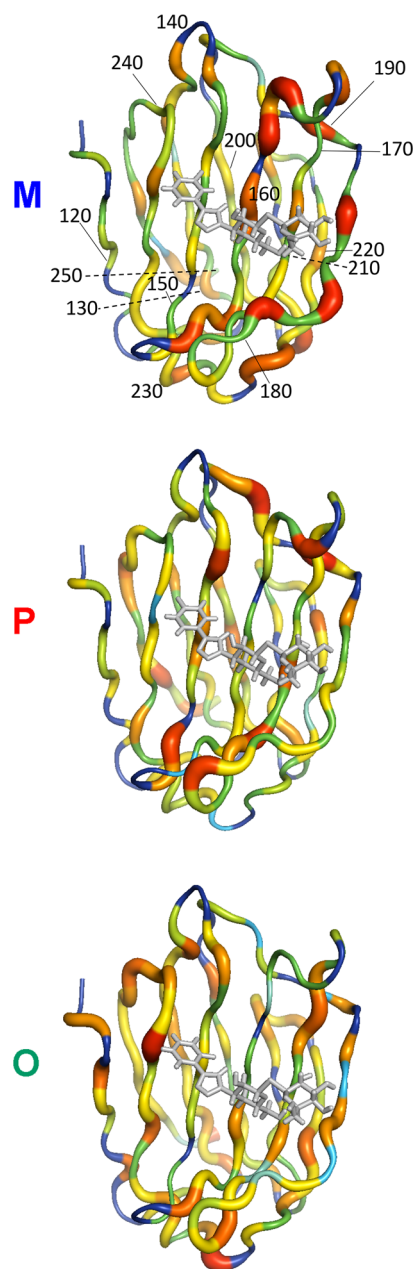
Figure 6. Comparison of the backbone order parameters for the three ligand−galectin-3C complexes. (A) Backbone order parameters for M− (blue), P− (red), and O−galectin-3C (green). (B) Backbone intercomplex ΔO^2 values. Error bars are removed to improve clarity, but a representative error bar, corresponding to the mean error, is placed in the upper left corner. The same 116 residues are shown for each complex. The black and gray bars have the same meaning as in Figure 5.

the backbone fluctuations. Further, residues Thr175, Asn179, Trp181, Arg186, and Ser237 also show differences in O^2. Trp181 and Arg186 contact the ligand with their side chains, while Thr175 forms a hydrogen bond with Gly182. The carbonyl of Ser237 is involved in a fluorine−amide interaction in the P complex, which shows the highest O^2, whereas O^2 is lowest in the O complex for this residue. The largest differences are observed for Arg183, Phe190, and Glu193, none of which are located within 6 Å from the ligand. All three backbone amides lack hydrogen bonding partners and appear in nonregular secondary structure elements. Despite being located in a more remote β-turn, residues 150−154 also show minor differences that again might arise from variations in ligand interactions relayed from the binding site via a network of side chains, see further below.

The intercomplex differences in backbone order parameters are mapped onto the structures in Figure 7, where red, wide tubes represent residues that are more flexible ($\Delta O^2 \leq -0.025$) in the given complex compared to the other two complexes, while cyan, thin tubes indicate residues that are less flexible ($\Delta O^2 \geq 0.025$). Residues that have greater flexibility in M−galectin-3C than in the other two complexes are found very close to the binding site (<3 Å away), see Figure 7. In P−galectin-3C, residues belonging to the corresponding category are located somewhat further away (<5 Å) from the binding site (Figure 7). In addition, the M and P complexes both have additional residues with $\Delta O^2 \leq -0.02$ (wide orange or red tubes in Figure 7) located within 11 Å of the binding site. In


Figure 7. Intercomplex comparison of backbone order parameters. ΔO^2 is color- and width-coded onto a tube representation of the protein backbone. Thin cyan tubes have $\Delta O^2 \geq 0.025$ (indicating less flexibility), followed by green, yellow, orange, and the wide red tubes, which have $\Delta O^2 \leq -0.025$ (indicating more flexibility). Dark blue indicates residues for which data are missing. The ligand is shown in gray. For reference, every 10th residue is marked in the M complex.

O−galectin-3C, the only residues with $\Delta O^2 \leq -0.02$ are Ser237 in the binding site and Lys227, which is located 18 Å from the ligand (Figure 7).

