Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3.

Diehl, Carl; Genheden, Samuel; Modig, Kristofer; Ryde, Ulf; Akke, Mikael

Published in:
Journal of Biomolecular NMR

DOI:
10.1007/s10858-009-9356-5

2009

Document Version:
Peer reviewed version (aka post-print)

Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3

Carl Diehl[a], Samuel Genheden[b], Kristofer Modig[a], Ulf Ryde[b,*], Mikael Akke[a,*]

[a] *Center for Molecular Protein Science, Biophysical Chemistry, Lund University, P. O. Box 124, SE-221 00 Lund, Sweden*

[b] *Theoretical Chemistry, Lund University, P. O. Box 124, SE-221 00 Lund, Sweden*

e-mail: mikael.akke@bpc.lu.se; ulf.ryde@teokem.lu.se

Abstract

The conformational entropy of proteins can make significant contributions to the free energy of ligand binding. NMR spin relaxation enables site-specific investigations of conformational entropy, via the local rank-2 tensor fluctuations that are parameterized in terms of order parameters. Here we have probed the conformational entropy of lactose binding to the carbohydrate recognition domain of galectin-3, a protein that plays an important role in cell growth, cell differentiation, cell cycle regulation, and apoptosis, making it a potential target for therapeutic intervention in inflammation and cancer. We used ^{15}N spin relaxation experiments and molecular dynamics simulations to monitor the backbone amides and secondary amines of the tryptophan and arginine side chains in the ligand-free and lactose-bound states of the galectin-3 carbohydrate binding domain. Overall, we observe good agreement between the experimental and computed order parameters of the ligand-free and lactose-bound states. Thus, the ^{15}N spin relaxation data indicate that the molecular dynamics simulations provide reliable information on the conformational entropy of the binding process. The molecular dynamics simulations reveal a correlation between the simulated order parameters and residue-specific backbone entropy, re-emphasizing that order parameters provide useful estimates of local conformational entropy. The present results show that the protein backbone exhibits an increase in conformational entropy upon binding lactose, despite the absence of structural changes.

Keywords: spin relaxation, order parameters, molecular dynamics simulations, ligand binding, entropy

Introduction

Conformational entropy potentially plays an important role in ligand binding to proteins (Cooper and Dryden 1984; Karplus and McCammon 2002). Nuclear spin relaxation is the only experimental technique available to date that can provide site-specific information on conformational entropy (Akke et al. 1993; Homans 2005; Igumenova et al. 2006). Previous NMR studies have indicated that the conformational fluctuations can either decrease or increase upon ligand binding, with the corresponding changes in entropy contributing unfavorably or favorably, respectively, to the binding free energy (Jarymowycz and Stone 2006). Hence, the dynamical response of the protein to ligand binding can tune the binding thermodynamics (Frederick et al. 2007).

Typically, NMR spectroscopic investigations of conformational dynamics focus on a relatively limited subset of bond vectors. Thus, investigations based solely on experimental data inevitably undersample the conformational entropy of the system. Interpretation of NMR order parameters in terms of entropy typically rely on simplifying assumptions, including the additivity of contributions from individual bond vectors (Akke et al. 1993). While it is still an open question to what extent intramolecular motions are correlated, results reported to date suggest that correlations introduce only a minor bias in the thermodynamic interpretation of conformational fluctuations (Prompers and Brüschweiler 2000; Prabhu et al. 2003). Recent work shows that side-chain dynamics must be probed not only by the terminal groups, but also by additional sites along the aliphatic chain to address the effect of motional decoupling of the different degrees of freedom (Trbovic et al. 2009). Thus, a comprehensive experimental investigation of conformational entropy would require an extensive set of data that together sample the backbone and side-chain dynamics using a number of different isotope labeling schemes (Lian and Middleton 2001; Lundström et al. 2007). To date, the most commonly used experiments probe the backbone or methyl groups by ^{15}N or ^{2}H relaxation (Palmer 2001; Igumenova et al. 2006), respectively, although approaches have been devised to monitor additional backbone and side-chain sites (Berglund et al. 1995; Boyd 1995; Yang et al. 1998; Jin et al. 2003; Zheng and Yang 2004; Teilum et al. 2006; Iwahara et al. 2007; Paquin et al. 2008).

Molecular dynamics (MD) simulations provide insights into motional mechanisms at the level of individual atoms. Thus, MD simulations in principle offer a route to interpret in great detail the results from spin relaxation experiments, provided that the two techniques yield commensurate results (Showalter and Brüschweiler 2007). In particular, MD simulations can provide the probability distribution of the conformational substates, including those degrees of freedom that are not probed by spin relaxation measurements. Once an MD-generated conformational ensemble has been validated by experimental NMR data, it is therefore possible to calculate the total conformational entropy of the system and to address other issues such as the degree of coupling between bond vector motions. In addition, MD simulations offer a high-resolution view of the motional mechanisms that cannot be determined directly from the NMR relaxation data. Here, we present a study along these lines towards determining the conformational entropy of ligand binding from the combination of ^{15}N relaxation data and MD simulations.

As a model system, we have studied the carbohydrate recognition domain of galectin-3 (Gal3). Galectins represent a family of proteins that preferentially bind β-galactoside-containing glycans composed of N-acetyllactosamine (Galβ1,4GlcNAc; LacNAc). The Gal3 structure consists of two anti-parallel β-sheets containing seven and five strands. The saccharide binding site is defined by a shallow groove formed by the seven-stranded β-sheet and surrounding loops (Fig. 1). Galectin–monosaccharide interactions are relatively weak, with dissociation constants on the order of 0.1–1 mM. The binding free energy is generally dominated by enthalpic contributions and has a minor unfavorable entropic contribution (Bachhawat-Sikder et al. 2001). Typically, two to five hydrogen bonds are formed between the carbohydrate ligand and Gal3, in addition to favorable van der Waals interactions. The binding site includes one tryptophan (W181) and two arginines (R162 and R186) whose side chains form close interactions with the bound ligand (Fig. 1). Both of these residue types have suitable ^{15}N spin probes in their side chains, the secondary amines (N^{ε}-H^{ε}) in the tryptophan indole and arginine guanidino groups, enabling us to probe detailed interactions in the ligand-binding site. The high-resolution crystal structures of apo (1.08 Å; to be published) and lactose-bound Gal3 (1.35 Å; PDB ID 2nn8, (Collins et al. 2007) are highly similar; the backbone RMSD is 0.22 Å and the

RMSD of all heavy atoms is 0.59 Å. The structural differences between these two states are limited to the ligand-binding region with minor changes observed only for the side chains of E184 and R186. Thus, the structural changes induced by ligand binding are very subtle, suggesting that lactose binds into an essentially preformed site. In this report we show that the protein backbone responds to ligand binding with a significant favorable change in conformational entropy, despite the absence of structural changes.

