Using Molecular Dynamics Simulations To Provide New Insights into Protein Structure on the Nanosecond Timescale: Comparison with Experimental Data and Biological Inferences for the Hyaluronan-Binding Link Module of TSG-6

Andrew Almond,*,† Charles D. Blundell,†,§ Victoria A. Higman,§
Alexander D. MacKerell, Jr.,‖ and Anthony J. Day‡,§

Manchester Interdisciplinary Biocentre, Faculty of Life Sciences, University of Manchester, Princess Street, Manchester M1 7DN, U.K., Michael Smith Building, Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, U.K., MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and Department of Pharmaceutical Chemistry, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201

Abstract: Link module domains play an essential role in extracellular matrix assembly and remodeling by binding to the flexible glycosaminoglycan hyaluronan. A high-resolution NMR-structure of the Link module from the protein product of tumor necrosis factor-stimulated gene-6 (Link_TSG6) has been determined, but a fuller appreciation of protein dynamics may be necessary to understand its hyaluronan-binding. Therefore, we have performed a 0.25 μs MD simulation, starting from the lowest-energy NMR-derived solution structure of Link_TSG6, with explicit water and ions, using the CHARMM22 protein force field. The simulation was as good a fit to the NMR data as the ensemble from simulated annealing, except in the $\beta5$-$\beta6$ loop. Furthermore, analysis revealed that secondary structure elements extended further than previously reported and underwent fast picosecond time scale dynamics, whereas nanosecond dynamics was found in certain loops. In particular, surface side chains proposed to interact with glycosaminoglycans were predicted to be highly mobile and be directed away from the protein surface. Furthermore, the hyaluronan-binding $\beta4$-$\beta5$ loop remained in a closed conformation, favoring an allosteric interaction mechanism. This enhanced view of the Link module provides general insight into protein dynamics and may be helpful for understanding the dynamic molecular basis of tissue assembly, remodeling, and disease processes.

1. Introduction

Submicrosecond conformational rearrangements in proteins provide the specificity and plasticity required to associate with other molecules.[1−4] Unfortunately, the two principal techniques used for defining protein structure at atomic resolution, X-ray crystallography (XRC), and nuclear magnetic resonance (NMR) spectroscopy have a limited capacity to obtain such information on conformational rearrangements.[5,6] While, XRC B-factors can, in principle, provide information about the distribution of conformers present in the crystal,[7] the nonphysiological conditions (i.e., the protein is in the solid state and often at very low temperatures) and crystal packing artifacts can perturb protein structure, particularly at surface-exposed side chains and loops.[8,9] In

* Corresponding author phone: +44 161 30 64 199; fax: +44 161 30 68918; e-mail: andrew.almond@manchester.ac.uk.
† Manchester Interdisciplinary Biocentre, Faculty of Life Sciences, University of Manchester.
‡ Michael Smith Building, Faculty of Life Sciences, University of Manchester.
§ University of Oxford.
‖ University of Maryland.

the case of NMR, while performed in the solution state at physiological temperatures, it typically relies on nuclear Overhauser enhancements (NOEs) to provide local geometric information between neighboring protons. These NOEs are resultant from dynamic molecular processes but are generally interpreted as a set of static structures through simulated annealing, which can lead to unrealistic conformational bias.[10]

A more detailed analysis of protein dynamics in solution can be performed using experimental techniques such as fluorescence anisotropy,[11] measurements of NMR relaxation parameters, and residual dipolar couplings.[12−14] While NMR relaxation is in principle capable of investigating ns-time scale protein dynamics at residue-specific positions, a serious problem is encountered because both internal motion and overall tumbling contribute to dipolar relaxation. Since overall tumbling in proteins occurs on the nanosecond time scale, separation of the two processes cannot be achieved effectively even in the ideal case of isotropic motion. Therefore, ^{15}N-relaxation (and other NMR-relaxation based experiments) is more effective at studying fast picosecond-time scale dynamics rather than those on the nano- to microsecond timescales, on which local protein conformational rearrangements occur.[15]