Differences in Side-Chain Order Parameters

The side chains are naturally more flexible than the backbone, as reflected by their generally lower order parameters. The order parameters are listed in Table S3, and Figure S9 shows histograms of the O^2 distributions for each complex. The mean values and standard error of the means are <O^2> = 0.675 ± 0.022, 0.684 ± 0.022, and 0.670 ± 0.021 ($N = 73$) for M−, P−, and O−galectin-3C, respectively. As in the case of the
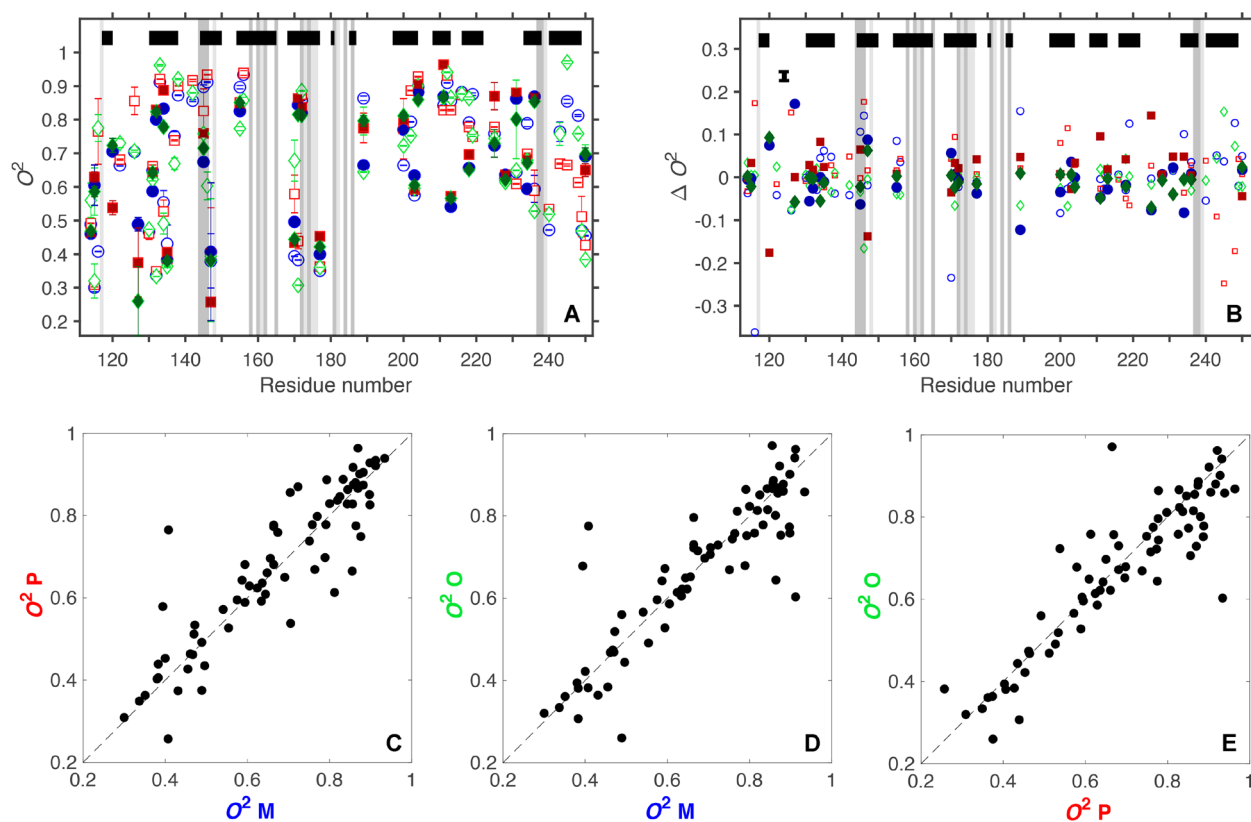
Figure 8. Comparison of methyl-axis order parameters for the various ligand−galectin-3C complexes. (A) O− (green diamonds), M− (blue circles), and P−galectin-3C (red squares) order parameters, reported for 73 methyl groups. Open markers represent residues with a single methyl group (Ala, Thr, and Met) and methyl groups Cγ1 (Val), Cδ1 (Leu), or Cδ1 (Ile), whereas filled markers refer to Cγ2 (Val), Cδ2 (Leu), or Cγ2 (Ile). Error bars indicate 1 SD of (B) methyl ΔO^2 values representing intercomplex comparisons. For clarity, individual error bars are not shown, but an error bar representing the mean error is placed in the upper left-hand corner. (C−E) Correlation plots of the data presented in panel (A); the dotted lines have a slope of 1.

backbone order parameters, one-way ANOVA does not reject the null hypothesis of equal means. Figure 8 presents an overview of the residue-specific values of O^2 as well as the intercomplex differences, ΔO^2, while Figure S11 shows the pairwise comparisons.

We investigated the detailed differences observed for individual residues between the complexes. There are only 5 methyl-containing residues among the 28 residues within 6 Å of the ligand, where the greatest variations among complexes in backbone order parameters are seen. Side-chain ^{2}H relaxation data are available for four of these residues, namely Ile145, Ala146, Leu147, and Val172. Among these, Ala146 shows the greatest variation with O^2 values of 0.92, 0.93, and 0.60 for the M, P, and O complexes, respectively. Interestingly, Ala146 interacts directly with the fluorinated phenyl ring of the ligand. The higher flexibility of Ala146 in the O complex is in agreement with the ensemble-refined crystal structure and the ^{19}F relaxation data, which both indicate significantly greater fluctuations of the fluorophenyl ring in the O complex than in the other two complexes (cf., Figure 4). Clearly, the Ala146 methyl group senses the fluctuations of the ring. Ile145 Cγ1 also shows a lower O^2 of 0.76 in the O complex than in the other two complexes (0.90 and 0.83 for M and P). Ile145 Cδ2 shows O^2 values of 0.67, 0.76, and 0.72 for the M, P, and O complexes. Leu147 does not show any significant differences in O^2, and Val172 shows only very minor differences.

Outside of the immediate binding site, Val116 Cγ1 also shows significant differences with O^2 of 0.41, 0.77, and 0.78 for

the M, P, and O complexes (data are not available for Val116 Cγ2 due to spectral overlap). Ala146 and Val116 are located on either side of Ser237, the backbone of which responds to differences in the fluorine position (as described above). Further, Val155 Cγ1 and Ala156 show minor ΔO^2, as was also observed for the backbone of residues 150−154. The side chain of Val225, which packs against Ala156, also exhibits a difference between complexes with Cγ2 having a higher O^2 for the P complex. Leu120, Leu234, and Ala245 form a cluster of interacting side chains that all show minor ΔO^2. Both Leu120 Cδ2 and Ala245 have the lowest O^2 in the P complex, but these results cannot be explained by inspecting the crystal structures. Further away from the binding site, Val126 Cγ1 shows the opposite behavior compared with Leu120 and Ala 245. Finally, Val170 Cγ1, which packs against V172, has O^2 values of 0.39, 0.58, and 0.68 for the M, P, and O complexes, respectively. Similar to previous results for a number of galectin-3C complexes,[5,8] differences in order parameters do not generally correspond to obvious structural differences between the complexes. This finding reconfirms that subtle but statistically significant differences in dynamics can arise among structurally similar complexes.

Conformational Entropy Differences Determined by NMR

The nonlinear dependence of entropy on the backbone order parameter, together with residue-specific variations in ΔO^2, lead to significant differences in entropy between complexes, despite $\langle O^2 \rangle$ not showing significant variation among M−,

P−, and O−galectin-3C. We estimated the differences between complexes in their conformational entropy contributions to ligand binding, using eqs 1−3 (Table 2). M− and P−galectin-

Table 2. Intercomplex Differences in Protein Conformational Entropy Estimated from NMR Order Parameters

complex	$-T\Delta\Delta S_{conf}$ (kJ mol^{-1})		
	bb[a]	sc[b]	bb + sc
M	−5.7 ± 0.6	−0.5 ± 2	−6 ± 2
P	−6.0 ± 0.8	4 ± 2	−2 ± 2
O	11.7 ± 0.8	−4 ± 2	8 ± 2

[a]Backbone: based on 116 backbone NH groups, errors are given as ±1 SD. [b]Side chains: based on 73 methyl groups, errors are given as ±1 SD.