Results and discussion

The chemical shift differences between lactose-bound Gal3 (lac-Gal3) and apo-Gal3 in the ^{1}H-^{15}N HSQC spectrum are limited, with RMSD values of 0.02 ppm and 0.14 ppm for the backbone amide ^{1}H and ^{15}N resonances, respectively, indicating that the close structural similarity observed between the two crystal structures persists in solution. Weighted ^{1}H and ^{15}N chemical shift differences (Cavanagh et al. 2007) greater than 0.05 ppm are observed for 13 residues, which are located primarily across the saccharide binding β-sheet and in loop regions. Cross-peaks are observed for six out of the nine arginine $H^{\varepsilon}N^{\varepsilon}$ groups (residues 129, 162, 168, 183, and 224) in the spectrum of apo-Gal3, while five cross-peaks are observed for lac-Gal3 (129, 162, 168, 186, and 224). Upon binding of lactose the $H^{\varepsilon}N^{\varepsilon}$ cross-peak for R186 appears, R183 disappears and R162 shifts downfield in both dimensions with a weighted chemical shift difference of 0.11 ppm. The W181 $H^{\varepsilon}N^{\varepsilon}$ cross-peak is observed in both states, but has rather weak peak intensity, which makes for large uncertainties in the relaxation data.

Relaxation data

We acquired relaxation data comprising R_1, R_2, and $\{^{1}H\}$-^{15}N NOE at 14.1 T, and R_1 and R_2 at 11.7 T for both apo- and lac-Gal3 at a temperature of 28.0 ± 0.1 °C. Separate sets of experiments were conducted for the amide backbone, together with the tryptophan side chain indole, and the arginine side chain guanidino groups.

Relaxation rates could be measured for 112 and 118 backbone amides in the apo and lactose-bound states, respectively. Overall, the relaxation data are relatively homogeneous across the sequence. The trimmed mean values of the relaxation data are: R_1, 1.64 ± 0.05 s^{-1} (apo) and 1.47 ± 0.06 s^{-1} (lac); R_2, 10.2 ± 0.7 s^{-1} (apo) and 10.6 ± 0.7 s^{-1} (lac); $\{^{1}H\}$-^{15}N NOE, 0.78 ± 0.04 (apo) and 0.78 ±

0.07 (lac). Similar results are obtained at 11.7 T: R_1, 2.02 ± 0.08 s^{-1} (apo) and 1.85 ± 0.08 s^{-1} (lac); R_2, 9.6 ± 0.6 s^{-1} (apo) and 9.9 ± 0.7 s^{-1} (lac). The large majority of residues excluded by trimming is located in loops. The mean R_1 value is significantly greater for apo-Gal3 than for lac-Gal3, whereas the R_2 and NOE are both identical within errors.

Not surprisingly, the variability in the relaxation data is higher among the side chains than what is observed for the backbone. The relaxation rates of the arginine guanidino groups cover the following ranges at 14.1 T: R_1, 0.35–1.49 s^{-1} (apo) and 1.13–1.33 s^{-1} (lac); R_2, 3.2–15.3 s^{-1} (apo) and 5.5–10.6 s^{-1} (lac); $\{^1H\}$-^{15}N NOE, –2.60–1.18 (apo) and 0.11–0.84 (lac). The tryptophan indole ring is very flexible in the apo state, as observed from low R_1 and R_2 relaxation rates and a negative NOE, but as rigid as the backbone in the lactose-bound state.

NMR-derived order parameters

The diffusion tensor was optimized iteratively together with the model-free parameters for the internal dynamics. The best-fit diffusion tensor for apo-Gal3 is slightly asymmetric with a global correlation time of 7.2 ns, anisotropy of 1.07, and rhombicity of 0.51, while that for lac-Gal3 is axially symmetric with a global correlation time of 7.9 ns and an anisotropy of 1.10. These deviations from a spherical diffusion tensor are marginal and do not impact significantly on the fitted order parameters. Nonetheless, the diffusion tensors reported above were used in the model-free parameter optimization, since these models were selected in the fitting procedure.

The resulting order parameters for apo-Gal3 exhibit only modest variation along the sequence (Fig. 2a). The mean value is S^2 = 0.87 ± 0.05, including all backbone amides (N = 111), or 0.88 ± 0.03 for residues in β-strands only. Residues outside of secondary structure elements have a mean S^2 of 0.85 ± 0.07.

The order parameters for lac-Gal3 are lower and show a greater variation than those for apo-Gal3 with a mean of 0.83 ± 0.14 (N = 116), see Fig. 2b. Excluding residues 114, 124 and 142, which have very low order parameters, the mean is 0.84 ± 0.07. These three residues are located at the N-terminus (114) or in loops (124 and 142). Their amide resonances are not observed in apo-Gal3, presumably due to fast amide proton exchange with solvent. By comparison, the mean S^2 calculated for all β-strands is 0.86 ± 0.04, and that for residues outside of secondary structure elements is 0.75 ± 0.07.

The side-chain order parameters are quite diverse (Fig. 2c–d). For apo-Gal3, order parameters were obtained for R129, R162, R168, R183 and R224, as well as for W181. The order parameters indicate that some side chains are highly ordered (R162 and R224; $S^2 > 0.85$), whereas others are flexible (R129 and R168; $0.55 < S^2 < 0.75$), or highly disordered (R183 and W181; $S^2 < 0.1$), see Fig. 2c.

The side-chain order parameters of lac-Gal3 exhibit significantly less diversity than those for apo-Gal3. No side chain is highly disordered, but R129 exhibits some flexibility, while R162, R168, R186, R224 and W181 have order parameters comparable to the highly ordered backbone (Fig. 2d).

MD-derived order parameters

Order parameters of the N-H bond vectors were obtained from the MD simulation by fitting the autocorrelation function (ACF) of the N-H vector. The calculated order parameters are compared with the experimental values in Figs. 2 and 3a–b, revealing a good agreement in general. The RMSD between the experimental and simulated order parameters is 0.049 (apo) and 0.054 (lac; excluding the highly flexible residues 114, 124 and 141), which is on par with results from a previous bench-marking of state-of-the-art MD simulations against experimental NMR data for ubiquitin (Showalter and Brüschweiler 2007). Several of the largest deviations between the NMR- and MD-derived order parameters can be attributed to incomplete convergence of the ACFs after 1 ns. For instance, this holds for residues 114, 115, 165, 166, 225 and 227 in apo-Gal3 and residues 115, 116 and 232 in lac-Gal3; these residues are marked with boxes in Fig 3. We note that each of these residues has, or is proximal to another residue that has, two alternative conformations in the X-ray structure. Since the MD simulations were initiated from one of the two conformations, these observations suggest that the convergence problem can be rationalized by gradual, but incomplete, sampling of both conformations in the MD trajectories. The inferred relationship between the multiple conformations in the X-ray structures and the imperfect agreement between the experimental and simulated S^2 values will be subject to detailed investigations in future work.

