Interplay between Conformational Entropy and Solvation Entropy in Protein-Ligand Binding

Verteramo, Maria Luisa; Stenström, Olof; Ignjatović, Majda Misini; Caldararu, Octav; Olsson, Martin A.; Manzoni, Francesco; Leffler, Hakon; Oksanen, Esko; Logan, Derek T.; Nilsson, Ulf J.; Ryde, Ulf; Akke, Mikael

Published in:
Journal of the American Chemical Society

DOI:
10.1021/jacs.8b11099

2019

Document Version:
Publisher's PDF, also known as Version of record

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Cite This: J. Am. Chem. Soc. 2019, 141, 2012−2026

pubs.acs.org/JACS

Interplay between Conformational Entropy and Solvation Entropy in Protein−Ligand Binding

Maria Luisa Verteramo,[†,¶] Olof Stenström,[‡,¶] Majda Misini Ignjatović,[§,¶] Octav Caldararu,[§,¶] Martin A. Olsson,[§,¶] Francesco Manzoni,[∥,¶,▽] Hakon Leffler,[⊥] Esko Oksanen,[#] Derek T. Logan,[∥] Ulf J. Nilsson,[†] Ulf Ryde,[§] and Mikael Akke*[,‡]

[†]Centre for Analysis and Synthesis, Department of Chemistry, Lund University, 221 00 Lund, Sweden

[‡]Biophysical Chemistry, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden

[§]Theoretical Chemistry, Department of Chemistry, Lund University, 221 00 Lund, Sweden

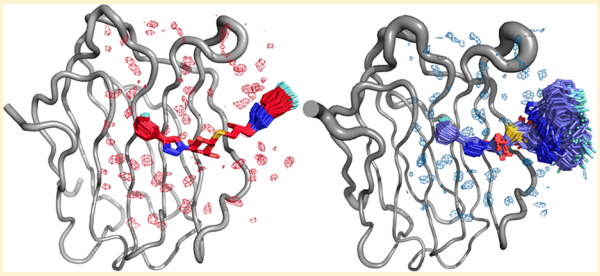
[∥]Biochemistry and Structural Biology, Center for Molecular Protein Science, Department of Chemistry, Lund University, 221 00 Lund, Sweden

[⊥]Microbiology, Immunology, and Glycobiology, Department of Laboratory Medicine, Lund University, 221 00 Lund, Sweden

[#]European Spallation Source ESS ERIC, 225 92 Lund, Sweden

Supporting Information

ABSTRACT: Understanding the driving forces underlying molecular recognition is of fundamental importance in chemistry and biology. The challenge is to unravel the binding thermodynamics into separate contributions and to interpret these in molecular terms. Entropic contributions to the free energy of binding are particularly difficult to assess in this regard. Here we pinpoint the molecular determinants underlying differences in ligand affinity to the carbohydrate recognition domain of galectin-3, using a combination of isothermal titration calorimetry, X-ray crystallography, NMR relaxation, and molecular dynamics simulations followed by conformational entropy and grid inhomogeneous solvation theory (GIST) analyses. Using a pair of diastereomeric ligands that have essentially identical chemical potential in the unbound state, we reduced the problem of dissecting the thermodynamics to a comparison of the two protein−ligand complexes. While the free energies of binding are nearly equal for the R and S diastereomers, greater differences are observed for the enthalpy and entropy, which consequently exhibit compensatory behavior, $\Delta\Delta H^{\circ}(\mathrm{R}-\mathrm{S}) = -5 \pm 1$ kJ/mol and $-T\Delta\Delta S^{\circ}(\mathrm{R}-\mathrm{S}) = 3 \pm 1$ kJ/mol. NMR relaxation experiments and molecular dynamics simulations indicate that the protein in complex with the S-stereoisomer has greater conformational entropy than in the R-complex. GIST calculations reveal additional, but smaller, contributions from solvation entropy, again in favor of the S-complex. Thus, conformational entropy apparently dominates over solvation entropy in dictating the difference in the overall entropy of binding. This case highlights an interplay between conformational entropy and solvation entropy, pointing to both opportunities and challenges in drug design.



INTRODUCTION

Molecular recognition is fundamental to biology in that it governs signaling within and between cells, with prominent examples provided by the immune system, hormonal control of distant organs in higher organisms, and specificity of enzyme reactions. Modern medicine is to a large extent based on the possibility to interfere with and control molecular recognition by the design of synthetic ligands or effectors that bind to a specific protein in a given signaling pathway. Drug design aims to generate such protein ligands that have high affinity and specificity for the target. Despite the enormous resources contributed by industry and academia over the past several decades, rational structure-based design of ligands by computational approaches remains extremely challenging. One reason is that the free energy of binding is in most cases a small difference between large numbers arising from the different interactions between the protein, ligand, other solutes, and solvent molecules. In addition, the energy terms are strongly dependent on the detailed molecular conformations, due to their sharp dependence on interatomic distances and orientations. Furthermore, entropic contributions can be significant because proteins have many degrees of freedom, are generally flexible, and consequently populate a wide range of conformations. Recent work has indeed highlighted the role of protein conformational entropy in ligand binding,[1−8] as well as the highly heterogeneous response of water molecules around binding sites[9−12] and ligands.[13]

Received: October 15, 2018
Published: January 8, 2019

We have identified the carbohydrate recognition domain (CRD) of galectin-3 (denoted galectin-3C) as an interesting system for investigating the role of conformational entropy[4,5] and solvation in ligand binding.[14] Galectin-3 has a relatively solvent-accessible binding site placed in a shallow groove across one of the two β-sheets, with water molecules forming an integral part of the binding site by bridging between the ligand and protein.[14,15] Galectin-3 is a member of the galectin family of mammalian lectins, defined by the CRD with its conserved sequence motif that confers affinity for β-galactoside containing glycans.[16,17] Galectins play important roles in cell growth, cell differentiation, cell cycle regulation, signaling, and apoptosis, which target them for pharmaceutical intervention to treat inflammation and cancer,[16−19] with specific examples reported for galectin-3.[20−22]

Here, we report a comparative analysis of galectin-3C in complex with two diastereomeric ligands. The advantage of this approach is that the differences in binding thermodynamics are dominated by the properties of the two ligand−protein complexes, while the unbound diastereomers have nearly identical chemical potential in the unbound state and thus cancel in the comparative analysis. We used a combination of experimental and computational approaches including isothermal titration calorimetry (ITC), competitive fluorescence polarization assay, X-ray crystallography, NMR spectroscopy including ^{15}N backbone and ^{2}H side-chain methyl relaxation, and molecular dynamics (MD) simulations followed by conformational entropy and grid inhomogeneous solvation theory (GIST) calculations. Following on our previous work,[4,5] we focus on entropic contributions to the free energy of binding. In the present work, we extend the analysis to include not only conformational entropy of the protein and ligand but also solvent entropy. Our results show that conformational entropy makes a greater contribution than solvent entropy to the difference between ligands in overall entropy of binding, and further highlight an interplay between conformational entropy and solvent entropy in contributing toward ligand binding affinity and specificity.

MATERIALS AND METHODS

Ligand Synthesis. The two diastereomeric compounds (2*R*)- and (2*S*)-2-hydroxy-3-(4-(3-fluorophenyl)-1*H*-1,2,3-triazol-1-yl)-propyl 2,4,6-tri-*O*-acetyl-3-deoxy-3-(4-(3-fluorophenyl)-1*H*-1,2,3-triazol-1-yl)-1-thio-β-D-galactopyranoside (denoted ligands R and S, respectively) were synthesized from triisoprolylsilyl 2,4,6-tri-*O*-acetyl-3-azido-3-deoxy-1-thio-β-D-galactopyranoside[23] and *R*- and *S*-glycidyl nosylate. Reaction conditions, physical data, and purity data are given in the Supporting Information.

Protein Expression and Purification. Galectin-3C was expressed and purified by the Lund Protein Production Platform (LP3) at Lund University following published protocols,[4,5] yielding a protein stock solution of 9.2 mg/mL in ME-PBS buffer (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 β-mercaptoethanol), and 150 mM lactose. The protein stock solution was stored at 278 K.

Isothermal Titration Calorimetry. Galectin-3C samples were prepared by extensive dialysis against 5 mM 4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid (HEPES) buffer to remove all lactose, followed by centrifugation at 14 000 rpm to remove any aggregates. Both ligands were dissolved in stock solutions of dimethyl sulfoxide (DMSO) to prepare stock solutions 20.7 mM and 20.3 mM for R and S, respectively, and sonicated immediately prior to experiments. Isothermal titration calorimetry (ITC) experiments were performed on MicroCal iTC200 and MicroCal PEAQ−ITC instruments (Malvern) at a temperature of 301 K by titrating the protein at a concentration of 0.22 mM into the cell containing the ligand at a concentration of 0.02 mM. The DMSO concentrations in the cell and the syringe were carefully matched to minimize the heat of dilution, and were the same for the two ligands. Five replicate experiments were performed for each complex. Peak integration was done using NITPIC.[24] A single-site binding model was fitted simultaneously to the 5 titrations curves to yield the binding enthalpy (ΔH), fraction of binding-competent protein (n), and dissociation constant (K_d), using in-house MATLAB routines with Monte Carlo error estimation.[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 that accounts for heats of mixing, and $Q(i)$ is the heat function 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 subsequently determined using the relationships $\Delta G^\circ = RT\ln(K_d)$ and $-T\Delta S^\circ = \Delta G^\circ - \Delta H^\circ$.

Competitive Fluorescence Polarization Experiments. The binding affinity between galectin-3C and each ligand was determined using competitive fluorescence polarization experiments described previously,[22] using the fluorescent probe 3,3′-dideoxy-3-[4-(fluorescein-5-yl-carbonylaminomethyl)-1*H*-1,2,3-triazol-1-yl]-3′-(3,5-dimethoxybenzamido)- 1,1′-sulfanediyl-di-β-D-galactopyranoside.[27]

X-ray Crystallography. Crystals of lactose-bound galectin-3C were grown with the hanging drop method in NeXtal plates and with the following reservoir condition: 28% (w/v) PEG 4000, Tris-HCl pH 7.5, 0.4 M NaSCN, 15 mM β-mercaptoethanol. The drop volume was 2 μL and the protein solution:reservoir ratio was varied between 0.5:1, 1:1, and 2:1. The crystals were then moved to drops containing the same reservoir with the addition of 10 mM of the ligand (R or S), made from a 100 mM stock solution in neat DMSO. Soaking lasted for 7 h for the R diastereomer and 20 h for S. Before data collection, crystals were placed for a couple of seconds in a drop containing 1 volume of 100% PEG400 and 3 volumes of crystallization solution as a cryoprotectant, before cryocooling to 100 K in a stream of gaseous N_2. Data were collected at 100 K at beamline I911−3 of the MAX-II synchrotron, Lund, Sweden.[28] All data were integrated using XDS.[29] Diffraction data for R were collected in a single pass, while that for S involved two passes, one at low resolution with lower exposure time followed by one at high resolution, and subsequently scaled and merged with XSCALE.[29]

MTZ files were generated with Aimless.[30] Cross-validation during refinement was based on 10% of the reflections. An initial structure solution was determined through rigid-body refinement in Refmac5[31] using as a starting model the lactose−galectin-3C structure[14] with lactose and water molecules removed and with the resolution limit set to 3.5 Å. The structures of the R and S ligand stereoisomers were built manually using Chimera[32] and geometric restraints for the ligands were obtained through phenix.eLBOW.[33] Restrained refinement was then performed using phenix.refine[34] using data to the diffraction limit. Manual rebuilding, including addition of water molecules, was done using Coot.[35]

Ensemble Refinement of Crystal Structures. Ensemble refinement of the X-ray diffraction data was performed using the module phenix.ensemble_refinement in the Phenix software suite.[36] The X-ray crystal structures of the *S*-galectin-3C and *R*-galectin-3C complexes from the previous section were used as starting structures. The crystallographic water molecules were kept and hydrogen atoms and missing atoms in the protein were added using the Leap module from the Amber 14 software.[52] Ligand restraints and coordinates were the same as those used in the original refinement.