3C have almost identical conformational backbone entropy, yielding intercomplex differences of $-T\Delta\Delta S_{bb,M} = -5.7 \pm 0.6$ kJ/mol and $-T\Delta\Delta S_{bb,P} = -6.0 \pm 0.8$ kJ/mol, whereas O−galectin-3C has a lower backbone entropy yielding $-T\Delta\Delta S_{bb,O} = 11.7 \pm 0.8$ kJ/mol. Hence, the backbone conformational entropy favors the binding of M and P, whereas the binding of O is disfavored in the intercomplex comparison. The side-chain entropy differences are less pronounced and have greater uncertainty: $-T\Delta\Delta S_{sc,M} = -0.4 \pm 2$ kJ/mol, $-T\Delta\Delta S_{sc,P} = 4 \pm 2$ kJ/mol, and $-T\Delta\Delta S_{sc,O} = -4 \pm 2$ kJ/mol. The empirically calibrated "entropy meter", proposed by Wand and co-workers,[95] is based on the difference between complexes in their average order parameter of the methyl axes and consequently does not capture the minor effects observed here, because $<O^2>$ is not significantly different between the M, P, and O complexes.

Summing up the conformational entropy of the backbone and the side chains results in intercomplex differences: $-T\Delta\Delta S_{bb+sc,M} = -6 \pm 2$ kJ/mol, $-T\Delta\Delta S_{bb+sc,P} = -2 \pm 2$ kJ/mol, and $-T\Delta\Delta S_{bb+sc,O} = 8 \pm 2$ kJ/mol. The differences between M and P are not significant, while O differs significantly from the other two. This suggests that the conformational entropy of the backbone and methyl-bearing side chains determined by NMR favors binding of M and, to a minor extent, also P, whereas it disfavors binding of O.

Conformational Fluctuations and Entropy Differences Determined by MD

MD simulations can provide information on the total conformational entropy of the protein and ligand. Therefore, we performed MD simulations to complement the entropies estimated from the NMR experiments, which are limited to sampling the backbone and methyl-bearing side chains.

The coordinate RMSFs from the MD simulations provide a first overview of differences between the complexes. The MD simulations confirm the results from the ensemble-refined crystal structures and the NMR data, showing that the RMSF is largest for the O ligand in the complex (1.16 ± 0.08 Å) and smallest for P (0.90 ± 0.09 Å). However, the RMSF for the protein barely differs between complexes, where P−galectin-3C has 0.84 ± 0.01 Å compared to 0.85−0.86 ± 0.01 Å for the M and O complexes.

To further validate the simulations against experimental data, we calculated order parameters for the backbone and compared with those measured by NMR. The simulated order parameters reproduce the experimental ones with a pairwise

RMSD between 0.043 and 0.056, which is in line with previous results.[4,8,96] The simulated order parameters are modestly higher than the experimental parameters on average, with $<O^2>_{MD} - <O^2>_{NMR} = 0.006, 0.013,$ and 0.007 for the M, P, and O complexes, respectively. However, there are clear discrepancies in the details (Figure S12). Clearly, it is a great challenge for MD simulations to reproduce the variations in dynamics based on the minor differences in structure among the three complexes.
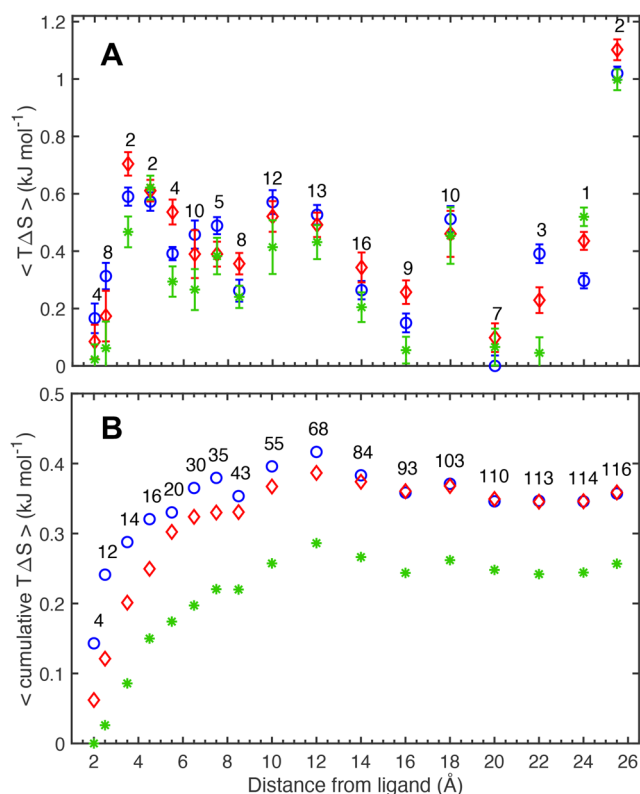
We estimated the conformational entropy from dihedral angle distributions in the MD trajectories. The MD simulations do not yield any clearly significant difference between the three complexes in the conformational entropy of the protein (Table S4) but suggest that it is slightly higher in the M complex, in line with the NMR results. However, the MD simulations do indicate a significant difference in the conformational entropy of the ligands within the complexes, where O deviates from M and P (Table S4). The ligand contributions to the total entropy change are 16.1 ± 0.4 kJ/mol for O, 19.6 ± 0.4 kJ/mol for M, and 20.4 ± 0.3 kJ/mol for P. The less unfavorable entropy contribution from the O ligand primarily reflects a higher degree of flexibility in the bound state, which agrees with the ensemble-refined crystal structures (Figure 4), the ^{19}F relaxation data (Figure S8), and the lower order parameters of methyl-bearing side chains located in direct proximity to the ligand, mentioned above. By contrast, the conformational entropies of the free ligands agree within 1 kJ/mol (Table S5). Similar, but less reliable, results are obtained by molecular mechanics with generalized Born and surface area solvation (MM/GBSA) and interaction entropy calculations (Table S6); see the Supporting Information for details.