In general, the S^2 values determined by fitting the ACF agree well with those calculated by the iRED approach (Prompers and Brüschweiler 2002), indicating that the overall and internal motions are separable; the RMSDs are 0.028 (apo) and 0.029 (lac) and the correlation coefficients are $r_C = 0.97$ (apo) and 0.92 (lac),

see Fig 3c–d. However, in the case of apo-Gal3 there is a minor systematic deviation between the S^2 values obtained by the iRED approach and by fitting the ACF (Fig. 3c). Paired *t*-tests indicate that the iRED-derived S^2 values are significantly different from the ACF-derived ones for apo-Gal3, $p = 4 \cdot 10^{-11}$, but that they agree with the experimental results, $p = 0.17$; conversely, the ACF results differ from the latter, $p = 0.001$. The situation is less clear-cut in the case of lac-Gal3, where the corresponding comparisons yield $p = 0.026$, 0.061, and 0.085. The deviations between the iRED and ACF results for apo-Gal3 are more pronounced for amides with lower values of S^2 (Fig. 3c), which are typically associated with more slowly decaying ACFs, possibly suggesting that imperfect convergence of the ACF leads to an unreliable fit of the plateau value for a larger number of residues than those identified above. By contrast, the iRED approach extracts S^2 from the eigenvalues of the covariance matrix describing the rank-2 tensor reorientations along the trajectory, and is therefore not subject to the type of fitting error that can affect the ACF approach. However, incomplete sampling of conformational space is a general problem that affects both approaches by introducing systematic deviations from the experimental data. The following analyses of backbone fluctuations primarily involve iRED-derived S^2 values, since these show the best agreement with the experimental results.

Order parameters for the arginine and tryptophan side chains are shown in Fig. 2c–d. The order parameters compare relatively well to the experimental ones, except for those of W181 in apo-Gal3, and R168 in both states. It is possible that the low order parameters observed experimentally originate from conformational changes that take place on a time scale of nanoseconds, which may not be adequately sampled by the 20 ns long MD simulations. In addition, the high-resolution X-ray structures of apo- and lac-Gal3 both include electron density for two different conformations of the R168 side chain.

Differences in order parameters between the apo and lactose-bound states

As outlined above, backbone order parameters for apo-Gal3 are slightly higher in general than those for lac-Gal3. Paired *t*-tests show that the decrease in S^2 upon lactose binding is significant, as gauged by both the experimental and simulated data, with $p = 0.3 \cdot 10^{-12}$ (NMR), 0.006 (iRED), and 0.05 (ACF), demonstrating that the backbone becomes more flexible upon ligand binding. The mean pairwise difference in order parameter between lac-Gal3 and apo-Gal3 is $\langle \Delta S^2 \rangle = -0.031 \pm$

0.004 (mean ± standard error of the mean), see Fig. 4a. The corresponding values obtained from the MD simulations, –0.017 ± 0.009 (iRED; Fig. 4b) and $<\Delta S^2>$ = –0.011 ± 0.009 (ACF), agrees qualitatively with the experimental value, although the residue-by-residue comparison reveals individual discrepancies (Figs. 4a–b) as outlined in the previous section. Both NMR and MD show that differences in order parameters between the two states are not localized to any specific region of the protein structure, but occur throughout the sequence. Residues that become significantly more flexible or rigid upon ligand binding are highlighted in Fig. 5.

The side-chain order parameters are similar for R129, R162, R168 and R224, of which only R162 is associated with ligand binding. The high S^2 for R162 is explained by the presence of a salt-bridge between the guanidino group of R162 and the carboxylate of E165 in the crystal structures of both apo- and lac-Gal3, which persists throughout the major part of the MD trajectory in both states. Unfortunately, of the other two side chains coordinating lactose, only W181 is observed in both states. W181 is significantly more ordered in lac-Gal3 than in apo-Gal3, as gauged from the experimental S^2 values. The poor agreement between the experimental and simulated order parameters rules out further quantitative interpretation in terms of entropy for this specific probe. On a qualitative level, however, the significant change in the experimental S^2 for the indole NH of W181 reveals an unfavorable entropic contribution to the free energy of lactose binding.

Conformational entropy

The conformational entropy was calculated from MD simulations using three different approaches based on dihedral angle fluctuations, quasi-harmonics, and normal modes. The results of these calculations are shown in Table 1. Even though there is a large difference between the absolute values, the estimated differences in entropy between apo-Gal3 and lac-Gal3 are similar for the dihedral and quasi-harmonic methods. By contrast, the normal-mode approach predicts a much smaller difference between the two states. The good agreement between the quasi-harmonic and dihedral analyses is somewhat unexpected. It is common that entropy estimates based on quasi-harmonic analysis do not converge even after very long simulations (Gohlke and Case 2004). Furthermore, the quasi-harmonic approach often overestimates the entropy since it enforces a harmonic approximation upon dihedrals that have several distinct conformations (i.e. local

minima), giving a wide and flat well, rather than several narrow ones (Chang et al. 2005). Only the dihedral analysis, Eqn [2], is expected to treat such rotamer averaging correctly. Estimating entropy based on dihedral angle fluctuations has the drawback that correlations between the various degrees of freedom are ignored. However, it has been found that the backbone entropy is relatively insensitive to motional correlations (Prompers and Brüschweiler 2000), so one might actually expect the dihedral analysis to provide the most accurate estimate. Normal-mode analysis, on the other hand, should underestimate the entropy of dihedral angles that undergo rotameric averaging, because it restricts the analysis to a single local minimum.

We observe an increase in the number of dihedrals that assume several conformations when lactose is bound from 544 for apo-Gal2 to 710 for lac-Gal3. This can explain the low estimate of the normal-mode analysis, because this approach neglects that a dihedral can have many energy minima. The similarity between the results of the quasi-harmonic and dihedral approaches, suggests that the former method actually gives reasonable estimates also for dihedrals with several distinct conformations.