Due to these experimental limitations, protein nanosecond-time scale dynamics are relatively unexplored in proteins, which is of significant concern considering its crucial role in ligand-binding events. In the absence of detailed experimental data about molecular conformational rearrangements, computer simulations can be employed. Molecular dynamics (MD) simulation using a molecular-mechanics force field is one of the best all-atom theoretical techniques presently available for investigating ns-time scale dynamic atomic motions in proteins and their attendant solvent molecules and ions. With increases in computing power, simulations of proteins are now being performed more routinely on the 0.1 μs time scale.[15] Such simulations are sufficient to investigate nanosecond-time scale dynamics effectively. Furthermore, MD is able to predict with relatively high accuracy the rotamer conformation of the side chains important for molecular recognition, and even in simulations of unbound protein, these frequently visit the rotamer conformations seen in the complex.[4,16] However, MD simulations are based on assumptions about molecular potential energies and at present need to be compared carefully against experimental data to ensure that realistic results are being obtained. In this regard, two recent studies that compared MD simulations of proteins against a variety of experimental data are illustrative.[15,17] The first concerns two simulations of ubiquitin, which were compared against NMR data,[15] and the second is a 3 ns simulation of bacterial cytochrome *c*, which was compared against amide exchange rates and XRC B-factors.[17] These studies show that MD simulations are accurate enough to predict a wide array of experimental data and therefore provide crucial insight into nanosecond dynamics.

Such promising results now give the impetus to understand dynamics in a wider range of experimentally characterized proteins and particularly in proteins that bind to flexible ligands. The Link module, for example, is a structural domain of approximately 100 amino acids that plays a fundamental role in extracellular matrix assembly via its association with the high molecular weight polysaccharide hyaluronan.[18−21] This domain is found in extracellular matrix proteins such as cartilage link protein, aggrecan, cell surface receptors (including CD44), and the inflammation-associated protein tumor necrosis factor-stimulated gene-6 (TSG-6). TSG-6 is a secreted protein comprising a Link and a CUB module that is not constitutively expressed in normal adult tissues. It has roles in inflammation, matrix remodeling, and ovulation via interactions with hyaluronan[21] and also interacts with a wide range of other extracellular matrix molecules including sulfated glycosaminoglycans and various proteins.[22,23] The Link module from human TSG-6, which has been expressed in bacteria (as an 11-kDa, 98-residue construct termed Link_TSG6) and shown to support hyaluronan binding,[24] comprises two α-helices and two antiparallel triple-stranded β-sheets arranged around a large hydrophobic core,[24,25] as illustrated in Figure 1(a). Link_TSG6 interacts with hyaluronan via a binding groove that is opened by rearrangements within the $\beta1$-$\alpha1$ and $\beta4$-$\beta5$ loops that allow the key residues (K11, Y12, Y59, F70, Y78, and R81) to bind to hyaluronan.[25,26] Bikunin, a serine protease inhibitor that is potentiated through its association with TSG-6, also binds to a similar region (i.e., K11, Y12, Y59, and Y78) via a protein−protein interaction.[27] In contrast, heparin interacts with surface-exposed basic residues (including K20, K34, K41, and K54) on a different face of Link_TSG6.[27] Since hyaluronan and heparin are locally dynamic molecules,[28−31] it is reasonable to suggest that the entropic penalty associated with their binding to the Link module may be lessened by binding them using a dynamic platform of loops and surface residues in the interaction sites. Nanosecond time scale dynamics is therefore likely to be important to the biological roles of Link modules in the extracellular matrix.