Radial Distribution of Conformational Entropy around the Binding Site

Although $<O^2>$ for the backbone is close to identical among the three complexes, the radial distribution of the backbone entropy (Figure 9) reveals differences in mobility for residues close to the binding site. Table S7 lists which residues belong to each shell. The NMR-based entropies reveal that the four residues of the innermost shell, His158, Arg162, Asn174, and Glu184, are among the most rigid in all complexes. These residues also form hydrogen bonds with the ligand (Figure 3). The next shell, 2−3 Å from the ligand, includes eight residues that are slightly less rigid. However, the following two shells (3−4 Å and 4−5 Å) comprise two residues each, whose flexibility is much higher than that for residues in the two innermost shells. The general pattern in Figure 9 reflects a dependence of secondary structure elements, with β-strands having low entropy, and loops and coils contributing to higher entropy; outside of the innermost shells, the proportion of residues in loops and coils is higher.

The distance dependence of the backbone entropy mirrors the greater backbone fluctuations in the binding sites of M− and P−galectin-3C compared to O−galectin-3C (cf., Figure 7). Since many residues have similar O^2, the differences between the cumulative entropy curves are dominated by just a few residues (Figure 9B). In the innermost shell, the main difference is attributed to Arg162, with an O^2 of 0.83, 0.85, and 0.87 in M−, P−, and O−galectin-3C. In the second shell (2−3 Å), Trp181, Arg186, Asn160, and Val172 show similar trends with O^2 progressively increasing in the order M−, P−, and O−galectin-3C.

Figure 10. Differences in solvation around the binding site. Regions with a higher density of water molecules relative to bulk water (8 times the bulk water density) are represented as blue, red, and green meshes for M−, P−, and O−galectin-3C, respectively. Differences are observed nearby the side chain of Arg144, where water sites 1 and 2 are occupied in the M and P complexes, whereas in the O complex, only one water molecule is observed in site 3 because of the slightly different orientation of the Arg144 guanidino group.

Figure 9. Radial distribution of the backbone conformational entropy in concentric shells around the binding site. The average backbone conformational entropy per residue $T\Delta S$, estimated from NMR order parameters, is plotted as a function of the closest distance between any atom in the protein residue and the ligand. Entropies are referenced to the lowest entropy of any shell. (A) Average backbone conformational entropy per residue in each shell. (B) Cumulative conformational entropy per residue. Blue circles, M−galectin-3C; red diamonds, P−galectin-3C; green stars, O−galectin-3C. The number of residues included in each shell is stated above the symbols. The width of the shell is 1 Å between 2 and 9 Å, and 2 Å between 9 and 27 Å. The distance is evaluated at the midpoint of the shell. For clarity, error bars are only reported in panel A. Table S5 lists the residues included in each shell.

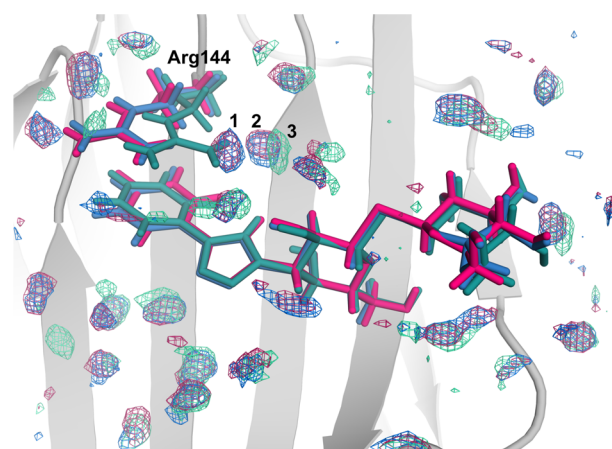## Water Binding Sites and Solvation Thermodynamics

We identified water sites around the ligand-binding site and determined the solvent thermodynamics using GIST calculations based on MD simulations, in which the solute was restrained to the crystal structure.[77] The results for the O, M, and P complexes reveal differences in their hydration thermodynamics, relative to bulk water (Table 3).

M− and P−galectin-3C have similar solvation entropies, $-T\Delta S_{water}$, of 432.4 ± 0.3 and 428.4 ± 0.3 kJ/mol, respectively, whereas O−galectin-3C differs by having a less unfavorable solvation entropy of 402.1 ± 0.3 kJ/mol. Figure 10

Table 3. Solvation Entropy from GIST Calculations[a]

complex	$-T\Delta S_{rot}$	$-T\Delta S_{trans}$	$-T\Delta S_{water}$
M−galectin-3C	249.1 ± 0.1	183.3 ± 0.2	432.4 ± 0.3
P−galectin-3C	245.3 ± 0.2	183.0 ± 0.2	428.4 ± 0.3
O−galectin-3C	235.6 ± 0.2	166.5 ± 0.2	402.1 ± 0.3

[a]Units of kJ/mol; $-T\Delta S_{rot}$, solvent rotational entropy; $-T\Delta S_{trans}$, solvent translational entropy; $-T\Delta S_{water} = -T\Delta S_{rot} - T\Delta S_{trans}$. Errors are given as ±1 SD.

shows an overview of the identified water sites, i.e., regions with higher density than bulk water, surrounding the bound ligands. Most water sites are highly similar among the three complexes, but a few distinct differences are observed. The less favorable solvation entropy of the M and P complexes primarily arises from the difference in the average conformation of Arg144 (cf., Figure 3B), which allows two water molecules to bridge between the protein and the ligand (positions 1 and 2, Figure 10), whereas in the O complex, the Arg144 side chain partly occupies this region of the binding site, leaving room for only one water molecule (position 3, Figure 10). These water molecules are observed also in the crystal structures (two for M and P but one for O, at positions within 0.5−0.6 Å of the centers of the GIST densities), although water molecule 2 is replaced by a chloride ion in the structures of the M and P complexes. The MD simulations indicate that these water-binding sites are highly dynamic with residence times for individual water molecules of only ~0.3 ns for position 1 and 1−3 ns for positions 2 and 3. These residence times are in perfect agreement with experimental results obtained on lactose-bound galectin-3C.[13]

Solvation free energies of the free ligands in water solution, obtained by COSMO-RS calculations, are identical within errors (Table S5). Moreover, the total solvation entropies, calculated by GIST, for the free ligands agree within 0.7 kJ/mol (Table S5). Thus, the differences in total solvation entropy of binding among the three systems is governed by the observed differences in hydration of the complexes.