The bootstrap standard errors are small (< 1%) for the entropies calculated from dihedral fluctuations. As shown in Fig. 6, the calculated entropies show reasonable convergence for both apo- and lac-Gal3, although the entropy of the latter state might be presumed to increase slightly for an extended trajectory.

The MD simulations indicate that approximately 23–36% of the conformational entropy originates from backbone fluctuations in Gal3 (Table 1). The relative backbone contribution obtained here is in general agreement with previous results for staphylococcal nuclease, 25% (Wrabl et al. 2000), and the molten-globule state of α-lactalbumin, 33% (Schafer et al. 2002). The sizeable backbone entropy implies that changes in backbone fluctuations between the free and bound states potentially can contribute significantly to the entropy of ligand binding. Indeed, backbone fluctuations contribute 20–28% of the difference in total conformational entropy between apo- and lac-Gal3 (Table 1).

The entropy difference between apo- and lac-Gal3 was estimated from the experimental order parameters using analytical relationships (Akke et al. 1993; Yang and Kay 1996), Eqn. [1]. The resulting value agrees well with that estimated from the backbone dihedral angle fluctuations (Table 1). The agreement can be

explained in part by the observation that the N–H bond vector motions are dominated by dihedral angle fluctuations, rather than by librations relative to the peptide plane (Buck and Karplus 1999). However, the entropy calculated from the MD-derived order parameters using Eqn. [1] is lower for both the ACF and iRED approaches, as expected from Fig. 4 and the results discussed above.

Figure 7 shows the backbone dihedral entropy plotted against the order parameter for all residues. The present plots reveal a clear correlation between the backbone order parameter and entropy, with $r_C = 0.86$ and 0.77 for apo- and lac-Gal3, respectively. Similar results have been reported recently for arginine guanidino groups (Trbovic et al. 2009). The good correlation is not necessarily expected because the order parameter reports on the N–H bond vector fluctuations, which might not depend directly on the $\psi(i-1)$, $\omega(i)$ and $\phi(i)$ dihedral angle fluctuations. Correspondingly, this caveat explains the scatter in the data of Fig. 7. The relationship observed for the backbone indicates that ^{15}N order parameters provide useful estimates of the backbone entropy beyond that associated with the N–H bond vector fluctuations.

Concluding remarks

The experimental and simulated results agree qualitatively, demonstrating that the change in conformational entropy of Gal3 contributes favorably to the binding of lactose. The MD simulations indicate that the backbone fluctuations contribute as much as one third of the total conformational entropy. Notably, the estimated conformational entropy is significantly greater than the total entropy determined by isothermal titration calorimetry (Table 1), implying that changes in the protein fluctuations make an important contribution to the free energy of ligand binding.

Materials and methods

Expression and purification

The galectin-3 carbohydrate recognition domain (Gal3; amino acid residues 113–250) was expressed as a thioredoxin fusion construct. The Gal3-thioredoxin plasmid was transformed into *E. coli* BL21 DE3 pLysS Star using electroporation. A single colony was used for inoculation of an overnight culture in LB medium. For expression of ^{15}N or ^{15}N/^{13}C-labeled protein, minimal M-9 medium was used with autoclaved tap water. The culture was started by adding 1% (vol) of overnight culture to pre-warmed medium. The bacterial growth was monitored by

absorbance at 600 nm (A_{600}). At A_{600} ~ 0.6, the culture was induced by adding 0.4 mM IPTG and harvested by centrifugation after 3 hours. The cells were resuspended in MEPBS pH 7.2 (100 mM $NaHPO_4$, 100 mM NaCl, 2 mM beta-mercaptoethanol, 4 mM EDTA) and frozen overnight at –80 °C. The cells were thawed and sonicated 5×1 min, followed by centrifugation and collection of the supernatant.

The supernatant was loaded onto an equilibrated lactosyl sepharose column, washed with MEPBS pH 7.2, and eluted with 200 mM lactose (Massa et al. 1993). The eluted Gal3-thioredoxin fusion protein was cleaved using recombinant enterokinase (Novagen) in MEPBS pH 7.2 at 37 °C. The cleavage reaction was monitored by Phast Gel (GE). The cleavage product was dialyzed twice overnight against doubly distilled water. The dialysate was repurified on a lactosyl sepharose column. The purity of Gal3 was confirmed by SDS-PAGE. Typical yields ranged between 25–50 mg/l culture. The extinction coefficient at 280 nm was determined to 11,000 $cm^{-1}mol^{-1}$ using amino acid analysis.

NMR sample preparation

Lactose-bound Gal3 samples were prepared by concentration and buffer exchange into 5 mM HEPES pH 7.4, 200 mM lactose using Vivaspin-20 concentrators (M_w cutoff 10 kDa). The concentration of the ^{15}N-labeled sample was 0.4 mM, while that of the $^{15}N/^{13}C$ sample was 2.2 mM, as determined by absorbance. Apo-Gal3 samples were prepared as the lac-Gal3 samples, except that lactose was not added and repeated buffer exchange was performed until the concentration of lactose (remaining from the purification step) was less than 1 μM. The concentrations of the ^{15}N and $^{15}N/^{13}C$ samples were 0.4 mM and 0.2 mM, respectively. All NMR samples included small amounts of NaN3, DSS, and ^2H_2O.

Chemical shift assignments

Assignments of lac-Gal3 have been reported previously for a Gal3 construct covering residues 107–250 (Umemoto and Leffler 2001). The present assignments were performed de novo using standard approaches (Cavanagh et al. 2007) based on the following 3D experiments acquired at 11.7 T: HNCA (Kay et al. 1990; Farmer et al. 1992; Grzesiek and Bax 1992), HNCOCA (Bax and Ikura 1991; Grzesiek and Bax 1992), HNCACB (Wittekind and Mueller 1993), CBCACONH (Grzesiek and Bax 1992), and CCONH (Farmer and Venters 1995). The backbone chemical shifts of apo-Gal3 are similar to those of lac-Gal3, and were transfered

based on HNCA and CCONH experiments. Arginine $N^{\varepsilon}H^{\varepsilon}$ and $N^{\eta}H^{\eta}$ resonances were assigned using CCC-TOCSY-NEHE and HHNZCZHE experiments (Yamazaki et al. 1995), respectively, with the ^{15}N carrier frequency set to 85 ppm in both cases. Spectra were processed using nmrPipe (Delaglio et al. 1995). The processing protocol involved a solvent filter, squared cosine apodization functions, zero filling to twice the number of increments in all dimensions, and baseline correction in the ^{1}H dimension. Assignments were carried out using the CcpNmr suite (Vranken et al. 2005). Chemical shift changes between apo- and lac-Gal3 were evaluated as the weighted ^{1}H and ^{15}N chemical shift differences, $([\Delta\delta(^{1}H)]^2 + [0.1\Delta\delta(^{15}N)]^2)^{1/2}$ (Cavanagh et al. 2007).