NMR structures of Link_TSG6 have been determined in both the free and hyaluronan-bound state.[24,26,32] Since virtually every atom was assigned in free Link_TSG6, the solution structure was determined to high resolution (over 13 NOE constraints per residue) and is therefore a good starting conformation for MD simulation. These assignments also enabled a considerable amount of other NMR data to be collected, making a large experimental data set that can be used to test simulations. Here, we present a 0.25 μs MD simulation of Link_TSG6 that was started from the lowest energy NMR-derived structure[26] with additional side-chain refinements including salt-bridges. The results are compared against NOE intensities, ^{15}N-NMR-relaxation measurements, vicinal scalar couplings, amide hydrogen exchange times, and temperature coefficients. By suitable analysis, the simulation is used to investigate nanosecond time scale dynamics and provide an enhanced view of loops and surface side chains that could not be obtained by NMR alone. The resultant dynamic picture provides a more realistic molecular view of the TSG-6 Link module and serves as a basis for understanding how this might affect its interactions with glycosaminoglycans, compared to the static and conforma-
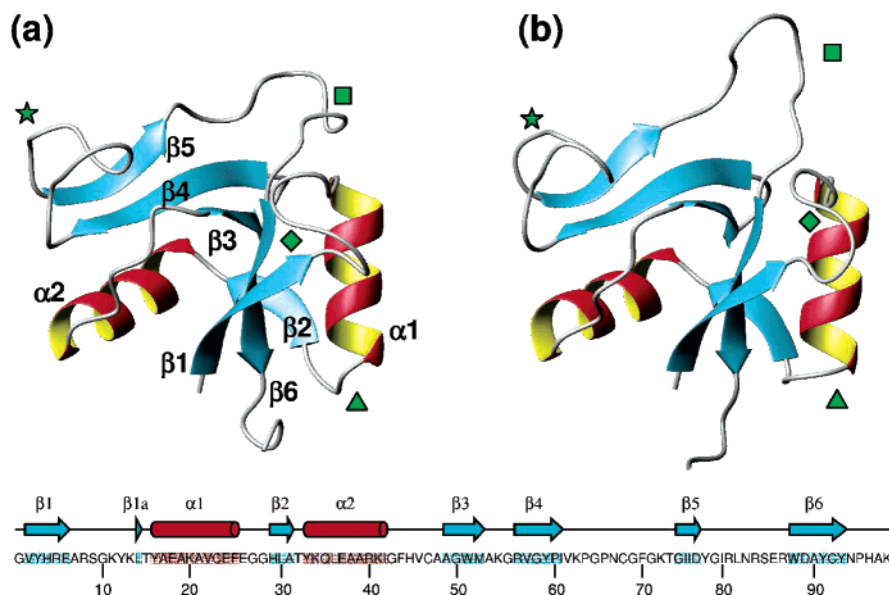
Figure 1. (a) The average solution backbone structure for Link_TSG6 of the lowest-energy 125 (of 250) minimized conformations derived from simulated annealing[26] and (b) the average backbone structure derived by averaging a 0.25 μs aqueous simulation of Link_TSG6 (125,000 frames). The following loops are marked: β1-α1 (green diamond), α1-β2 (green triangle), β4-β5 (green star), and β5-β6 (green square). In the lower portion of the figure the relationship between primary sequence and secondary structure is shown.

tionally biased views produced by XRC and NMR structure determination.

2. Materials and Methods

2.1. Molecular Dynamics Simulation. In all calculations, the CHARMM program[33] was used together with the CHARMM22 parameters for proteins.[34,35] Nonbonded interactions were truncated using the switching function between 0.8 nm and 1.0 nm, and nonbonded list generation was stopped at 1.2 nm, with the dielectric constant set to unity. Periodic boundary conditions were applied to the whole system by application of a face-centered cubic lattice transformation (the unit cell corresponds to a truncated octahedron) and using the Ewald summation convention with order set to 6, $\kappa = 0.44$ and 64 points over the box. Potential energy minimizations were performed using an implementation of the adopted-basis Newton−Raphson algorithm. The molecular dynamics simulation produced an NPT ensemble (variable volume) by application of the Nosé−Hoover algorithm with suitable parameters to maintain the pressure at 1 atm and the temperature at 298 K. The SHAKE methodology was used to keep bonds to hydrogen atoms rigid and hence permit an integration time-step of 2 fs. Numerical integration of the equations of motion applied the Leapfrog variant of the original Verlet algorithm.