■ CONCLUSIONS

In this study, we compared the thermodynamic signature of binding three congeneric ligands, differing only in the position of a fluorine atom on a phenyl ring, to the carbohydrate binding domain of galectin-3. We used a multipronged approach to investigate in great detail the statistical thermodynamic driving forces for selectivity in binding the three ligands. Specifically, we break down the total entropy of binding into individual contributions from the different

degrees of freedom, viz., protein, ligand, and solvent, as summarized in Figure 11.
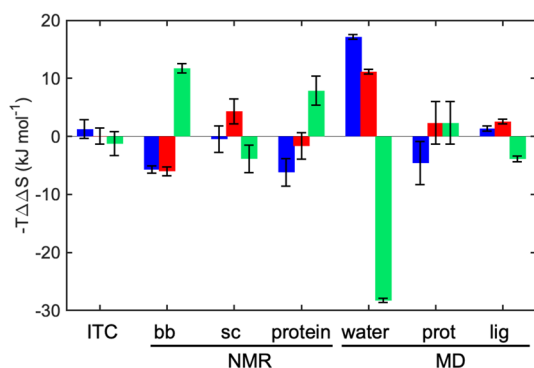
Figure 11. Entropy contributions to the differential binding of ligands M (blue), P (red), and O (green) to galectin-3C. The colored bars represent intercomplex differences in entropy; hence, a negative value corresponds to a favorable entropic contribution to binding for the specified complex, relative to the other two complexes. ITC reports the total entropy of binding, Table 1. NMR reports estimates of the conformational entropy of the backbone (bb) and methyl-bearing side chains (sc), Table 2. MD reports the conformational entropy of the protein (prot) and ligand (lig), Table S4, and the solvation entropy determined by GIST (water), Tables 3 and S5. Error bars indicate ± 1 SD.

The minor structural difference between the ligands renders their chemical potentials in the free state highly similar, as confirmed by calculations of conformational entropies, GIST solvation entropies, and COSMO-RS solvation free energies; the results do not differ, within the calculated uncertainties, between the three ligands (Table S5). Hence, the difference in binding thermodynamics can be related primarily to differences between the ligand–galectin-3C complexes. As might be expected from their structural similarity, the three ligands bind to galectin-3C with similar free energy. However, the O complex deviates from the other two in having less favorable binding enthalpy and less unfavorable binding entropy. The difference in binding enthalpy can ostensibly be rationalized in a straightforward manner by a reduced number of interactions involving the fluorine atom and surrounding protein in the O complex, compared to the M and P complexes. The difference in total entropy of binding arises as the net sum of opposing contributions from different components. The experimental results from NMR relaxation and ensemble-refined X-ray diffraction data suggest that the backbone of galectin-3C has reduced entropy when bound to O, especially for residues close to the binding site. (The MD-derived estimates of conformational entropy have large uncertainties and do not yield any significant differences between the three complexes.) One can speculate that the less favorable interactions of the *ortho*-fluorine atom with the protein create strain and restrict the flexibility of the protein backbone. However, this unfavorable entropic contribution to binding is apparently offset in the O complex, but not in the other two complexes, by increased entropy of the ligand and protein side-chains in immediate proximity with the ligand. The higher degree of flexibility of O can intuitively be associated with the fewer favorable protein–ligand interactions in this complex. Finally, the three complexes differ markedly in solvation entropy: A large difference in water entropy for the protein–ligand complexes favors the O complex.

Thus, the present results indicate an interesting case of *entropy–entropy compensation* between the conformational degrees of freedom of the protein, which disfavor binding of O over M or P, and the ligand and solvation entropies, which favor binding of O. These results highlight intricate details of binding thermodynamics and the importance of accounting for all components of the system in order to advance prospects for rational drug design.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.0c00094.

Synthesis of ligands M and P, crystallography data collection and refinement statistics, electron density image for ligands, assigned $^{1}H-^{15}N$ and $^{1}H-^{13}C$ HSQC spectra, intercomplex chemical shift perturbations, ^{15}N R_1, R_2, and {$^{1}H-^{15}N$}-NOE data together with back-calculated relaxation rates based on optimized model-free parameters, ^{19}F R_2 data, histograms of backbone and methyl O^2 distributions, pairwise comparison of NMR order parameters for the complexes, comparison of backbone order parameters determined by MD and NMR, tables of backbone and methyl order parameters, table of conformational entropy differences between free and bound states determined by MD, table of conformational entropy and solvation free energy of the free ligands, table of results from MM/GBSA and interaction entropy (IE) calculations, methods section describing the MM/GBSA and IE calculations, table listing residues in each shell of the radial distribution of conformational entropy (PDF)

Accession Codes

The crystal structures and the diffraction data have been deposited in the Protein Data Bank under accession IDs 6RZG (M complex), 6RZH (P complex), and 6RZF (O complex). The chemical shift assignments, relaxation rate constants, and order parameters have been deposited in the Biological Magnetic Resonance Bank (BMRB) under accession codes 50283 (M complex), 50284 (P complex), and 50285 (O complex).