Relaxation experiments and data analysis

R_1, R_2, and $\{^{1}H\}$-^{15}N NOE experiments (Farrow et al. 1994; Cavanagh et al. 2007) targeting the ^{15}N spins of the backbone and tryptophan side chain indole were performed on apo- and lac-Gal3 with the carrier placed in the center of the backbone amide region. Typically, 8–12 data points were acquired interleaved with relaxation delays ranging between 0–1 s (R_1) and 0–0.192 s (R_2), and a delay between experiments of 2.0 s. $\{^{1}H\}$-^{15}N heteronuclear NOEs were measured in an interleaved manner using a ^{1}H saturation time of 5 s and a recycle delay of 10 s between acquisition and the first ^{15}N pulse in both the NOE and control experiments. The spectral widths were 8013 Hz and 1835 Hz in the ^{1}H and ^{15}N dimensions, respectively, covering 1024 and 128 points. Experiments targeting the arginine side chain guanidino groups were performed essentially as described for the backbone, except that the ^{15}N carrier was placed at 72 ppm, and the spectral widths were 8013 Hz (1024 points) and 2880 Hz (128 points). The recycle delay in the NOE experiment was 15 s in order to allow full relaxation of flexible side chains.

All spectra were processed using two different apodization schemes. Most peaks were resolved using squared cosine window functions in both dimensions, while a subset of peaks benefited from a cosine window function in indirect dimension. In the case of the arginine side chains in lac-Gal3, the high-resolution apodization scheme involved Lorentzian-to-Gaussian window functions in both dimensions with a inverse exponential width of 5.0 Hz and a Gaussian width of 5.0 Hz. Peak intensities were measured as the summed signal in windows of 5×3 ($^{1}H/^{15}N$) points centered on the peak. The signal-to-noise ratio (S/N) was

estimated by calculating the standard deviation of 200 samples of integrated 5×3-point windows in empty regions of each spectrum. Mono-exponential functions were fitted to the R_1 and R_2 decays using the Levenberg–Marquardt minimization routine (Press et al. 1986) implemented in C-programs developed in-house. Errors in the fitted parameters were estimated from 1000 synthetic data sets created using Monte-Carlo simulations (Press et al. 1986; Mandel et al. 1995). NOEs were calculated as the ratio between peak intensities in the saturated and unsaturated experiments. The S/N was estimated as described above and the errors in the NOEs were determined by error propagation. The trimmed mean and standard deviation were calculated for each dataset of R_1, R_2 and NOE, where residues outside of two standard deviations were excluded in a single pass.

Model-free optimization

The global correlation time was initially estimated by fitting the R_2/R_1 ratio to an isotropic tensor using in-house routines implemented in Matlab (The Mathworks). Model-free parameters were fitted using the program suite relax (d'Auvergne and Gooley 2008) with in-house modified scripts. Four different diffusion tensors (spherical, oblate, prolate, and asymmetric) were considered. In each case, the optimization was performed iteratively by fixing the diffusion tensor and optimizing five different models of the local motion for each residue, as described (d'Auvergne and Gooley 2008). Next, the parameters describing the diffusion tensor and local motion were optimized simultaneously. The optimized parameters were taken as input for the next round of optimization. This procedure was iterated until the diffusion tensor and model-free parameters converged. Following convergence of the diffusion tensor parameters, the Akaike information criterion (AIC) was used to select models (Akaike 1974). Errors in the fitted model-free parameters were estimated using the Monte-Carlo approach with 1000 synthetic data sets and a fixed diffusion tensor.

In the case of the arginine and tryptophan side chain $H^{\varepsilon}N^{\varepsilon}$ groups, only the local motional models were fitted to each residue, while the diffusion tensor parameters were fixed at those arrived at in the model-free analysis of the backbone relaxation data.

The model-free analysis used an N-H bond length of 1.02 Å and a CSA of –172 ppm for the ^{15}N backbone spins (Kroenke et al. 1999). The CSA values of the

N^{ε} spins were set to –114 ppm for arginine (Trbovic et al. 2009) and 89 ppm for tryptophan (Ramamoorthy et al. 1997).

Calculation of entropy from NMR order parameters

The difference in conformational entropy between the two states was estimated from the experimental S^2 values using expressions presented previously (Akke et al. 1993; Yang and Kay 1996). As expected, these treatments yielded identical results within errors. The values reported herein were calculated as (Yang and Kay 1996)

$$\Delta S_{BA} = k\sum_{j=1}^{N} \ln\left[\frac{3-(1+8S_{j,B})^{1/2}}{3-(1+8S_{j,A})^{1/2}}\right] \qquad [1]$$

where ΔS_{BA} is the change in entropy upon going from state A to state B, $S_{j,A}$ is the order parameter of residue j in state A, N is the number of residues observed, and k is Boltzmann's constant.

MD simulations

The MD simulations were based on the crystal structure of Gal3 in complex with lactose (PDB code 2nn8 (Collins et al. 2007)). The lactose molecule was described by the glycam06 force field (Kirschner et al. 2008) and the protein by the AMBER99-SB force field (Hornak et al. 2006). The apo state was obtained by removing the lactose molecule from the crystal structure. The close similarity between the X-ray structures of lactose-bound (Collins et al. 2007) and apo states (to be published) rationalize this approach, which yielded better agreement with the NMR data than did MD simulations initiated from the lower-resolution NMR structure of the apo state (Umemoto et al. 2003). In those cases where alternative conformations are reported for individual residues in the PDB structure, we consistently used conformer A (the two conformations have the same occupancy in all cases). Protons were added with the leap module of AMBER and the protonation state was selected to mimic an apparent pH of 7, as suggested by the PROPKA software (Li et al. 2005). The protonation of His residues was selected by a detailed study of the solvent accessibility, the hydrogen-bond pattern, and the local surroundings. This indicated that H158 is protonated on the $N^{\delta 1}$ atom, whereas H208, H217 and H223 are protonated on the $N^{\varepsilon 2}$ atom. The total charge of the protein was +4. The protein was solvated in an octahedral box of TIP4P-Ewald water molecules (Horn et al. 2004), extending at least 9 Å from the protein.