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

The collective dynamics of the protein were described using a TLS model with a single group, which included both the protein and the ligand atoms. A model including two TLS groups was also tested—one for the ligand and one for the protein—but it gave worse results (R_{free} values of 0.20 compared to 0.17 for the single TLS model). 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} = 0.7 for both 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. The X-ray weight-coupled temperature bath offset was kept at the default value of 5 K. 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.[37]

NMR Sample Preparation. The galectin-3C concentration was 0.32, 0.2, and 0.34 mM for the ^{15}N, $^{15}N/^{13}C$, and $^{15}N/^{13}C/^{2}H$ samples, respectively. The ligands were dissolved in neat DMSO to a concentration of 8.2 mM for S and 35 mM for R. The protein−ligand complexes were prepared by titrating the ligand into the protein, while monitoring the ^{15}N heteronuclear single-quantum correlation (HSQC) spectra. The final DMSO content in the NMR sample was 4.3% for S and 1.2% for R.

NMR Resonance Assignments and Chemical Shift Analysis. Backbone chemical shift assignments were based on HNCACB[38] spectra and previous assignments for various galectin-3C complexes.[5] Methyl groups were assigned using CCH-TOCSY and HCCH-TOCSY experiments.[39,40] All spectra were processed using NMRPipe,[41] employing a processing protocol including a solvent filter, square cosine apodization, and zero filling to twice the number of points in all dimensions. All spectra were analyzed using the CCPNmr program suite.[42] Chemical shift differences were evaluated as weighted distances: $([\Delta\delta(^{1}H)]^2 + [0.1\Delta\delta(^{15}N)]^2)^{1/2}$ for backbone amides and $([\Delta\delta(^{1}H)]^2 + [0.25\Delta\delta(^{13}C)]^2)^{1/2}$ for methyls.

NMR Relaxation Experiments and Data Analysis. ^{15}N R_1, R_2, and $\{^{1}H\}-^{15}N$ nuclear Overhauser effect (NOE) experiments targeting the backbone amides were performed at magnetic field strengths of 11.7, 14.1, and 21.1 T, and a temperature of 301 K. Spectral widths were 14−16 ppm 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 acquired at 11.7 and 14.1 T, 0−3 s for R_1 acquired at 21.1 T, and 0−0.2 s for R_2 (at all fields) with a 1.2 ms delay between refocusing pulses. The NOE was measured using a ^{1}H saturation time of 7 s and a recycle delay between experiments of 3 and 7 s for experiments acquired at 11.7 and 14.1 T, respectively, while the reference experiment was acquired using a recycle delay of 10 and 14 s at 11.7 and 14.1 T, respectively. NOE experiments performed at 21.1 T employed a ^{1}H saturation time of 6 s and a recycle delay between experiments of 2 s, while the reference experiment was acquired with a recycle delay of 14 s. Peak intensities were evaluated as partial peak volumes calculated over 3 × 5 points in the direct and indirect dimension, respectively. Monoexponential functions were fitted to the R_1 and R_2 relaxation decays using the CCPNmr program suite and bootstrap error estimation. NOEs were calculated as the ratio of the peak intensities in the saturated and reference experiments, and the standard errors were determined by propagating the errors of intensities estimated from the baseplane noise.

$R_1(D_Z)$, $R(3D_Z^2 - 2)$, $R_2(D_+)$, and $R(D_+D_Z + D_ZD_+)$ ^{2}H relaxation experiments[43] targeting the methyl groups 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 70 and 84 points for ^{13}C at 11.7 and 14.1 T, respectively. The number of points recorded were 1024 for ^{1}H at both static magnetic field strengths. Relaxation decays were sampled by 9 points covering 0−0.1 s for $R_1(D_Z)$ and $R(3D_Z^2 - 2)$, 0−20 ms for $R_2(D_+)$ and $R(D_+D_Z + D_ZD_+)$. The recycle delay was 1.8−2 s. Peak volumes were evaluated using the program suite PINT.[44] Monoexponential functions were fitted to the relaxation decays using an in-house MATLAB script with Monte Carlo error analysis.[25]

^{15}N CPMG relaxation dispersion experiments were performed at 301 K and static magnetic field strengths of 11.7 and 14.1 T on *S*-galectin-3C and 14.1 T on *R*-galectin-3C, using a single experimental data point per refocusing frequency.[45,46] A series of 19 relaxation dispersion spectra were acquired with CPMG refocusing frequencies ranging from 50 to 800 Hz, and in addition a single reference spectrum was recorded without any CPMG refocusing pulses. The relaxation dispersion data were analyzed using the general equation for two-state exchange.[47−49]

Model-Free Analysis of NMR Relaxation Data. Backbone amide model-free parameters were fitted using the program suite relax,[50−52] using a N−H bond length of 1.02 Å and a ^{15}N chemical shift anisotropy of −172 ppm. The backbone optimization was restricted to five different models defined by the parameter sets: $\{O^2\}$, $\{O^2, \tau_e\}$, $\{O^2, R_{ex}\}$, $\{O^2, \tau_e, R_{ex}\}$, or $\{O^2_f, O^2_s, \tau_s\}$, where O^2, O^2_f, and O^2_s denote the order parameter with subscripts *f* and *s* indicating that the order parameter can be resolved into amplitudes of fluctuation taking place on separate time scales (fast and slow), τ_e and τ_s denote effective correlation times for the internal motion with subscript *s* indicating that the correlation time is associated with the slower time scale, and R_{ex} denotes exchange contributions to R_2; in addition, the correlation time for overall rotational diffusion, τ_c, was also fitted.[53] Side-chain methyl-axis model-free optimization was performed using in-house routines implemented in MATLAB. The ^{2}H quadrupolar coupling constant was set to 167 kHz.[54] Three different models were fitted using two $\{O^2, \tau_f\}$, three $\{O^2, \tau_f, \tau_{eff}\}$, or four $\{O_f^2, O_s^2, \tau_f, \tau_{eff}\}$ 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 was fixed to the value obtained from the backbone model-free optimization. Model selection was performed using an *F*-test at the level $\alpha = 0.95$ ($p < 0.05$).[56]

Conformational Entropy Estimates from Order Parameters. The backbone conformational entropy change, going from state A to B, was estimated from the NMR order parameters using the relationship[1,57]

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

where $O_{X,k}$ is the order parameter for residue *k* in state *X*, and the sum runs over all residues. In a similar way, the conformational entropy change of the side chain methyl-axis was determined using[57]

$$\Delta S_{AB} = R \sum_m C_m \sum_n (O^2_{B,n} - O^2_{A,n}) \quad (2)$$

where C_m is a function of the residue type. The sums run over all residues *n* of type *m*. C_m = 1.32 for Val and Thr, 3.1 for Ile and Leu, and 2.31 for Met. The entropy for Ala side chains was calculated using eq 1.

The entropy estimated from eqs 1−2 rests on a number of assumptions that have been discussed in the literature.[1,58,59] Most importantly, the approach does not account for contributions from conformational fluctuations with correlation times greater than τ_c, and does not consider the effects of correlated motion. An alternative approach, based on empirical calibration, has been proposed recently.[60] Here, the total conformational entropy is estimated from the average methyl-axis order parameters:

$$\Delta S_{AB} = s_d N_\chi \Delta\langle O^2\rangle_{AB} \quad (3)$$

where $s_d = -(4.8 \pm 0.5) \times 10^{-3}$ kJ/mol/K is an empirically determined constant,[60] N_χ denotes the number of dihedral angles, and $\Delta\langle O^2\rangle_{AB}$ is the difference between states A and B in their average methyl-axis order parameter. This empirically calibrated estimate of conformational entropy is believed to capture also the effects of correlated motion and motions occurring on time scales greater than τ_c.[60]

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

Molecular Dynamics Simulations and Analysis. All MD simulations were run with the Amber 14 software suite.[61] The X-ray crystal structures of the *S*-galectin-3C and *R*-galectin-3C complexes were used as the starting points for MD simulations. The PDB structure 3ZSL was used for the simulations of apo galectin-3C. Separate simulations were run for the two different conformations observed for ligand S. 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, so that 4965−5593 water molecules were included in the simulations. The simulations were set up in the same way as in our previous studies of galectin-3C.[4,62,63] 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 active-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 of the lactose-bound state,[15] NMR measurements, and previous extensive test calculations with MD.[64] 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,[65] water molecules with the TIP4P-Ewald model,[66] whereas the ligands were treated with the general Amber force field.[67] Charges for the ligands were obtained with the restrained electrostatic potential method.[68] 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.[69] These calculations were performed with the Gaussian 09 software.[70] The potentials were then used by antechamber to calculate the charges. A few missing parameters were obtained with the Seminario approach:[71] The geometry of the ligands was optimized at TPSS/def2-SV(P) level, followed by a frequency calculation using the aoforce module of Turbomole 7.01.[72] From the resulting Hessian matrix, parameters for the missing angles and dihedrals were extracted with the Hess2FF program.[73] These parameters are given in Table S1 in the Supporting Information.

For each complex, 10 000 steps of minimization were used, followed by 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 209 kJ/mol/Å^2. Finally, the system was equilibrated for 2 ns, followed by 10 ns of production simulation, both performed with constant pressure and without any restraints. For each protein−ligand complex, 10 independent simulations were run, employing different solvation boxes and different starting velocities.[74] Consequently, the total simulation time for each complex was 100 ns. All bonds involving hydrogen atoms were constrained to the equilibrium value using the SHAKE algorithm,[75] allowing for a time step of 2 ps. The temperature was kept constant at 300 K using Langevin dynamics,[76] with a collision frequency of 2 ps^{-1}. The pressure was kept constant at 1 atm using a weak-coupling isotropic algorithm[77] with a relaxation time of 1 ps. Long-range electrostatics were handled by particle-mesh Ewald (PME) summation[78] 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.[37]

Conformational Entropy Estimates from MD Simulations. To validate the MD trajectories by NMR, we calculated order parameters from the MD trajectories. The N−H order parameters were obtained using isotropic reorientational eigenmode dynamic analysis.[79] The covariance matrix of the NH bond vectors was obtained from the trajectories by the cpptraj module[37] in the Amber 14 software.[61]

A total of 10 000 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. Conformational entropies were calculated from the ensemble of configurations of the protein and ligands by analyzing the dihedral angle fluctuations.[4,63,80,81] 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.

Both entropies and order parameters were calculated as averages over 50 simulations of 2 ns each (with 200 snapshots in each, i.e., each of the 10 simulations were divided into five parts of equal length). The 2 ns time window is similar to the rotational correlation time of the protein. This procedure yields more stable entropy estimates by restricting the dependence on rare events.[63] The reported uncertainties are standard errors over these 50 simulations.