■ AUTHOR INFORMATION

Corresponding Authors

Mikael Akke − *Biophysical Chemistry, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden;* ⓞ orcid.org/0000-0002-2395-825X; Email: mikael.akke@bpc.lu.se

Ulf Ryde − *Theoretical Chemistry, Department of Chemistry, Lund University, 221 00 Lund, Sweden;* ⓞ orcid.org/0000-0001-7653-8489; Email: ulf.ryde@teokem.lu.se

Authors

Johan Wallerstein − *Biophysical Chemistry, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden*

Vilhelm Ekberg − *Theoretical Chemistry, Department of Chemistry, Lund University, 221 00 Lund, Sweden;* ⓞ orcid.org/0000-0002-2813-3564

Majda Misini Ignjatović − *Theoretical Chemistry, Department of Chemistry, Lund University, 221 00 Lund, Sweden*

Rohit Kumar − *Biochemistry and Structural Biology, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden*

Octav Caldararu − *Theoretical Chemistry, Department of Chemistry, Lund University, 221 00 Lund, Sweden*

Kristoffer Peterson − *Centre for Analysis and Synthesis, Department of Chemistry, Lund University, 221 00 Lund, Sweden*

Sven Wernersson − *Biophysical Chemistry, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden*

Ulrika Brath − *The Swedish NMR Center, University of Gothenburg, 405 30 Gothenburg, Sweden*

Hakon Leffler − *Microbiology, Immunology, and Glycobiology, Department of Experimental Medicine, Lund University, 221 00 Lund, Sweden;* orcid.org/0000-0003-4482-8945

Esko Oksanen − *European Spallation Source ESS ERIC, 225 92 Lund, Sweden*

Derek T. Logan − *Biochemistry and Structural Biology, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden;* orcid.org/0000-0002-0098-8560

Ulf J. Nilsson − *Centre for Analysis and Synthesis, Department of Chemistry, Lund University, 221 00 Lund, Sweden;* orcid.org/0000-0001-5815-9522

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacsau.0c00094

Author Contributions

Notes

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Akke, M.; Brüschweiler, R.; Palmer, A. G. NMR Order Parameters and Free Energy: An Analytical Approach and Its Application to Cooperative Ca²⁺ Binding by Calbindin D9k. *J. Am. Chem. Soc.* **1993**, *115*, 9832−9833.

(2) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The Role of Conformational Entropy in Molecular Recognition by Calmodulin. *Nat. Chem. Biol.* **2010**, *6*, 352−358.

(3) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational Entropy in Molecular Recognition by Proteins. *Nature* **2007**, *448*, 325−330.

(4) Diehl, C.; Genheden, S.; Modig, K.; Ryde, U.; Akke, M. Conformational Entropy Changes upon Lactose Binding to the Carbohydrate Recognition Domain of Galectin-3. *J. Biomol. NMR* **2009**, *45*, 157−169.

(5) Diehl, C.; Engström, O.; Delaine, T.; Håkansson, M.; Genheden, S.; Modig, K.; Leffler, H.; Ryde, U.; Nilsson, U. J.; Akke, M. Protein Flexibility and Conformational Entropy in Ligand Design Targeting the Carbohydrate Recognition Domain of Galectin-3. *J. Am. Chem. Soc.* **2010**, *132*, 14577−14589.

(6) Tzeng, S. R.; Kalodimos, C. G. Protein Activity Regulation by Conformational Entropy. *Nature* **2012**, *488*, 236−240.

(7) Gill, M. L.; Byrd, R. A.; Palmer, A. G., III Dynamics of GCN4 Facilitate DNA Interaction: A Model-Free Analysis of an Intrinsically Disordered Region. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5839−5849.

(8) Verteramo, M. L.; Stenström, O.; Misini, M.; Ignjatović, I.; Caldararu, O.; Olsson, M. A.; Manzoni, F.; Leffler, H.; Oksanen, E.; Logan, D. T.; et al. Interplay between Conformational Entropy and Solvation Entropy in Protein-Ligand Binding. *J. Am. Chem. Soc.* **2019**, *141*, 2012−2026.

(9) Syme, N. R.; Dennis, C.; Bronowska, A.; Paesen, G. C.; Homans, S. W. Comparison of Entropic Contributions to Binding in a "Hydrophilic" versus "Hydrophobic" Ligand-Protein Interaction. *J. Am. Chem. Soc.* **2010**, *132* (25), 8682−8689.

(10) Haider, K.; Wickstrom, L.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Enthalpic Breakdown of Water Structure on Protein Active-Site Surfaces. *J. Phys. Chem. B* **2016**, *120*, 8743−8756.

(11) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. Entropy-Enthalpy Transduction Caused by Conformational Shifts can Obscure the Forces Driving Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20006−20011.

(12) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808−813.

(13) Saraboji, K.; Håkansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; et al. The Carbohydrate-Binding Site in Galectin-3 Is Preorganized to Recognize a Sugar-like Framework of Oxygens: Ultra-High Resolution Structures and Water Dynamics. *Biochemistry* **2012**, *51*, 296−306.

(14) Manzoni, F.; Wallerstein, J.; Schrader, T. E.; Ostermann, A.; Coates, L.; Akke, M.; Blakeley, M. P.; Oksanen, E.; Logan, D. T. Elucidation of Hydrogen Bonding Patterns in Ligand-Free, Lactose- and Glycerol-Bound Galectin-3C by Neutron Crystallography to Guide Drug Design. *J. Med. Chem.* **2018**, *61*, 4412−4420.

(15) Leffler, H.; Carlsson, S.; Hedlund, M.; Qian, Y. N.; Poirier, F. Introduction to Galectins. *Glycoconjugate J.* **2002**, *19*, 433−440.

(16) Dumic, J.; Dabelic, S.; Flögel, M. Galectin-3: An Open-Ended Story. *Biochim. Biophys. Acta, Gen. Subj.* **2006**, *1760*, 616−635.

(17) Liu, F. T.; Rabinovich, G. A. Galectins: Regulators of Acute and Chronic Inflammation. *Ann. N. Y. Acad. Sci.* **2010**, *1183*, 158−182.

(18) Liu, F.-T.; Rabinovich, G. A. Galectins as Modulators of Tumour Progression. *Nat. Rev. Cancer* **2005**, *5*, 29−41.

(19) Di Lella, S.; Sundblad, V.; Cerliani, J. P.; Guardia, C. M.; Estrin, D. A.; Vasta, G. R.; Rabinovich, G. A. When Galectins Recognize Glycans: From Biochemistry to Physiology and Back Again. *Biochemistry* **2011**, *50*, 7842−7857.