All simulations were run using the AMBER 10 sander module (Case et al. 2008) with a 2 fs time step. The SHAKE algorithm (Ryckaert et al. 1977) was used to constrain bonds involving hydrogen atoms. The temperature was kept constant at 300 K using Langevin dynamics (Wu and Brooks 2003) with a collision frequency of 2.0 ps^{-1}. Electrostatics were calculated using particle-mesh-Ewald summation (Darden et al. 1993), with a fourth-order B-spline interpolation and a tolerance of 10^{-5}. The non-bonded cut-off was 8 Å and the non-bonded pair list was updated every 50 fs.

The system was energy minimized for 1000 steps, keeping all atoms except water molecules and hydrogen atoms restrained to their crystal positions with a harmonic force constant of 418 $kJ \cdot mol^{-1} \cdot Å^{-2}$. This was followed in turn by: 20 ps molecular dynamics equilibration at constant pressure, while retaining the restraints; 50 ps equilibration at constant pressure, without restraints; 200 ps equilibration at constant volume, without restraints; and finally 20 ns production run with coordinates saved every ps. Stability of the backbone atoms was monitored by calculating the root-mean-square deviation relative to the first trajectory. The trajectories stabilized after about 5 ns of the production run, and therefore only the last 15 ns were used in subsequent calculations.

Order parameters estimated from the MD simulation

Order parameters were estimated by first calculating the time autocorrelation function (ACF) of the vector of interest, using the AMBER 10 ptraj module. The values of the ACF are correlated with each other, even at large time windows, owing to the finite sampling time. Therefore, only a fraction of the autocorrelation is useful in the calculation (Zwansig and Ailawadi 1969; Lu and Bout 2006; Madsen 2008). Hence, a 1 ns long time window was used to extract the ACF from the last 15 ns of the simulation. Also a 1.5 ns long time window was tested, but the ACF turned out to be slightly more unstable in this case. The overall tumbling was removed by fitting the backbone heavy atoms to the first snapshot. The order parameters were obtained by fitting the ACF to an exponential function of the form

[2]

where $\{x\}$ is the autocorrelation values and *A, B, C, D* and *E* are fitted coefficients (Lipari and Szabo 1982). The order parameter can be identified with the plateau value *A* of the exponential (Buck and Karplus 1999; Case 2002). Statistical errors

of the order parameters were estimated using a bootstrap procedure on the residuals from the exponential fit, using 1000 samples (Efron and Tibshirani 1986). Order parameters were also extracted using the iRED approach (Prompers and Brüschweiler 2002).

Calculation of entropy from the MD simulations

Three different approaches were used to calculate the entropy of the protein . In the first approach, we transformed the Cartesian coordinates to internal coordinates (bond–angle–torsion coordinates) and calculated the entropy for each of these coordinates by the statistical mechanical formula (Edholm and Berendsen 1984; Chang et al. 2008; Trbovic et al. 2009)

$$S = -R\sum p \ln p \qquad [3]$$

where the sum is over all possible states of a coordinate and p is the probability of that state. The probability distribution function was estimated by histograms. For bonds and angles, the number of bins was 50 and the bin size was chosen to fit the largest range for any bond or angle. For the dihedrals, the number of bins was 72, giving a bin size of 5°. The dihedral entropy was normalized to that of a free rotor. Hence, $R/2 - R \ln n$ was added to the entropy in Eq. [3], where R is the gas constant and n is the number of bins (note that this additive constant will cancel when taking the difference between apo- and lac-Gal3). Several numbers of bins from 3 to 120 were tested, including the average optimum number (Shimazaki and Shinomoto 2007), which was determined to 111. However, the results converged after 24 bins, indicating that the results are insensitive to the actual bin size. The calculations showed that the contributions of the bond and angle fluctuations to the entropy were negligible, as observed before (Chang et al. 2008; Trbovic et al. 2009). Therefore, these terms will not be discussed further. Errors were estimated with a bootstrap algorithm, using 1000 samples (Efron and Tibshirani 1986).

Second, we used the quasi-harmonic approach (Karplus and Kushick 1981), which assumes that the probability distribution is a multivariate Gaussian distribution. The normal-mode frequencies, ω, were calculated from the determinant

$$\det\left(M^{1/2} s M^{1/2} - \frac{kT}{w^2} 1 \right) = 0 \qquad [4]$$

where M is a diagonal matrix with atom masses on the diagonal, σ is the covariance matrix, k is Boltzmann's constant, T is the absolute temperature (300 K in the present case) and ν are the harmonic frequencies. The covariance matrix, σ, is defined as

$$\sigma_{ij} = \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle \quad [5]$$

where $\{x_i\}$ are the Cartesian coordinates of the system. Once the frequencies are obtained, the entropy is calculated within the harmonic oscillator approximation (Hill 1986; Carlsson and Aqvist 2005). Prior to the calculations, translation and rotation were removed by fitting the backbone atoms to the first snapshot.

The third approach was standard normal-mode analysis (Kollman et al. 2000). Snapshots of the protein were extracted every ten picosecond, waters were stripped off, and the remaining structure was minimized. To ensure that the protein structures do not change significantly during the energy minimization, a fixed buffer region of the water molecules closest to the protein were included in the minimization. The number of water molecules was 300, which corresponds to all waters within 3 Å from the protein. For lac-Gal3 the lactose molecule was also included in the buffer region. The buffer atoms were kept fixed while the protein was minimized. Normal-mode frequencies of the minimized protein were then calculated by the AMBER nmode module and these were used to calculate the entropy.

Acknowledgments

This work was supported by the Swedish Research Council (MA, UR), The Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (MA), and the FLÄK Research School for Pharmaceutical Sciences at Lund University (MA, UR). Computer resources were provided by Lunarc at Lund University and HPC2N at Umeå University. We thank Hakon Leffler for the plasmid harboring the Gal3-thioredoxin fusion construct, and HL, Ulf Nilsson, and Gunnar Karlström for discussions.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Control AC-19: 716-723

Akke M, Brüschweiler R, Palmer AG (1993) NMR order parameters and free energy: An analytical approach and its application to cooperative Ca^{2+} binding by calbindin D_{9k}. J Am Chem Soc 115: 9832-9833

Bachhawat-Sikder K, Thomas CJ, Suriola A (2001) Thermodynamic analysis of the binding of galactose and poly-N-acetyllactoseamine derivatives to human galectin-3. FEBS Lett 500: 75-79

Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the α-carbon of the preceding residue in uniformly15N/13C enriched proteins. J Biomol NMR 1: 99-104

Berglund H, Baumann H, Knapp S, Ladenstein R, Härd T (1995) Flexibility of an arginine side chain at a DNA-protein interface. J Am Chem Soc 117: 12883-12884

Boyd J (1995) Measurement of ^{15}N relaxation data from the side chains of asparagine and glutamine residues in proteins. J Magn Res, B 107: 279-285

Buck M, Karplus M (1999) Internal and overall peptide group motion in proteins: molecular dynamics simulations for lysozyme compared with results from X-ray and NMR spectroscopy. J Am Chem Soc 121: 9645-9658

Carlsson J, Aqvist J (2005) Absolute and relative entropies from computer simulation with applications to ligand binding. J Phys Chem B 109: 6448-6456

Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. Acc Chem Res 35: 325-31

Case DA, Darden TA, Cheatham I, T.E. , Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu XW, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER 10. University of California, San Francisco.

Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (2007) Protein NMR Spectroscopy: principles and practice. Second edn. Elsevier Academic Press, San Diego

Chang CE, Chen W, Gilson MK (2005) Evaluating the accuracy of the quasiharmonic approximation. J Chem Theory Comput 1: 1017-1028

Chang CEA, McLaughlin WA, Baron R, Wang W, McCammon JA (2008) Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association. Proc Natl Acad Sci USA 105: 7456-7461

Collins PM, Hidari KIPJ, Blanchard H (2007) Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: possible implications for ligand-exchange protocols. Acta Crystallogr D 63: 415-419

Cooper A, Dryden DTF (1984) Allostery without conformational change. A plausible model. Eur Biophys J 11: 103-109

d'Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. J Biomol NMR 40: 107-119

Darden T, York D, Pedersen L (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J Chem Phys 98: 10089-10092
Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6: 277-293
Edholm O, Berendsen HJC (1984) Entropy Estimation from Simulations of Non-Diffusive Systems. Mol Phys 51: 1011-1028
Efron B, Tibshirani R (1986) Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statist sci 1: 54-77
Farmer BT, Venters RA (1995) Assignment of Side-Chain 13C Resonances in Perdeuterated Proteins. J Am Chem Soc 117: 4187-4188
Farmer BT, Venters RA, Spicer LD, Wittekind MG, Muller L (1992) A Refocused and Optimized Hnca - Increased Sensitivity and Resolution in Large Macromolecules. J Biomol NMR 2: 195-202
Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33: 5984-6003
Frederick KK, Marlow MS, Valentine KG, Wand AJ (2007) Conformational entropy in molecular recognition by proteins. Nature 448: 325-330
Gohlke H, Case DA (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 25: 238-250
Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. Journal of Magnetic Resonance 96: 432-440
Hill TL (1986) An Introduction to Statistical Thermodynamics. Dover Publications, Inc., New York
Homans SW (2005) Probing the binding entropyof ligand-protein interactions by NMR. Chembiochem 6: 1585-1591
Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120: 9665-78
Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65: 712-25
Igumenova TI, Frederick KK, Wand AJ (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. Chem Rev 106: 1672-1699
Iwahara J, Jung YS, Clore GM (2007) Heteronuclear NMR spectroscopy for lysine NH3 groups in proteins: Unique effect of water exchange on N-15 transverse relaxation. J Am Chem Soc 129: 2971-2980
Jarymowycz VA, Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem Rev 106: 1624-1671
Jin C, Prompers JJ, Brüschweiler R (2003) Cross-correlation suppressed T1 and NOE experiments for protein side-chain 13CH2 groups. J Biomol NMR 26: 241-247
Karplus M, Kushick JN (1981) Method for Estimating the Configurational Entropy of Macromolecules. Macromolecules 14: 325-332
Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9: 646-652

Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. Journal of Magnetic Resonance 89: 496-514

Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. J Comput Chem 29: 622-655

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, 3rd (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33: 889-97

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 14: 51-55

Kroenke CD, Rance M, Palmer AG (1999) Variability of the 15N chemical shift anisotropy in *Escherichia coli* ribonuclease H in solution. J Am Chem Soc 121: 10119-10125

Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins 61: 704-21

Lian L-Y, Middleton DA (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. Progr NMR Spectrosc 39: 171-190

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity. J Am Chem Soc 104: 4546-4559

Lu C-Y, Bout DAV (2006) Effect of finite trajectory length on the correlation function analysis of single molecule data. J Chem Phys 125: 124701-124709

Lundström P, Teilum K, Carstensen T, Bezsonova I, Wiesner S, Hansen F, Religa TL, Akke M, Kay LE (2007) Fractional 13C enrichment of isolated carbons using [1-13C]- or [2-13C]-glucose facilitates the accurate measurement of dynamics at backbone Ca and side-chain methyl positions in proteins. J Biomol NMR 38: 199-122

Madsen H (2008) Time Series Analysis. Chapman & Hall/CRC, New York

Mandel AM, Akke M, Palmer AG (1995) Backbone dynamics of *Eschericia coli* Ribonuclease HI: Correlations with structure and function in an active enzyme. J Mol Biol 246: 144-163

Massa SM, Cooper DN, Leffler H, Barondes SH (1993) L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity. Biochemistry 32: 260-7

Palmer AG (2001) NMR probes of molecular dynamics: overview and comparison with other techniques. Annu Rev Biophys Biomol Struct 30: 129-155

Paquin R, Ferrage F, Mulder FAA, Akke M, Bodenhausen G (2008) Multiple-Timescale Dynamics of Side-Chain Carboxyl and Carbonyl Groups in Proteins by C-13 Nuclear Spin Relaxation. J Am Chem Soc 130: 15805-+

Prabhu NV, Lee AL, Wand AJ, Sharp KA (2003) Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. Biochemistry 42: 562-570

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical Recipes. The Art of Scientific Computing. Cambridge University Press, Cambridge

Prompers JJ, Brüschweiler R (2000) Thermodynamic interpretation of NMR relaxation parameters in proteins in the presence of motional correlations. J Phys Chem B 104: 11416-11424

Prompers JJ, Brüschweiler R (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. J Am Chem Soc 124: 4522-4534

Ramamoorthy A, Wu CH, Opella SJ (1997) Magnitudes and orientations of the principal elements of the H-1 chemical shift, H-1-N-15 dipolar coupling, and N-15 chemical shift interaction tensors in N-15(epsilon 1)-tryptophan and N-15(pi)-histidine side chains determined by three-dimensional solid-state NMR spectroscopy of polycrystalline samples. J Am Chem Soc 119: 10479-10486

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comp Phys 23: 327-341

Schafer H, Smith LJ, Mark AE, van Gunsteren WF (2002) Entropy calculations on the molten globule state of a protein: Side-chain entropies of alpha-lactalbumin. Proteins 46: 215-224

Shimazaki H, Shinomoto S (2007) A method for selecting the bin size of a time histogram. Neural Comput 19: 1503-1527