To estimate the effect of correlation, entropies were also calculated employing the maximum information spanning tree algorithm[82,83] (MIST), with the pdb2entropy program.[84] Entropies were calculated to the tenth nearest neighbor to account for high-order correlations, whereas entropies calculated to the first nearest neighbor were considered correlation-free.

Water Structure and Solvation Thermodynamics. We analyzed the structure and thermodynamics of the solvent around the two ligands (R and S) bound to galectin-3C, using GIST,[85] implemented in the cpptraj module of the Amber 14 software. The method requires snapshots from MD simulations in which the solute is kept restrained. Therefore, we first performed clustering of the trajectories from the unrestrained simulations described above, using the hierarchical agglomerative clustering approach, implemented in the cpptraj module, with average-linkage criteria and the ligand RMS as distance metric. The minimum distance between clusters was set to 3.5 Å. Subsequently we performed 10 independent 10 ns long MD simulations for each identified cluster. In these simulations the protein was kept restrained toward the starting crystal structure, and the ligand was kept restrained toward the conformation which best represents the cluster, both with a force constant of 10 kcal/mol/Å^2.

For each cluster, the water−water interaction energy, E_{w-w}, and solute−water interaction energy, E_{s-w}, as well as translational, S_{trans}, and rotational, S_{rot}, entropy contributions were calculated for a rectangular grid of dimensions 30 Å × 21 Å × 21 Å, centered on the ligand and extended at least 3 Å on each side of the ligand. The grid was divided into cubic boxes (0.5 Å × 0.5 Å × 0.5 Å), for which the thermodynamic properties were calculated. The sum of these properties over the entire region reveals changes in the hydration thermodynamics of the region for each cluster, relative to the thermodynamics of the bulk water. For each of the two ligands, the solvation free energy, ΔG_{solv}, was calculated as a sum over solvation free energies for each cluster, $\Delta G_{solv}(i)$, multiplied by the probability of finding the ligand in conformation i, $p(i)$:

$$\Delta G_{solv} = \sum_i \Delta G_{solv}(i) p(i)$$

A separate set of solute-restrained MD simulations was performed in which both the protein and the ligand were restrained toward the crystal structure. To analyze these simulations, we used a 27 Å × 14 Å × 15 Å grid.

RESULTS AND DISCUSSION

Ligand Design and Synthesis. We investigated the driving forces underlying affinity and selectivity in ligand binding by carrying out a comparative analysis involving the binding of two diastereomeric ligands R and S (Figure 1) to galectin-3C. The design of ligands R and S was inspired by the high-affinity (K_d = 2 nM) galectin-3 ligand 1−1′-sulfanediyl-bis-{3-deoxy-3-[4-(3-fluorophenyl)-1*H*-1,2,3-triazol-1-yl]-β-D-galactopyranoside}[22,86] (Figure 1A). The high-affinity ligand interacts with galectin-3 via one of the galactose residues (that on the left-hand side in Figure 1A) in the conserved galactose binding site and the fluorophenyltriazolyl moieties interacts via face-to-face stacking with arginine side chains and one fluorine−amide orthogonal multipolar interaction.[22] The second galactose moiety interacts with only a single hydrogen

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

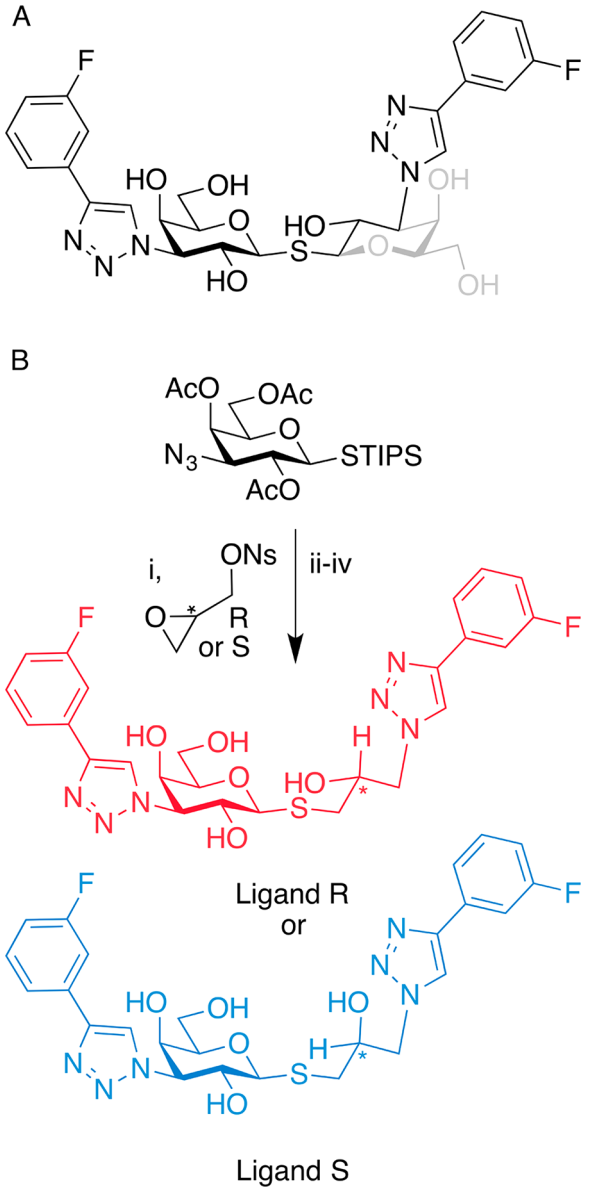


Figure 1. Chemical structures and synthesis of ligands. (A) Chemical structure of the parent, high-affinity ligand 1−1′-sulfanediyl-bis-{3-deoxy-3-[4-(3-fluorophenyl)-1*H*-1,2,3-triazol-1-yl]-*β*-D-galactopyranoside}. The noninteracting atoms of one galactopyranose are depicted in gray. (B) Synthesis and structures of the two ligands R (red) and S (blue). The stereocenter is located at the propyl C2 (marked by an asterisk). Reagents and conditions: (i) TBAF·3H_2O, dry THF. (ii) NaN_3, NH_4Cl, dioxane/H_2O 1:1. (iii) 1-Ethynyl-3-fluorobenzene, CuI, Et_3N, DMF. (iv) MeONa, MeOH.

bond (Figure 1A; the noninteracting parts are depicted in gray) to the protein, leading us to hypothesize that this galactopyranose ring could be mimicked by a 2-hydroxypropyl chain, which would open up for the synthesis of two diastereomeric ligands R and S (Figure 1B).

Synthesis of the ligands R and S relied on fine-tuning the reactivity between a 1-sulfhydryl-galactopyranose nucleophile and a doubly electrophilic glycidyl derivative: In situ fluoride-mediated activation of the masked nucleophilic triisopropylsilyl thiogalactoside and (*R*)- and (*S*)-glycidyl nosylate, respectively, proceeded stereoselectively in high yields, while other galactose nucleophiles (-SAc, -SH, thiouronium salts, and thioxanthate) and glycidyl electrophiles (glycidyl tosylate, *tert*-butyl dimethyl silyl glycidyl, and epi-chlorohydrin) gave lower yields and stereochemical scrambling due to nucleophilic attack occurring on both C1 and C3 of the glycidyl derivatives, or due to epoxide opening followed by intramolecular substitution to epoxide reclosing. Regioselective ring-opening of the epoxide with NaN_3, Cu(I)-catalyzed cycloadditions with 1-ethynyl-3-fluorobenzene, and finally Zemplen transesterification gave ligands R and S in 99+% purities.

Overall Binding Thermodynamics. We characterized the thermodynamics of ligand binding using ITC. We carried out five replicate titrations for each of ligands R and S, and analyzed the binding isotherms by performing a combined fit of the replicate data sets (Figure 2; Figure S1). Table 1 lists the

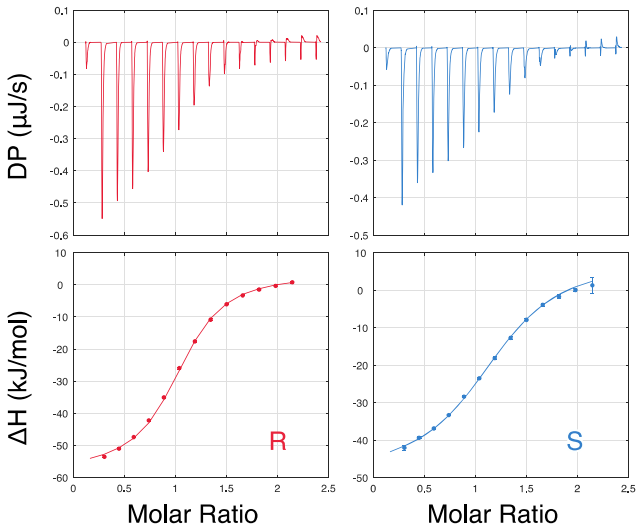


Figure 2. ITC experiments of ligand binding to galectin-3C. Example isotherms describing the titration of galectin-3C with ligand R (left-hand side) and ligand S (right-hand side). The top panels show the raw thermograms of differential power plotted versus the ligand:protein molar ratio, while the lower panels show the resulting isotherms. The binding curve results from global fitting of 5 replicate data sets. Error bars are smaller than the size of the symbols, except for the last titration point for ligand S. (Figure S1 shows all 5 isotherms for each ligand).

Table 1. Overall Binding Thermodynamics from ITC

complex	K_d (10^{-6} M)	ΔG°_{tot} (kJ/mol)	ΔH°_{tot} (kJ/mol)	$-T\Delta S^\circ_{tot}$ (kJ/mol)
R-galectin-3C	1.0 ± 0.03	−34.6 ± 0.1	−60.4 ± 0.4	25.8 ± 0.4
S-galectin-3C	2.1 ± 0.1	−32.7 ± 0.1	−55.7 ± 0.9	22.9 ± 0.9
difference (R − S)		−1.9 ± 0.1	−5 ± 1	3 ± 1

resulting binding thermodynamics. Both ligands have dissociation constants in the low micromolar range, $K_d(R) = (1.0 \pm 0.03) \times 10^{-6}$ M and $K_d(S) = (2.1 \pm 0.1) \times 10^{-6}$ M, and the results correlate well with those obtained in competitive fluorescence polarization experiments, $K_d(R) = (0.43 \pm 0.04) \times 10^{-6}$ M and $K_d(S) = (0.67 \pm 0.5) \times 10^{-6}$ M. As reported previously,[5] K_d values determined by ITC are typically found to be higher by a factor of 2−4 than those measured by fluorescence polarization, but the relative affinities are unchanged within errors. The free energies of binding differ

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

by only $\Delta\Delta G°(R - S) = -1.9 \pm 0.1$ kJ/mol, but the differences in $\Delta H°$ and $-T\Delta S°$ are greater and consequently opposite in sign, indicating enthalpy−entropy compensation: $\Delta\Delta H°(R - S) = -5 \pm 1$ kJ/mol and $-T\Delta\Delta S°(R - S) = 3 \pm 1$ kJ/mol.