(20) Blanchard, H.; Yu, X.; Collins, P. M.; Bum-Erdene, K. Galectin-3 Inhibitors: A Patent Review (2008-Present). *Expert Opin. Ther. Pat.* **2014**, *24*, 1053−1065.

(21) Delaine, T.; Collins, P.; MacKinnon, A.; Sharma, G.; Stegmayr, J.; Rajput, V. K.; Mandal, S.; Cumpstey, I.; Larumbe, A.; Salameh, B. A.; et al. Galectin-3-Binding Glycomimetics That Strongly Reduce

Bleomycin-Induced Lung Fibrosis and Modulate Intracellular Glycan Recognition. *ChemBioChem* **2016**, *17*, 1759−1770.

(22) Boza-Serrano, A.; Ruiz, R.; Sanchez-Varo, R.; García-Revilla, J.; Yang, Y.; Jimenez-Ferrer, I.; Paulus, A.; Wennström, M.; Vilalta, A.; Allendorf, D.; et al. Galectin-3, a Novel Endogenous TREM2 Ligand, Detrimentally Regulates Inflammatory Response in Alzheimer's Disease. *Acta Neuropathol.* **2019**, *138*, 251−273.

(23) Peterson, K.; Kumar, R.; Stenström, O.; Verma, P.; Verma, P. R.; Håkansson, M.; Kahl-Knutsson, B.; Zetterberg, F.; Leffler, H.; Akke, M.; et al. Systematic Tuning of Fluoro-Galectin-3 Interactions Provides Thiodigalactoside Derivatives with Single-Digit NM Affinity and High Selectivity. *J. Med. Chem.* **2018**, *61*, 1164−1175.

(24) Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis. *Anal. Chem.* **2012**, *84*, 5066−5073.

(25) Brautigam, C. A.; Zhao, H.; Vargas, C.; Keller, S.; Schuck, P. Integration and Global Analysis of Isothermal Titration Calorimetry Data for Studying Macromolecular Interactions. *Nat. Protoc.* **2016**, *11*, 882−894.

(26) Freiburger, L.; Auclair, K.; Mittermaier, A. Global ITC Fitting Methods in Studies of Protein Allostery. *Methods* **2015**, *76*, 149−161.

(27) Cianci, M.; Bourenkov, G.; Pompidor, G.; Karpics, I.; Kallio, J.; Bento, I.; Roessle, M.; Cipriani, F.; Fiedler, S.; Schneider, T. R. P13, the EMBL Macromolecular Crystallography Beamline at the Low-Emittance PETRA III Ring for High- and Low-Energy Phasing with Variable Beam Focusing. *J. Synchrotron Radiat.* **2017**, *24* (1), 323−332.

(28) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.

(29) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204−1214.

(30) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213−221.

(31) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658−674.

(32) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. Electronic Ligand Builder and Optimization Workbench (ELBOW): A Tool for Ligand Coordinate and Restraint Generation. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65* (10), 1074−1080.

(33) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 352−367.

(34) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.

(35) Burnley, B. T.; Afonine, P. V.; Adams, P. D.; Gros, P. Modelling Dynamics in Protein Crystal Structures by Ensemble Refinement. *eLife* **2012**, *1*, No. e00311.

(36) Roe, D. R.; Cheatham, T. E., III PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Synamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(37) Wittekind, M.; Mueller, L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *J. Magn. Reson., Ser. B* **1993**, *101*, 201−205.

(38) Grzesiek, S.; Bax, A. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.* **1992**, *114*, 6291−6293.

(39) Bax, A.; Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Ikura, M.; Kay, L. E. Practical Aspects of Proton-Carbon-Carbon-Proton Three-Dimensional Correlation Spectroscopy of 13C-Labeled Proteins. *J. Magn. Reson.* **1990**, *87*, 620−627.

(40) Bax, A.; Clore, G. M.; Gronenborn, A. M. 1H-1H Correlation via Isotropic Mixing of 13C Magnetization, a New Three-Dimensional Approach for Assigning 1H and 13C Spectra of 13C-Enriched Proteins. *J. Magn. Reson.* **1990**, *88*, 425−431.

(41) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.

(42) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, P.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 687−696.

(43) Ahlner, A.; Carlsson, M.; Jonsson, B. H.; Lundström, P. PINT: A Software for Integration of Peak Volumes and Extraction of Relaxation Rates. *J. Biomol. NMR* **2013**, *56*, 191−202.

(44) Mosteller, F.; Tukey, J. W. *Data Analysis and Regression. A Second Course in Statistics*; Addison-Wesley: Reading, MA, 1977.

(45) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in 13C-Labeled and Fractionally 2H-Enriched Proteins in Solution. *J. Am. Chem. Soc.* **2002**, *124*, 6439−6448.

(46) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes. The Art of Scientific Computing*; Cambridge University Press: Cambridge, 1986.

(47) Halle, B.; Wennerström, H. Interpretation of Magnetic Resonance Data from Water Nuclei in Heterogeneous Systems. *J. Chem. Phys.* **1981**, *75*, 1928−1943.

(48) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. I. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

(49) d'Auvergne, E. J.; Gooley, P. R. The Use of Model Selection in the Model-Free Analysis of Protein Dynamics. *J. Biomol. NMR* **2003**, *25*, 25−39.

(50) d'Auvergne, E. J.; Gooley, P. R. Optimisation of NMR Dynamic Models II. A New Methodology for the Dual Optimisation of the Model-Free Parameters and the Brownian Rotational Diffusion Tensor. *J. Biomol. NMR* **2008**, *40*, 121−133.

(51) d'Auvergne, E. J.; Gooley, P. R. Optimisation of NMR Dynamic Models I. Minimisation Algorithms and Their Performance within the Model-Free and Brownian Rotational Diffusion Spaces. *J. Biomol. NMR* **2008**, *40*, 107−119.

(52) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(53) Mandel, A. M.; Akke, M.; Palmer, A. G., III Backbone Dynamics of Eschericia Coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **1995**, *246*, 144−163.