Showalter SA, Brüschweiler R (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. J Chem Theory Comput 3: 961-975

Teilum K, Brath U, Lundström P, Akke M (2006) Biosynthetic 13C labeling of aromatic side-chains in proteins for NMR relaxation measurements. J Am Chem Soc 128: 2506-2507

Trbovic N, Cho J-H, Abel R, Friesner RA, Rance M, Palmer AG (2009) Protein side-chain dynamics and residual conformational entropy. J Am Chem Soc 131: 615-622

Umemoto K, Leffler H (2001) Assignment of 1H, 15N and 13C resonances of the carbohydrate recognition domain of human galectin-3. J Biomol NMR 20: 91-92

Umemoto K, Leffler H, Venot A, Valafar H, Prestegard JH (2003) Conformational differences in liganded and unliganded states of Galectin-3. Biochemistry 42: 3688-3695

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59: 687-696

Wittekind M, Mueller L (1993) 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 Res, B 101: 201-205

Wrabl JO, Shortle D, Woolf TB (2000) Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: Molecular dynamics simulations of native and denatured staphylococcal nuclease. Proteins 38: 123-133

Wu XW, Brooks BR (2003) Self-guided Langevin dynamics simulation method. Chem Phys Lett 381: 512-518

Yamazaki T, Pascal SM, Singer AU, Formankay JD, Kay LE (1995) Nmr Pulse Schemes for the Sequence-Specific Assignment of Arginine Guanidino N-15 and H-1 Chemical-Shifts in Proteins. J Am Chem Soc 117: 3556-3564

Yang D, Kay LE (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. J Mol Biol 263: 369-382

Yang D, Mittermaier A, Mok Y-K, Kay LE (1998) A study of protein side-chain dynamics from new ^{2}H auto-correlation and ^{13}C cross-correlation NMR experiments: Application to the N-terminal SH3 domain from drk. J Mol Biol 276: 939-954

Zheng Y, Yang D (2004) Measurement of dipolar cross-correlation in methylene groups in uniformly 13C-, 15N-labeled proteins. J Biomol NMR 28: 103-116

Zwansig R, Ailawadi NK (1969) Statistical Error Due to Finite Time Averaging in Computer Experiments. Phys Rev 1982: 280-282

Table 1. Conformational entropies (J/mol/K)

Method	Total Entropy[a]			Backbone Entropy		
	lac-Gal3	apo-Gal3	ΔS	lac-Gal3	apo-Gal3	ΔS
Dihedral[b]	–23347 ± 2	–24228 ± 2	941 ± 3	–5457 ± 1	–5644 ± 1	187 ± 1
QH[c]	31351	30488	863	11253	11008	245
NM[d]	21252 ± 1	21233 ± 1	18 ± 2			
S^2 ACF[e]						61 ± 22
S^2 iRED[f]						106
S^2 NMR[g]						204 ± 11
ITC[h]			–15.6			

[a]Total entropy refers to the total conformational entropy of the protein, except in the case of that listed for the ITC metod, which is the total entropy of lactose binding to Gal2. [b]Dihedral, Eqn. [3] normalized to the entropy of a free rotor; [c]QH, quasi-harmonic; [d]NM, normal mode; [e]S^2 ACF, simulated S^2 obtained by fitting the autocorrelation function; [f]S^2 iRED, simulated S^2 obtained by the iRED approach; [g]S^2 NMR, experimental S^2; [h]ITC, total entropy obtained by isothermal titration calorimetry (Bachhawat-Sikder et al. 2001).

Figure Legends

Fig. 1 Structure of lactose-bound Gal3. The backbone trace is shown in ribbon representation, while side chains ligating lactose are shown as sticks, and lactose is shown as ball-and-stick. The ligand-binding site is defined by the shallow groove formed by the seven-stranded β-sheet consisting of residues 118-121 (β1), 233-238 (β11), 145-151 (β3), 154-162 (β4), 170-177 (β5), 180-181 (β6), and 185-187 (β7). The side chains directly ligating lactose are His158, Asn160, Arg162, Val172, Asn174, Trp181, Glu184 and Arg186. The opposite five-membered sheet consists of residues 216-222 (β10), 208-213 (β9), 197-204 (β8), 130-138 (β2), and 240-249 (β12). The figure was prepared using Molmol (Koradi et al. 1996).

Fig. 2 Model-free order parameters of Gal3 determined by ^{15}N relaxation experiments (black boxes) and MD simulations (grey bars). Backbone S^2 for (a) apo-Gal3 and (b) lac-Gal3. Arginine and tryptophan side chain S^2 for (c) apo-Gal3 and (d) lac-Gal3. Asterisks identify simulated data that suffer from incompletely converged ACFs. Error bars were estimated as described in Materials and methods.

Fig. 3 Comparison of order parameters. (a–b) Experimental versus simulated order parameters for (a) apo-Gal3 and (b) lac-Gal3. Simulated S^2 values were obtained using the ACF method. S^2 values are shown for backbone (filled circles) and side chain (empty circles) N–H vectors. (c–d) ACF-derived versus iRED-derived S^2 for (c) apo-Gal3 and (d) lac-Gal3. S^2 values are shown for backbone N–H vectors only. Boxes identify outliers that suffer from incompletely converged ACFs. Error bars were estimated as described in Materials and methods.

Fig. 4 Difference in order parameters between lac-Gal3 and apo-Gal3, $\Delta S^2 = S^2_{lac} - S^2_{apo}$, determined by (a) ^{15}N relaxation experiments and (b) MD simulations using the iRED approach.

Fig. 5 Structural location of residues that show differences in the experimental S^2 values between lac-Gal3 and apo-Gal3, $\Delta S^2 = S^2_{lac} - S^2_{apo}$. The color ranges from cyan to blue with increasingly negative ΔS^2 values, and from magenta to red with increasingly positive ΔS^2. Gray color indicates residues for which there is no data

or ΔS^2 is less than one standard deviation. The figure was prepared using Molmol (Koradi et al. 1996).

Fig. 6 Convergence of the conformational entropy calculated from MD simulations. The total conformational entropy calculated from dihedral angle fluctuations is plotted versus the simulation time for apo- (full line) and lac-Gal3 (broken line). S_{dihed} was evaluated at every nanosecond.

Fig. 7 Backbone conformational entropy versus order parameter for (a) apo-Gal3 and (b) lac-Gal3. The entropy was calculated from dihedral fluctuations. Each data point represents the sum of the entropies associated with the $\psi(i-1)$, $\omega(i)$, and $\phi(i)$ dihedrals plotted versus $S^2(i)$, where i denotes the residue number. S^2 was calculated using the iRED approach.