Crystal Structures Reveal Subtle Differences in Binding Modes. The crystal structures of the *R*- and *S*-galectin-3C complexes were refined to resolutions of 1.34 and 1.19 Å, respectively (see Table S2 for a summary of refinement statistics). The quality of the electron density data is sufficient to reveal the chirality of the ligands unambiguously (Figure 3 and Figure S2). As shown in Figure 3, the two complexes have closely similar structures, with essentially no difference in the protein backbone conformation. The RMS deviation between the two structures is 0.13 Å for 473 backbone atoms and 0.59 Å when 2054 atoms are compared, including side chains.

Below we will denote the aromatic ring substituents on galactose C3 as the "left hand side" (LHS), while the aromatic rings connected to the propylic chain will be referred as the "right hand side" (RHS); this notation is according to the viewpoint of Figure 3 and all subsequent renditions of the structures. The LHS shows perfect overlap between the two complexes. The 3-fluorophenyl substituent sits in a pocket generated by the displacement of Arg144, with the fluorine atom pointing toward the protein backbone. Key interactions involving the meta-fluorinated phenyl triazole on the LHS have been described previously.[22]

The *B*-factors of the ligand atoms on the LHS are very similar in the two complexes (10−15 Å^2), and lower than those of the RHS (20−35 Å^2 in R and 20−40 Å^2 in S), indicating that the LHS is more ordered. The electron density for Arg144, which stacks with the fluorinated phenyl ring of the LHS, is slightly less well-defined in R than in S. The difference in mobility of Arg144 does not seem to be correlated to the minor differences in water structure (see below).

Although R and S have a different configuration at propyl C2, the conformation of the ligand adjusts to allow the hydroxyl group of the stereocenter to maintain a hydrogen bond with Glu184. The configuration of the R-stereoisomer enables the propyl linker to adopt the same conformation as the corresponding segment in the glucose ring of the parent compound (cf. Figure 1). Thus, the C2 hydroxyl group of R makes an H-bond to Glu184 with its hydrogen atom in a staggered conformation with respect to the aliphatic hydrogen atom on the C2 carbon, as observed in the lactose and glycerol complexes by neutron crystallography.[15] In contrast, the hydroxyl group in S is positioned in an eclipsed conformation with respect to the aliphatic hydrogen, which is expected to be energetically less favorable. This conformational adjustment results in different interactions of the two ligands with the protein at the RHS of the binding site. Furthermore, the RHS of R is modeled with a single conformation, whereas the RHS of S is modeled as two conformations in which the fluorinated ring has two orientations related by an 180° flip. At the RHS, both R and S interact with Arg186, despite the differences in conformation at this end of the ligand. S appears at first glance to have tighter interaction with Arg186 due to a better alignment between the π orbitals of the ligand phenyl ring and the face of the arginine guanidinium group. However, the results from ensemble refinement suggest that the S isomer in fact has higher mobility (see below).

Water molecules are well conserved around the binding site. Particularly, we see that waters around the LHS overlap very

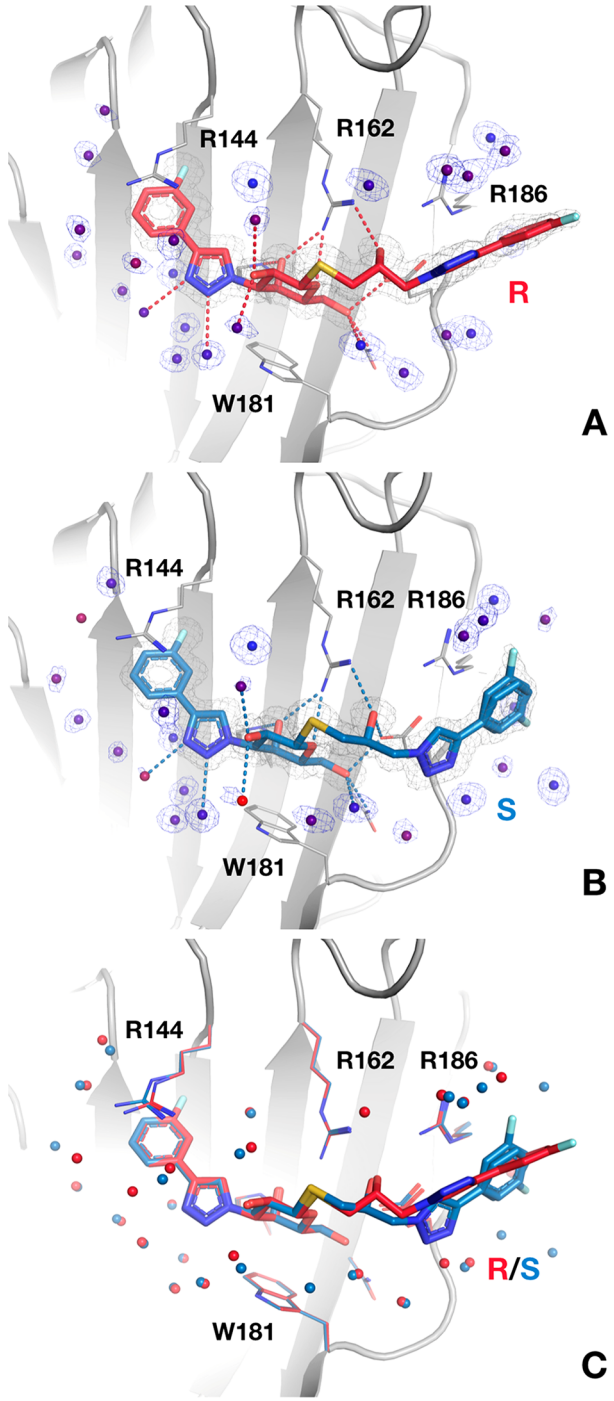


Figure 3. X-ray crystal structures of the ligand−galectin-3C complexes. (A) *R*-galectin-3C (PDB ID 6QGF). (B) *S*-galectin-3C (PDB ID 6QGE). (C) Overlay of the two complexes. The protein backbone is shown in ribbon representation (gray), key ligand-coordinating side chains are shown in stick representation, and hydrogen bonds to the ligands are shown as dashed lines. The 2m|F_o|−D|F_c| electron density map of the ligand and water molecules, contoured at 1.0 σ, is shown as a gray mesh. Carbon atoms of the R ligand are colored red, while those of the S ligand are blue. Water molecules that are within 5 Å of either ligand are represented as small spheres. In panels A and B, water molecules are colored by *B*-factor, on a spectrum from dark blue at 15 Å^2 to bright red at 70 Å^2. Water molecules shown without electron density are visible at <1.0 σ, but are poorly ordered. In panel C, water molecules belonging to *R*-galectin-3C are colored red and those belonging to *S*-galectin-3C are colored blue.

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

well between the two complexes. The minor differences observed could be due to the slightly different resolutions of the two complexes. For the RHS the different conformations of the ligands result in more distinct water structures.

Chemical Shift Mapping of Ligand Binding. Chemical shift assignments of *R*-galectin-3C and *S*-galectin-3C were based on a HNCACB experiment and the apo galectin-3C assignments reported previously.[4,5] Minor chemical shift differences are observed for the backbone amides throughout the protein; the RMS chemical shift difference between the ligand-bound and apo forms of galectin-3C in the ^{1}H and ^{15}N dimensions are 0.06 and 0.30 ppm for *R*-galectin-3C (Figure S3A) and 0.05 and 0.26 ppm for *S*-galectin-3C (Figure S3B). The methyl chemical shifts show changes similar to those of the backbone, with RMSDs of 0.03 ppm (^{1}H) and 0.1 ppm (^{13}C) for *S*-galectin-3C, and 0.04 and 0.1 ppm for *R*-galectin-3C. The largest chemical shift changes induced by ligand binding are observed for residues in close proximity to the ligand in the crystal structure (Figure S3C), demonstrating that the binding mode observed in the crystal structure is maintained in solution.

Significant chemical shift differences between the *R*- and *S*-galectin-3C complexes are observed in the binding site (Figure 4). The overall chemical shift RMSD is 0.02 and 0.14 ppm for backbone ^{1}H and ^{15}N, respectively, and 0.04 and 0.02 ppm for methyl ^{1}H and ^{13}C, respectively. Two methyl groups, Val170γ1 and Val172γ1, show a weighted chemical shift difference greater than 0.05 ppm between the two complexes. Furthermore, the ^{1}H and ^{15}N chemical shifts of the Arg162 and Arg186 guanidine groups differ between the two complexes. In both cases, the ^{1}H chemical shift is greater in *R*- than in *S*-galectin-3C, suggesting that the NHε atom forms a stronger hydrogen bond or that the population of hydrogen bonded conformations is greater in the R-complex. These four side chains are located closely together and in proximity of the stereocenter of the ligand. Notably, chemical shift differences are also observed in regions of the protein where the average structures are virtually identical between the two complexes, such as the backbone amides of Leu147 and Phe159, which form a pair of NH−CO hydrogen bonds. This observation indicates that subtle differences exist in the conformational ensembles sampled by the two complexes, a topic that we address in more detail below.

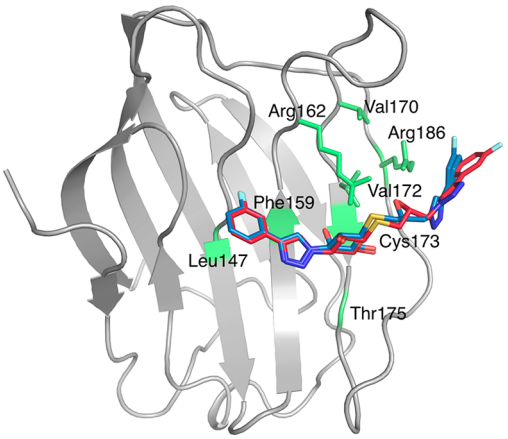


Figure 4. Chemical shift differences between the *R*- and *S*-galectin-3C complexes. Residues with weighted chemical shift differences $|\Delta\delta(R - S)| \geq 0.05$ ppm are highlighted in green on the structure of the *R*-galectin-3C complex with ligand S superimposed. These include the backbone amides of residues Leu147, Phe159, Cys173, Thr175, and Arg186, as well as the methyl groups of Val170 and Val172 and guanidine groups of Arg162 and Arg186, all located in the binding site. Val172 is situated beneath the side chain of Arg162 in the view of the figure.

Ensemble Refinement of Crystal Structures Highlights Differences in Flexibility. To investigate the conformational mobility of each complex in the crystal we carried out ensemble refinement of the structure against the X-ray diffraction data.[36] The resulting ensembles yield R_{free} = 0.1709 and R = 0.1358 for *R*-galectin-3C, and R_{free} = 0.1625 and R = 0.1339 for *S*-galectin-3C, values that are comparable to those resulting from traditional refinement (Table S2).

The results indicate that the S diastereomer shows larger fluctuations in the crystal than does R, due to a large variation in the RHS sp^3 dihedral angles, as shown in Figure 5. This result agrees with the dual conformation of ligand S observed in the traditionally refined crystal structure, although the conformational variation is much greater in the ensemble representation. In particular, the H-bond between the C2 hydroxyl group and Glu184 is broken in a much larger proportion of the ensemble structures for S than for R. The ensemble refinement also confirms that the R ligand stays in a single conformation, although with some translational movement of the RHS end.