(54) Ghose, R.; Fushman, D.; Cowburn, D. Determination of the Rotational Diffusion Tensor of Macromolecules in Solution from NMR Relaxation Data with a Combination of Exact and Approximate Methods—Application to the Determination of Interdomain Orientation in Multidomain Proteins. *J. Magn. Reson.* **2001**, *149*, 204−217.

(55) Skrynnikov, N. R.; Millet, O.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions. *J. Am. Chem. Soc.* **2002**, *124*, 6449−6460.

(56) Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, 5th ed.; Brooks/Cole Publishing Company: Monterey, 1999.

(57) Li, D. W.; Brüschweiler, R. A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters. *J. Am. Chem. Soc.* **2009**, *131*, 7226−7227.

(58) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; et al. *AMBER 2015*; University of California, San Francisco, 2015.

(59) Genheden, S.; Diehl, C.; Akke, M.; Ryde, U. Starting-Condition Dependence of Order Parameters Derived from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 2176−2190.

(60) Genheden, S.; Akke, M.; Ryde, U. Conformational Entropies and Order Parameters: Convergence, Reproducibility, and Transferability. *J. Chem. Theory Comput.* **2014**, *10*, 432−438.

(61) Uranga, J.; Mikulskis, P.; Genheden, S.; Ryde, U. Can the Protonation State of Histidine Residues Be Determined from Molecular Dynamics Simulations? *Comput. Theor. Chem.* **2012**, *1000*, 75−84.

(62) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(63) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(64) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(65) Bayly, C. C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(66) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(67) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.

(68) Genheden, S.; Ryde, U. A Comparison of Different Initialization Protocols to Obtain Statistically Independent Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 187−195.

(69) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(70) Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381* (3−4), 512−518.

(71) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(72) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(73) Prompers, J. J.; Brüschweiler, R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2002**, *124*, 4522−4534.

(74) Edholm, O.; Berendsen, H. J. C. Entropy Estimation from Simulations of Non-Diffusive Systems. *Mol. Phys.* **1984**, *51*, 1011−1028.

(75) Trbovic, N.; Cho, J.-H.; Abel, R.; Friesner, R. A.; Rance, M.; Palmer, A. G. Protein Side-Chain Dynamics and Residual Conformational Entropy. *J. Am. Chem. Soc.* **2009**, *131*, 615−622.

(76) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]Uril. *J. Chem. Phys.* **2012**, *137*, 044101.

(77) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10*, 2769−2780.

(78) Klamt, A.; Schüürmann, G. Cosmo - a New Approach To Dielectric Screening in Solvents With Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.

(79) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. COSMO Implementation in TURBOMOLE: Extension of an Efficient Quantum Chemical Code towards Liquid Systems. *Phys. Chem. Chem. Phys.* **2000**, *2* (10), 2187−2193.

(80) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(81) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE J.* **2002**, *48* (2), 369−385.

(82) Eckert, F.; Klamt, A. *COSMOtherm*; COSMOlogic GmbH & Co KG: Leverkusen, Germany, 2010.

(83) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. On the Performance of Continuum Solvation Methods. A Comment on "Universal Approaches to Solvation Modeling. *Acc. Chem. Res.* **2009**, *42*, 489−492.

(84) Reinisch, J.; Klamt, A. Prediction of Free Energies of Hydration with COSMO-RS on the SAMPL4 Data Set. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 169−173.

(85) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38* (6), 3098−3100.

(86) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33* (June), 8822−8824.

(87) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to Kr. *J. Chem. Phys.* **1992**, *97* (4), 2571−2577.

(88) Seetharaman, J.; Kanigsberg, A.; Slaaby, R.; Leffler, H.; Barondes, S. H.; Rini, J. M. X-Ray Crystal Structure of the Human Galectin-3 Carbohydrate Recognition Domain at 2.1-Angstrom Resolution. *J. Biol. Chem.* **1998**, *273* (21), 13047−13052.

(89) Sörme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J. Structural and Thermodynamic Studies on Cation-Pi Interactions in Lectin-Ligand Complexes: High-Affinity Galectin-3 Inhibitors through Fine-Tuning of an Arginine-Arene Interaction. *J. Am. Chem. Soc.* **2005**, *127*, 1737−1743.

(90) Olsen, J. A.; Banner, D. W.; Seiler, P.; Obst Sander, U.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an Enzyme Active Site: Evidence for C-F···C = O Interactions. *Angew. Chem., Int. Ed.* **2003**, *42* (22), 2507−2511.

(91) Zürcher, M.; Diederich, F. Structure-Based Drug Design: Exploring the Proper Filling of Apolar Pockets at Enzyme Active Sites. *J. Org. Chem.* **2008**, *73* (12), 4345−4361.

(92) Pollock, J.; Borkin, D.; Lund, G.; Purohit, T.; Dyguda-Kazimierowicz, E.; Grembecka, J.; Cierpicki, T. Rational Design of Orthogonal Multipolar Interactions with Fluorine in Protein-Ligand Complexes. *J. Med. Chem.* **2015**, *58* (18), 7465−7474.

(93) Kumar, R.; Misini Ignjatović, M.; Peterson, K.; Olsson, M. A.; Leffler, H.; Ryde, U.; Nilsson, U. J.; Logan, D. T. Structure and Energetics of Ligand−Fluorine Interactions with Galectin-3 Backbone and Side-Chain Amides − Insight into Solvation Effects and Multipolar Interactions. *ChemMedChem* **2019**, *14*, 1528−1536.

(94) Caldararu, O.; Kumar, R.; Oksanen, E.; Logan, D. T.; Ryde, U. Are Crystallographic B-Factors Suitable for Calculating Protein Conformational Entropy? *Phys. Chem. Chem. Phys.* **2019**, *21*, 18149−18160.

(95) Caro, J. A.; Harpole, K. W.; Kasinath, V.; Lim, J.; Granja, J.; Valentine, K. G.; Sharp, K. A.; Wand, A. J. Entropy in Molecular Recognition by Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 6563−6568.

(96) Showalter, S. A.; Bruschweiler, R. Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field. *J. Chem. Theory Comput.* **2007**, *3*, 961−975.