The protein also exhibits variable flexibility. The side chains of Asn160, Arg162, Glu165, and Arg186, which form hydrogen bonds with both ligands, have well-defined positions, whereas larger fluctuations are observed for Arg144, which interacts through $\pi-\pi$ stacking with the ligand phenyl ring at the LHS, and Arg168, which does not interact with the ligand. Arg144 shows slightly greater amplitudes of motion in the R-complex, in keeping with the difference in *B*-factors of the traditionally refined structures. The great variability in the side-chain orientation of Arg144 is also reflected by the NMR data (see below). On the other hand, the ensemble-refined crystal structure of *S*-galectin-3C shows higher fluctuations of several parts of the protein, e.g., the Asn164−Arg168 loop region (neighboring the RHS of the bound ligand), Lys138−Ala142, and Pro113−Va118 (Figure 5, insets).

The resulting ensembles indicate that the *S*-galectin-3C complex shows considerably higher mobility than does *R*-galectin-3C, providing qualitative evidence that protein and ligand conformational entropy is greater in *S*-galectin-3C. We attempted to quantitate the entropy difference from the ensembles, resulting in calculated values that were qualitatively consistent with our other results; however, the estimated standard errors were far greater than the difference between the R- and S-complexes (data not shown).

Differences in Conformational Fluctuations Measured by NMR. We carried out a suite of NMR relaxation experiments that probe conformational dynamics on the picosecond to nanosecond time scale to yield the amplitudes of conformational fluctuations in terms of order parameters, denoted O^2. We measured ^{15}N backbone relaxation rates at three static magnetic field strengths and methyl ^{2}H relaxation rates at two static magnetic field strengths. Out of 138 residues, ^{15}N relaxation data could be measured for 101 and 100 backbone amides in *R*-galectin-3C and *S*-galectin-3C, respectively. Likewise, out of a total of 85 methyl groups, ^{2}H relaxation rates could be measured for 65 and 47 methyl groups in *R*-galectin-3C and *S*-galectin-3C, respectively. The

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

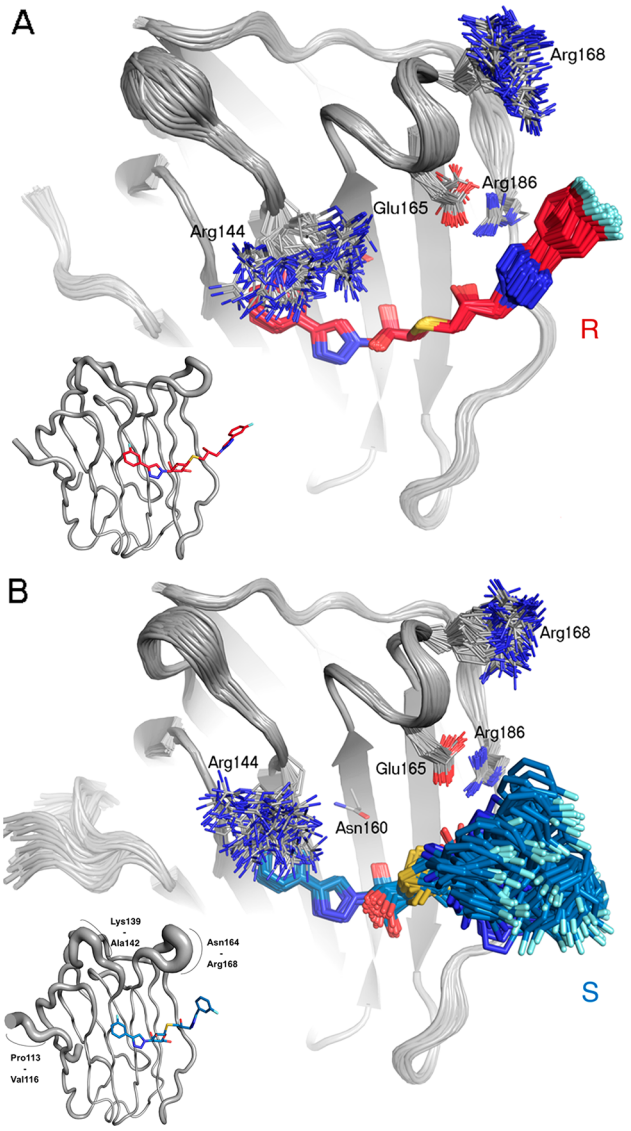


Figure 5. Ensemble refined X-ray crystal structures. Overlay of the 100 structures with the lowest R_{free} generated by ensemble refinement for (A) *R*-galectin-3C (red ligand) and (B) *S*-galectin-3C (blue ligand). Insets: The protein backbone is displayed as a tube with a diameter corresponding to the ensemble RMS fluctuations for all atoms of that residue (the ligand is shown in its crystal structure conformation).

missing residues had cross-peaks that were overlapped or too broadened to allow for quantitative analysis.

We characterized the amplitudes of conformational fluctuations using the model-free formalism.[87,88] The best-fit rotational diffusion tensor is anisotropic with a correlation time (τ_c) of 7.5 and 8.1 ns, anisotropy of 1.1 and 1.1, and rhombicity of 0.9 and 1.2 for *R*-galectin-3C and *S*-galectin-3C, respectively. The higher value of τ_c observed for *S*-galectin-3C is fully explained by the slightly higher concentration of DMSO in this sample, which increases the solvent viscosity.[89,90]

The backbone order parameters are very similar in the two complexes; the mean values and standard deviations are $\langle O^2 \rangle$ = 0.85 ± 0.05 and 0.84 ± 0.05 for *R*-galectin-3C and *S*-galectin-3C, respectively. A significant difference in O^2 is observed for residues Tyr118, Ile132, Ile171, Asp178, Arg183, and Leu242, none of which is located directly in the binding site (Figure 6).

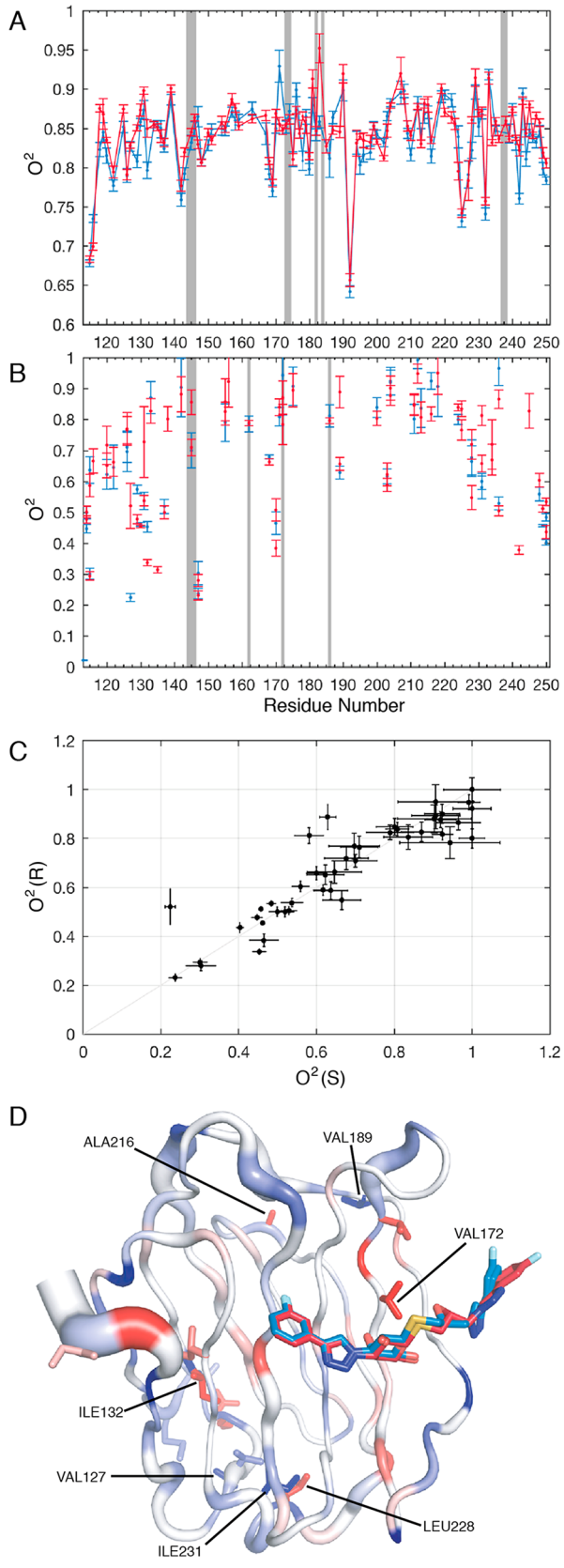


Figure 6. NMR order parameters for *R*- and *S*-galectin-3C. (A) Backbone O^2 values. (B) Side chain O^2 values for arginine $^{15}N\varepsilon$ and methyl axes. Data for *R*- and *S*-galectin-3C are shown in red and blue, respectively. Gray bars indicate residues in contact with the ligand (residues for which any backbone amide atom or methyl atom is within 5 Å of any ligand atom). (C) Scatter plot comparing the

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

Figure 6. continued

methyl-axis O^2 values for *R*- and *S*-galectin-3C presented in panel B. The straight line with slope of 1 is drawn to guide the eye. (D) ΔO^2 color coded onto the *R*-galectin-3C structure with ligand S superimposed. Residues with $\Delta O^2(\mathrm{R}-\mathrm{S}) > 0$ are colored blue, while those with $\Delta O^2(\mathrm{R}-\mathrm{S}) < 0$ are colored red. The intensity of the color scales with the magnitude of ΔO^2 from red via pink to white ($-0.1 \leq \Delta O^2(\mathrm{R}-\mathrm{S}) < 0$) and from white via light blue to dark blue ($0 < \Delta O^2(\mathrm{R}-\mathrm{S}) \leq 0.1$). Residues for which no data are available are colored white. Side chains are shown in stick representation for residues with a difference in side-chain order parameters of $|\Delta O^2(\mathrm{R}-\mathrm{S})| > 0.05$, and labeled residues have $|\Delta O^2(\mathrm{R}-\mathrm{S})| > 0.1$. Backbone and side-chain ΔO^2 are represented by the color of the tube and sticks, respectively. The width of the tube indicates the average backbone O^2 values in the two complexes: a wider tube indicates a lower order parameter and vice versa.

This result indicates that the different stereochemistry of the ligand and the associated differences in protein conformation affect the amplitudes of backbone fluctuations at remote locations. Backbone order parameters are relatively low in the loop regions at the top of the structure in the view of Figure 6. There is also a difference in order between the two complexes with O^2 being higher for the R-complex. Both of these observations agree well with the ensemble-refined crystal structures.

The order parameters for the methyl-bearing side chains vary significantly over the protein (Figure 6B,C). However, the differences between the two complexes are overall small, except for residues Val127, Ile132, Val172, Val189, Ala216, Leu228, and Ile231, which show $|\Delta O^2(\mathrm{R}-\mathrm{S})| > 0.10$; Figure S4A shows the distribution of $\Delta O^2(\mathrm{R}-\mathrm{S})$. Out of these residues, only Val172 is located in the binding site, next to the stereocenter of the bound ligand. The side chain of Val172 shows a greater degree of freedom in the *R*-galectin-3C complex. The mean values and standard errors of the mean for the methyl-axis order parameters are $\langle O^2 \rangle = 0.68 \pm 0.02$ and 0.64 ± 0.03 for *R*-galectin-3C and *S*-galectin-3C, respectively, when calculated over all residues, and $\langle O^2 \rangle = 0.66 \pm 0.03$ and 0.65 ± 0.03, when calculated over those residues for which data are available for both complexes.

Arginine side chains play a special role in ligand coordination by galectin-3C. Arg144, Arg162, and Arg186 form close interactions with the ligand (cf. Figure 3). However, the side-chain guanidine group of Arg144 is not observed in the NMR spectra, presumably as a consequence of intermediate exchange between alternative positions. This result is in agreement with the ensemble-refined crystal structures, in which Arg144 shows extensive flexibility. The fact that Arg168, which also is highly variable in the structure ensembles, is observed in the NMR spectra indicates that this side chain undergoes dynamic averaging on a faster time scale than does Arg144.

^{15}Nε side-chain order parameters could be measured for 5 out of 9 arginines. Arg162 and Arg186, which interact with the bound ligands, have O^2 values (0.78−0.81) that are higher than the average value of the guanidine groups and only slightly lower than the average value of the backbone. However, there is no significant difference in O^2 between the R- and S-complexes for these two side chains. Only Arg129 and Arg224 show minor differences between the two complexes, $|\Delta O^2(\mathrm{R}-\mathrm{S})|$ of 0.09 and 0.04, respectively (Figure 6B), and both of these residues are located peripherally to the binding site.

Order parameters derived from relaxation measurements report on conformational entropy due to fluctuations with correlation times shorter than τ_c. To investigate whether there are motions occurring on slower time scales, we performed ^{15}N CPMG relaxation dispersion experiments, which sample motions on the 100 μs to 100 ms time scales.[45] In both the R- and S-bound states, a single residue, Val189, exhibits conformational exchange. The exchange rate is identical, within error: k_{ex}= 6300 ± 1300 s^{-1} (R) and 4900 ± 300 s^{-1} (S), indicating that there are no major differences between the two complexes in the extent of conformational sampling on this time scale.

Differences in Conformational Fluctuations Determined by MD Simulations. To complement the information on conformational fluctuations obtained via NMR order parameters for the backbone, guanidine- and methyl-bearing side chains, we performed MD simulations that probe the intramolecular dynamics of all parts of the protein and ligand. Since the crystal structures of the ligand−galectin-3C complexes show two conformations of S, we initiated separate MD simulations for the two conformers. We validated the MD simulations by comparing order parameters calculated from the MD trajectories with those measured by NMR. There is reasonable, but variable, residue by residue agreement between the backbone O^2 values determined by NMR and MD for each complex. The RMSD is 0.05 in all 3 comparisons (Figure S4B,C), which is on par with previous results for other proteins.[65,91]

We studied how the conformation of the ligand varied in the MD simulations by following the dihedral angle representing the orientation of the RHS phenyl ring. In each of the three trajectories, the ligand samples a unimodal and equally wide (∼50°) distribution of the dihedral, indicating that the rotation barrier is high enough that the ligand does not change conformation on the nanosecond time scale.

Conformational Entropy Differences Estimated by NMR. On the basis of the experimental order parameters, we estimated the difference in the conformational entropy between the two complexes, see eqs 1−2. Despite the average values of O^2 being highly similar for the two complexes, residue-specific differences lead to a significant difference in backbone conformational entropy between galectin-3C in the R- and S-bound states, $-T\Delta\Delta S_{bb}(\mathrm{R}-\mathrm{S}) = 17 \pm 5$ kJ/mol. By contrast, the corresponding result for the methyl-axis O^2 is not statistically significant: $-T\Delta\Delta S_{sc}(\mathrm{R}-\mathrm{S}) = -5 \pm 6$ kJ/mol. Taken together, the NMR order parameters yield an estimate of $-T\Delta\Delta S_{bb+sc}(\mathrm{R}-\mathrm{S}) = 12 \pm 8$ kJ/mol, indicating that galectin-3C in the R-bound state has lower conformational entropy than in the S-bound state (Table 2). That is, the conformational entropy difference between the two complexes has the same sign as, but a greater magnitude than the difference in total entropy, $-T\Delta\Delta S°(\mathrm{R}-\mathrm{S})$, obtained by ITC, suggesting that the conformational entropy makes a significant contribution to the overall binding thermodynamics. It should be noted that the NMR-based estimate, $-T\Delta\Delta S_{bb+sc}(\mathrm{R}-\mathrm{S})$, covers only a subset of the dihedral angles in the protein. However, it serves as a useful reference for validating the MD simulations, which provide the total conformational entropy of both galectin-3C and the bound ligand.

We also used the empirically calibrated approach,[60] embodied in eq 3, to estimate the change in the total conformational entropy of the protein. The results yield $-T\Delta\Delta S_{conf}(\mathrm{R}-\mathrm{S}) = 16 \pm 14$ kJ/mol, suggesting, as might be

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

Table 2. Conformational Entropy Differences between *R*- and *S*-Galectin-3C

method	$-T\Delta\Delta S$ (kJ/mol)
NMR backbone + methyls[a]	12 ± 8
MD backbone + methyls[b]	8 ± 3
NMR protein[c]	16 ± 14
MD protein[d]	11 ± 5
MD protein + ligand[e]	10 ± 5

[a]Includes protein dihedrals of the backbone and methyl-bearing side chains, calculated using eqs 1−2. [b]Includes protein dihedrals of the backbone and methyl-bearing side chains. [c]Includes all protein dihedrals, calculated using eq 3. [d]Includes all protein dihedrals. [e]Includes all protein and ligand dihedrals.

expected, that $-T\Delta\Delta S_{bb+sc}$ underestimates the change in total conformational entropy (Table 2).

Conformational Entropy Differences Determined by MD. We calculated the conformational entropy of galectin-3C and the bound ligands in both complexes, using dihedral angle distributions from the MD simulations.[4,63] Table 2 shows the difference in conformational entropy between the two complexes. For both complexes, the dihedral flexibility of galectin-3C decreases upon ligand binding (Table S3). The effect yields a change in conformational entropy, $-T\Delta S_{conf}$, of 86 ± 5 kJ/mol and 74−75 ± 5 kJ/mol for the protein in the R- and S-bound states, respectively (separate MD simulations were initiated from the two conformations of S observed in the crystal structure and these resulted in entropies that agree within 1 kJ/mol). Comparing directly with the NMR-based estimate of conformational entropy associated with the backbone and methyl-bearing side chains, $-T\Delta\Delta S_{bb+sc}(R - S)$ = 12 ± 8 kJ/mol, the corresponding value obtained by MD is 8 ± 3 kJ/mol (Table 2).

The decrease in entropy is greatest for Arg186 in both complexes ($-T\Delta S_{conf}$ = 8−9 kJ/mol). This residue forms hydrogen bonds with Glu184, which interacts with both ligands R and S, and shows the second largest decrease in entropy when ligand S binds ($-T\Delta S_{conf}$ = 4 kJ/mol), but a rather small decrease upon binding ligand R ($-T\Delta S_{conf}$ = 1 kJ/mol). Arg144 also gives a rather large negative entropy contribution upon binding either ligand ($-T\Delta S_{conf}$ = 3−4 kJ/mol). Ile171 gives a large contribution ($-T\Delta S_{conf}$ = 4 kJ/mol) when binding S, but smaller when binding R (1 kJ/mol). This difference is also observed in the backbone O^2 determined by NMR (Figure 6), whereas there is no significant difference between the two complexes in the methyl-axis O^2 values for this residue, whose side chain is oriented away from the binding site. However, the NMR data reveal greater flexibility in the R-complex for the side chains of the neighboring residues Val170 and Val172, which are both oriented toward the binding site. Significantly increased conformational entropy is observed for 3−5 of the residues upon ligand binding, with the largest contribution coming from Asp148 ($-T\Delta S_{conf}$ = −1 kJ/mol).

The total conformational entropy of the protein is greater for *S*-galectin-3C than for the R-complex, $-T\Delta\Delta S_{conf}(R - S)$ = 11 ± 5 kJ/mol (taking into account both MD trajectories for *S*-galectin-3C), which is statistically significant at the 95% level. This result agrees well with the estimate obtained from NMR methyl-axis order parameters, $-T\Delta\Delta S_{conf}(R - S)$ = 16 ± 14 kJ/mol, which implicitly includes the effects of correlated motions and motions on time scales greater than τ_c. Thus, the general agreement supports the conclusion from MIST calculations that effects from correlated motions are minor, and further suggests that slower motions have no major bearing on $\Delta\Delta S_{conf}$ in keeping with the relaxation dispersion data.

The difference between complexes arises from small contributions from many residues (Figure 7). At the level of individual residues, 22−23% show a statistically significant contribution with the same sign as the total difference, whereas 9−11% show the opposite behavior. Among the latter, the largest contributions (−3 kJ/mol) come from Ile171 and Glu184 (Figure 7).

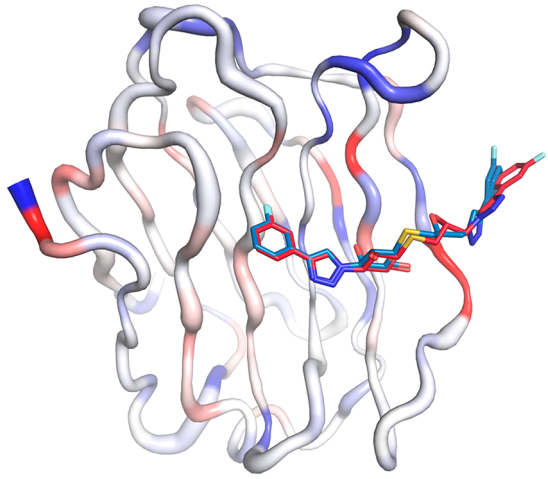


Figure 7. Conformational entropy contributions to $-T\Delta\Delta S_{conf}(R - S)$, reported per residue. $-T\Delta\Delta S_{conf}(R - S)$ is color coded onto the galectin-3C structure with blue hues indicating $-T\Delta\Delta S_{conf}(R - S) > 0$ and red hues indicating $-T\Delta\Delta S_{conf}(R - S) < 0$, with the color intensity ranging from weak (white) for $T\Delta\Delta S_{conf} = 0$ to intense (maximally blue or red) for $|T\Delta\Delta S_{conf}| = 2.8$ kJ/mol. The width of the tube indicates the average conformational entropy values per residue in the two complexes: a wider tube indicates higher average conformational entropy and vice versa. The figure is based on the crystal structure of *S*-galectin-3C.

The change in conformational entropy of the ligand upon complex formation is $-T\Delta S$ = 24 ± 1 kJ/mol and 25−26 ± 1 kJ/mol for *R*- and *S*-galectin-3C, respectively. The difference between R and S is not statistically significant, neither in the bound nor in the free states. The indistinguishable conformational entropy of the free ligands is in line with the expectation that they should have nearly identical chemical potential in the free state, based on their diastereomeric relationship.

We used the MIST approach[82,83] to investigate whether correlated motions affect the estimates of conformational entropy. The results show that the effect of correlation on $-T\Delta\Delta S_{conf}(R - S)$ is minimal, with 1 kJ/mol difference between the first- (without correlation) and tenth-order (with correlation) approximation. Thus, correlations are highly similar in the two states, in agreement with previous results for other proteins.[92]

Thus, taking into account the results for both ligand and protein, the difference in conformational entropy between the two complexes, $-T\Delta\Delta S_{conf}(R - S)$ = 10 ± 5 kJ/mol, is slightly greater than the difference in the net binding entropy, $-T\Delta\Delta S^{\circ}_{tot}(R - S)$ = 3 ± 1 kJ/mol, indicating that protein conformational entropy makes a dominant contribution to $\Delta\Delta G^{\circ}_{tot}(R - S)$. Note that we have designed this comparative

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

study in such a way that the only other contribution to the entropy of binding should originate from differences in solvation entropy of the two complexes, a topic we turn to next.

Grid Inhomogeneous Solvation Theory Reveals Key Differences in Solvation between the Two Complexes. In the standard GIST protocols, sampling of water sites is carried out while keeping the protein and ligand restrained.[85,93] The present case, where the protein and ligand show significant conformational fluctuations in the bound state, presents a challenge to calculations of hydration thermodynamics. We approached the problem by clustering trajectories from the unrestrained MD simulations, which resulted in three clusters for ligand R and four clusters for ligand S (two clusters for each of the two sets of unrestrained MD simulations in the latter case). The subsequent solute-restrained MD simulations, started from each of the clusters for *R*-galectin-3C and *S*-galectin-3C, reveal differences in their hydration thermodynamics (Table 3). Figure 8 affords an overview of water sites,

Table 3. Solvation Thermodynamics from GIST Calculations[a]

complex	*R*-galectin-3C	*S*-galectin-3C	difference (R − S)
$-T\Delta S_{rot}$	398.8 ± 0.6	397.7 ± 1.3	1.2 ± 1.5
$-T\Delta S_{trans}$	319.3 ± 0.4	317.4 ± 0.3	1.9 ± 0.5
$-T\Delta S_{solv}$	718.1 ± 0.9	715.0 ± 1.4	3.1 ± 1.6
ΔH_{s-w}	−2914.1 ± 2.0	−2805.4 ± 1.1	−108.7 ± 2.3
ΔH_{w-w}	−12813.8 ± 2.0	−12877.0 ± 2.0	63.2 ± 2.9
ΔH_{solv}	−15727.9 ± 2.3	−15682.4 ± 2.2	−45.5 ± 3.2
ΔG_{solv}	−15009.7 ± 1.8	−14967.3 ± 2.7	−42.4 ± 3.3

[a]Rotational, ΔS_{rot}, and translational, ΔS_{trans}, entropy as well as the solute−water interaction energy, ΔH_{s-w}, and water−water interaction energy, ΔH_{w-w}, of the studied region, shown relative to bulk water. $\Delta S_{solv} = \Delta S_{rot} + \Delta S_{trans}$, $\Delta H_{solv} = \Delta H_{s-w} + \Delta H_{w-w}$, and $\Delta G_{solv} = \Delta H_{solv} - T\Delta S_{solv}$. All terms are in kJ/mol. Reported uncertainties are the standard errors over the ten independent MD simulations.

i.e., regions with higher density than bulk water, surrounding the bound ligands. Overall, the distributions of highly populated water sites are similar in the two complexes (compare Figures 8A and 8B) and agree well with the crystal structures. However, the close-up view in Figure 8C reveal subtle differences in water positions, especially in the RHS region, where the two structures differ the most. Water molecules in the crystal structures with a low *B* factor overlap well with the highly populated water sites from the GIST analysis, whereas the overlap is poorer for water molecules with a higher *B* factor (Figure S5). Those GIST water densities also have a less spherical shape, indicating a larger mobility of the water structure.

There is a large difference in solvation enthalpy, which is compensated by protein−protein and protein−solvent enthalpies (outside the grid) that are large and hard to estimate accurately, whereas the difference in protein−ligand interaction energies between the R- and S-complexes is modest. Thus, we conclude that the higher binding affinity for the R diastereomer includes a contribution from favorable hydration enthalpy that is dominated by solute−water interactions around the binding site.

Focusing next on solvent entropy, we note that the difference between the two complexes amounts to only $-T\Delta\Delta S_{solv}(R - S) = 3 \pm 2$ kJ/mol. Although barely

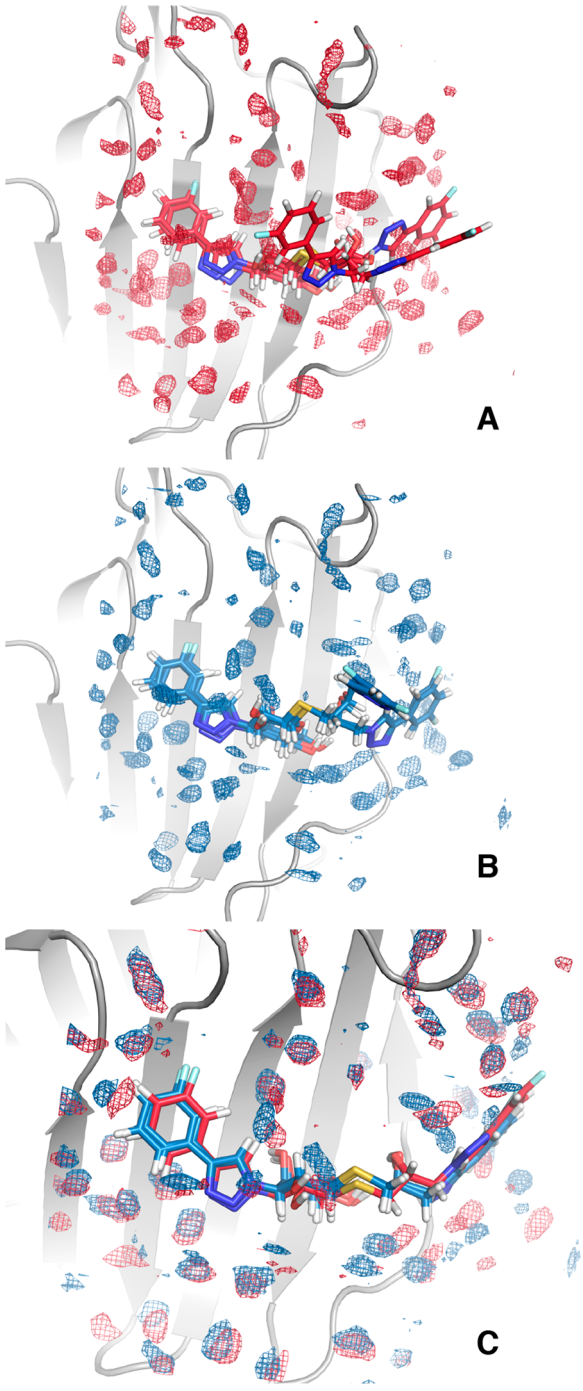


Figure 8. Differences in solvation around the binding site. Regions with high density of water relative to bulk water (six times the bulk water density) are represented as (A) red mesh for *R*-galectin-3C and (B) blue mesh for *S*-galectin-3C. (C) Close-up view of the binding site with the R- and S-complexes superimposed. For clarity, only the highest-occupancy clusters are shown for R and S (both conformations in the latter case) in panel C. See the text for details.

significant, the entropic contribution from solvation appears to add constructively to the conformational entropy ($-T\Delta\Delta S_{conf}(R - S) = 10 \pm 5$ kJ/mol). Arguably, this result is intuitive, as greater disorder in the protein and ligand conformations might be expected to translate to the surrounding water molecules. However, the opposite behavior has also been observed in MD simulations of other systems.[11]

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

The net contribution from conformational and solvent entropy, $-T\Delta\Delta S_{conf+solv}(R - S) = 13 \pm 5$ kJ/mol, is greater than the overall entropy difference determined by ITC, $-T\Delta\Delta S_{tot}(R - S) = 3 \pm 1$ kJ/mol (Figure 9), but the difference is not significant at the 95% confidence level. Taken together, the present results indicate that conformational entropy dominates over solvation entropy in determining the difference in binding entropy between the two ligand−galectin-3C complexes. It remains an open question to what extent these results are general, but we surmise that the relative contributions from conformational entropy and solvent entropy are highly system dependent.

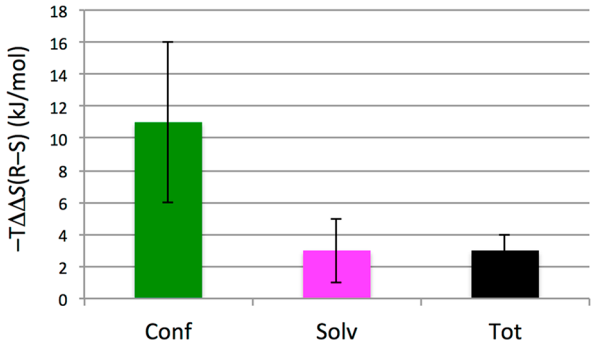


Figure 9. Entropy contributions to the differential binding of ligands R and S to galectin-3C. The bars indicate contributions from conformational entropy, $-T\Delta\Delta S_{conf}$ (green); solvation, $-T\Delta\Delta S_{solv}$ (magenta); and total entropy of binding determined by ITC, $-T\Delta\Delta S_{tot}$ (black). Error bars indicate the standard error (one standard deviation).

Galectin-3 has a relatively exposed and solvent-accessible binding site, which engages numerous water molecules, a feature that certainly contributes greatly to the present results. It would be of great interest to carry out future research to investigate other proteins with different types of binding sites, e.g., those that are less solvent accessible.[12]

CONCLUDING REMARKS

We have carried out a comparative analysis of ligand binding to galectin-3C using two diastereomeric ligands and a range of experimental techniques combined with computational methods. This approach has the important advantage that any differences in the thermodynamics of the two binding processes can be related to the bound state, while the contributions from the free states are expected to cancel—as borne out by the present results. Thus, on the basis of this experimental design, we were able to dissect the thermodynamics underlying the difference in ligand affinity.

The two ligands exhibit closely similar free energies of binding, as might be expected for diastereomers. However, the pair exhibits enthalpy−entropy compensation, so that the two complexes still manifest meaningful differences in both binding enthalpy and entropy that we investigated to pinpoint the driving forces underlying the thermodynamic signatures of binding. Our results demonstrate that the enthalpy−entropy compensation involves interplay between the protein and solvent degrees of freedom. GIST analyses of MD trajectories indicate that the difference in enthalpy includes a sizable contribution from solute−water interactions in favor of the *R*-galectin-3C complex. This contribution is counteracted by a difference in conformational entropy of the protein and a minor entropic component from the solvent that both favor the *S*-galectin-3C complex. Thus, conformational entropy dominates over solvation entropy in determining the difference in binding entropy between the two stereoisomers.

The sum of the conformational and solvation entropies, determined by NMR, MD simulations, and GIST calculations has the same sign as, but is greater than, the total entropy of binding, determined by ITC. Thus, the individual estimates of conformational and solvent entropy correctly identify which protein−ligand complex is favored, but the remaining deviation of $\Delta\Delta S_{conf+solv}$ from $\Delta\Delta S_{tot}$ suggests room for further methodological refinements.

The combination of high-resolution crystal structures, analyzed by ensemble refinement, NMR relaxation data, and MD simulations enable us to examine the structural origin of the thermodynamic differences outlined above. Differences in the interactions involving the hydroxyl group at the stereocenter of the diastereomers apparently lead to conformational strain and more pronounced conformational fluctuations in the S-stereoisomer at the RHS of the binding site, which couple with increased fluctuations of the surrounding protein. These results reinforce the notion that structure-based ligand design, when guided solely by static X-ray structures, addresses only one part of the picture and might be misleading.

In a broader perspective, improved knowledge about the sensitive interdependence of solvent entropy and protein conformational entropy adds to our understanding of molecular recognition. The phenomenon indicates both opportunities and challenges in rational drug design. On the one hand, contributions from solvation entropy to the free energy of binding are well-known, and the present results reiterate the concept of targeting individual water sites to achieve increased binding affinity.[12,94] On the other hand, efforts to design ligands that perturb the solvent structure around the binding site might not achieve the expected result due to changes in conformational entropy of the ligand and protein, as exemplified herein.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b11099.

Details of ligand synthesis and purification, including NMR and MS spectra; Tables detailing ligand force field parameters, X-ray crystallography data collection and refinements statistics, conformational entropy differences obtained from MD between the apo state and each of the R- and S-complexes; Figures showing replicate ITC data sets, a close-up view of the electron density at the stereocenter of the bound ligands, chemical shift changes upon ligand binding, and a comparison between GIST water densities and water molecules observed in the crystal structures (PDF)

AUTHOR INFORMATION

Corresponding Author

*mikael.akke@bpc.lu.se

ORCID

Hakon Leffler: 0000-0003-4482-8945
Derek T. Logan: 0000-0002-0098-8560
Ulf J. Nilsson: 0000-0001-5815-9522
Ulf Ryde: 0000-0001-7653-8489

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

Mikael Akke: 0000-0002-2395-825X

Author Contributions

[¶]MLV, OS, MMI, OC, MAO, and FM contributed equally.

Notes

The authors declare the following competing financial interest(s): UJN and HL are shareholders in Galecto Biotech AB, a company developing galectin inhibitors.

[▽]Francesco Manzoni passed away on March 12, 2017.

The crystal structures and diffraction data have been deposited in the Protein Data Bank with accession IDs 6QGE and 6QGF. The chemical shift assignments, relaxation rate constants, and order parameters have been deposited in the Biological Magnetic Resonance Bank (BMRB) under accession codes 27721 and 27722.

■ ACKNOWLEDGMENTS

We thank Ulrich Weininger and Göran Carlström for assistance with NMR experiments, Barbro Kahl-Knutson for help with FP experiments, Natalia Markova (Malvern) for helpful discussion on ITC measurements and analysis, and Malvern Instruments for access to ITC analysis software. Protein production was carried out by the Lund Protein Production Platform (LP3) at Lund University. High-field (900 MHz) NMR spectroscopy was carried out at the Swedish NMR Center at University of Gothenburg, supported by the Knut and Alice Wallenberg Foundation (NMR for Life). This work was supported by the Knut and Alice Wallenberg Foundation (KAW 2013.022), the Swedish Research Council (project 2014-5540 to UR), and the European Spallation Source (ERIC; UR and DL).

■ REFERENCES

(1) Akke, M.; Brüschweiler, R.; Palmer, A. G. NMR Order Parameters and Free Energy: An Analytical Approach and Its Application to Cooperative Ca^{2+} 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) MacRaild, C. A.; Daranas, A. H.; Bronowska, A.; Homans, S. W. Global Changes in Local Protein Dynamics Reduce the Entropic Cost of Carbohydrate Binding in the Arabinose-Binding Protein. *J. Mol. Biol.* **2007**, *368*, 822−832.

(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*, 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) Wienen-Schmidt, B.; Jonker, H. R. A.; Wulsdorf, T.; Gerber, H. D.; Saxena, K.; Kudlinzki, D.; Sreeramulu, S.; Parigi, G.; Luchinat, C.; Heine, A.; Schwalbe, H.; Klebe, G. Paradoxically, Most Flexible Ligand Binds Most Entropy-Favored: Intriguing Impact of Ligand Flexibility and Solvation on Drug-Kinase Binding. *J. Med. Chem.* **2018**, *61*, 5922−5933.

(14) Saraboji, K.; HÅkansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; Logan, D. T. 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.

(15) 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.

(16) Johannes, L.; Jacob, R.; Leffler, H. Galectins at a glance. *J. Cell Sci.* **2018**, *131*, 1−9.

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

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

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

(20) 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.

(21) 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.

(22) Delaine, T.; Collins, P.; MacKinnon, A.; Sharma, G.; Stegmayr, J.; Rajput, V. K.; Mandal, S.; Cumpstey, I.; Larumbe, A.; Salameh, B. A.; Kahl-Knutsson, B.; van Hattum, H.; van Scherpenzeel, M.; Peiters, R. J.; Sethi, T.; Schambye, H.; Oredsson, S.; Leffler, H.; Blanchard, H.; Nilsson, U. J. Galectin-3-Binding Glycomimetics That Strongly Reduce Bleomycin-Induced Lung Fibrosis and Modulate Intracellular Glycan Recognition. *ChemBioChem* **2016**, *17*, 1759−1770.

(23) Mandal, S.; Nilsson, U. J. Tri-Isopropylsilyl Thioglycosides as Masked Glycosyl Thiol Nucleophiles for the Synthesis of S-Linked Glycosides and Glyco-Conjugates. *Org. Biomol. Chem.* **2014**, *12*, 4816−4819.

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

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

(27) Salomonsson, E.; Larumbe, A.; Tejler, J.; Tullberg, E.; Rydberg, H.; Sundin, A.; Khabut, A.; Frejd, T.; Lobsanov, Y. D.; Rini, J. M.; Nilsson, U. J.; Leffler, H. Monovalent Interactions of Galectin-1. *Biochemistry* **2010**, *49*, 9518−9532.

(28) Ursby, T.; Unge, J.; Appio, R.; Logan, D. T.; Fredslund, F.; Svensson, C.; Larsson, K.; Labrador, A.; Thunnissen, M. M. The Macromolecular Crystallography Beamline I911−3 at the MAX IV Laboratory. *J. Synchrotron Radiat.* **2013**, *20*, 648−653.

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

(29) Kabsch, W. XDS. In *International Tables for Crystallography, Vol. F: Crystallography of Biological Macromolecules*; Rossmann, M. G., Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, 2010; Vol. *66*, pp 125−132.
(30) 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.
(31) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355−367.
(32) 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.
(33) 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*, 1074−1080.
(34) 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.
(35) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(36) 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.
(37) 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.
(38) 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.
(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) 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.
(44) 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.
(45) Loria, J. P.; Rance, M.; Palmer, A. G. A Relaxation-Compensated Carr-Purcell-Meiboom-Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* **1999**, *121*, 2331−2332.
(46) Mulder, F. A. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Measurement of Slow (Microseconds-Milliseconds) Time Scale Dynamics in Protein Side Chains by 15N Relaxation Dispersion NMR Spectroscopy: Application to Asn and Gln Residues in a Cavity Mutant of T4 Lysozyme. *J. Am. Chem. Soc.* **2001**, *123*, 967−975.
(47) Davis, D. G.; Perlman, M. E.; London, R. E. Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes via the T1r and T2 (CPMG) Methods. *J. Magn. Reson., Ser. B* **1994**, *104*, 266−275.
(48) Carver, J. P.; Richards, R. E. A General Two-Site Solution for the Chemical Exchange Produced Dependence of T2 upon the Carr-Purcell Pulse Separation. *J. Magn. Reson.* **1972**, *6*, 89−105.
(49) Palmer, A. G.; Kroenke, C. D.; Loria, J. P. Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. *Methods Enzymol.* **2001**, *339*, 204−238.
(50) 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.
(51) 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.
(52) 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.
(53) Mandel, A. M.; Akke, M.; Palmer, A. G. Backbone Dynamics of Escherichia Coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **1995**, *246*, 144−163.
(54) Mittermaier, A.; Kay, L. E. Measurement of Methyl H-2 Quadrupolar Couplings in Oriented Proteins. How Uniform Is the Quadrupolar Coupling Constant? *J. Am. Chem. Soc.* **1999**, *121*, 10608−10613.
(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) Jarymowycz, V. A.; Stone, M. J. Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chem. Rev.* **2006**, *106*, 1624−1671.
(59) Akke, M. Conformational Dynamics and Thermodynamics of Protein-Ligand Binding Studied by NMR Relaxation. *Biochem. Soc. Trans.* **2012**, *40*, 419−423.
(60) 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.
(61) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T E.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*; University of California: San Francisco, 2015.
(62) 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.
(63) Genheden, S.; Akke, M.; Ryde, U. Conformational Entropies and Order Parameters: Convergence, Reproducibility, and Transferability. *J. Chem. Theory Comput.* **2014**, *10*, 432−438.
(64) 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.
(65) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026

Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(66) 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.
(67) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(68) 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.
(69) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(70) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; 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, N. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2016.
(71) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996**, *60*, 1271−1277.
(72) *Turbomole* V7.01; Turbomole GmbH, 2007; http://www.turbomole.com.
(73) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. An Automatic Method to Generate Force-Field Parameters for Hetero-Compounds. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 274−289.
(74) Genheden, S.; Ryde, U. A Comparison of Different Initialization Protocols to Obtain Statistically Independent Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 187−195.
(75) 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.
(76) Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381* (3−4), 512−518.
(77) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(78) 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.
(79) 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.
(80) Edholm, O.; Berendsen, H. J. C. Entropy Estimation from Simulations of Non-Diffusive Systems. *Mol. Phys.* **1984**, *51*, 1011−1028.
(81) 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.
(82) King, B. M.; Tidor, B. MIST: Maximum Information Spanning Trees for Dimension Reduction of Biological Data Sets. *Bioinformatics* **2009**, *25*, 1165−1172.
(83) King, B. M.; Silver, N. W.; Tidor, B. Efficient Calculation of Molecular Configurational Entropies Using an Information Theoretic Approximation. *J. Phys. Chem. B* **2012**, *116*, 2891−2904.
(84) Fogolari, F.; Maloku, O.; Dongmo Foumthuim, C. J.; Corazza, A.; Esposito, G. PDB2ENTROPY and PDB2TRENT: Conformational and Translational-Rotational Entropy from Molecular Ensembles. *J. Chem. Inf. Model.* **2018**, *58*, 1319−1324.
(85) 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*, 973−980.
(86) Peterson, K.; Kumar, R.; Stenström, O.; Verma, P.; Verma, P. R.; Håkansson, M.; Kahl-Knutsson, B.; Zetterberg, F.; Leffler, H.; Akke, M.; Logan, D. T.; Nilsson, U. J. 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.
(87) Halle, B.; Wennerström, H. Interpretation of Magnetic Resonance Data from Water Nuclei in Heterogeneous Systems. *J. Chem. Phys.* **1981**, *75*, 1928−1943.
(88) 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.
(89) Schichman, S. A.; Amey, R. L. Viscosity and Local Liquid Structure in Dimethyl Sulfoxide-Water Mixtures. *J. Phys. Chem.* **1971**, *75*, 98−102.
(90) Catalan, J.; Diaz, C.; Garcia-Blanco, F. Characterization of Binary Solvent Mixtures of DMSO with Water and Other Cosolvents. *J. Org. Chem.* **2001**, *66*, 5846−5852.
(91) 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.
(92) Li, D. W.; Showalter, S. A.; Brüschweiler, R. Entropy Localization in Proteins. *J. Phys. Chem. B* **2010**, *114*, 16036−16044.
(93) Haider, K.; Cruz, A.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **2018**, *14*, 418−425.
(94) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.

DOI: 10.1021/jacs.8b11099
J. Am. Chem. Soc. 2019, 141, 2012−2026