Simulations were started from a low-energy conformer derived from NMR data as described previously.[26] However, slight structural refinements were made (including hydrogen bonds to side chains and intramolecular salt-bridges inferred from experiments conducted over a wide-range of pH values), and the majority charge state of each histidine residue was inferred at pH 6.0 by chemical-shift titrations (C. D. Blundell, I. D. Campbell, and A. J. Day, manuscript in preparation); H29 and H96 were made positively charged, while H4 and H45 were left uncharged. Basic (R, K) and

acidic (D, E) amino acids and the N- and C-termini were given the relevant charge at pH 6.0; all other amino acids were neutral giving an overall charge of $+10e$. The starting structure was centered inside a pre-equilibrated truncated octahedral water box, which initially contained 6000 TIP3P water molecules,[36] and water molecules that overlapped with protein were deleted. Ions (19 Na$^+$ and 29 Cl$^-$) were added in a random fashion clear of the protein surface, to explicitly neutralize the charge of the protein and simulate a 0.2 M background solution of NaCl (i.e., comparable to physiological ionic strength). The protein was then fixed, and the surrounding water and ions were minimized (100 steps) and subjected to 500 ps of equilibrating dynamics at 300 K (strong coupling to heat bath and pressure). Following this, the whole system was minimized (100 steps) and equilibrated with free dynamics for a further 500 ps. The simulation was continued for a further 0.25 μs at constant pressure and temperature (by weak coupling to heat bath and pressure). Coordinates were saved at 2 ps intervals for later analysis.

2.2. Analysis of Global Protein Conformation. The 0.25 μs simulation resulted in 125 000 separate frames, containing protein, water, and ions. For the majority of analyses, these data sets were not used directly but aligned prior to analysis. In order to achieve this, first water and ions were removed, and then each frame from the simulation was translated and rotated such that the root-mean-square-deviation (RMSD) of backbone atoms (N, C′, C$^\alpha$) within secondary structural elements from the NMR-derived starting structure was minimized; this set of structures will be referred to as being "in the molecular frame". A similar molecular frame calculation was performed for the 125 lowest-energy minimized structures derived from NMR data.[26] Although this is more than normally reported, none of these structures had serious steric clashes, which was not true for many of the remaining 125 higher-energy structures calculated. Average structures

for each ensemble were generated by calculating the average atomic positions in the molecular frame. Surface racer 3.0[37] was used to calculate the surface area of each coordinate set in the simulation, with a probe radius of 0.1 nm to reveal solvent accessible areas of the protein. Calculations were performed on the whole protein and repeated with the $\beta4$-$\beta5$ loop (residues V62-T73) omitted. The radius of gyration was calculated using standard equations (without mass weighting).

2.3. Prediction of Hydrogen Bonds and Comparison with Experimental Data. Hydrogen bonds were defined by distance and angular criteria, namely that the A-H...B angle deviated no more than 60° from direct and the distance between A and B did not exceed 0.35 nm (A and B are the electronegative donor and acceptor, respectively). Backbone amide hydrogen and oxygen atoms were examined for possible hydrogen bonds in each frame of the simulation, and average occupancies were calculated from this. Hydrogen bonds in side chains and salt-bridges were calculated in an analogous way.

Amide hydrogen exchange rates were calculated from the simulation using an empirical relationship,[38] which has recently been used to provide structural restraints.[39] In the EX2-limit of exchange,[40] it is assumed that both the presence of intramolecular hydrogen bonds (dominant factor) and the burial of amide hydrogen atoms in the interior of the protein (subsidiary factor) leads to protection, which slows the rate of exchange. Therefore, if N_h is the average number of hydrogen bonds that each amide hydrogen partakes in and N_c is the number of contacts made by each residue, then the protection factor, P, can be calculated by $\ln P = \beta_c N_c + \beta_h N_h$. The original values ($\beta_c = 1$ and $\beta_h = 5$) have been used unmodified here,[38] but it should be noted that other authors have tried to optimize them on the basis of molecular dynamics simulations.[17] The value of N_c was calculated by finding all residues that had a closest-approach distance less than 0.25 nm (any atom), and the value of N_h was calculated from the average number of hydrogen bonds made to and from amides during the simulation (using the definition described above).

Measurement of the amide hydrogen residence times using a H_2O/D_2O exchange NMR experiment has been described previously.[26] Temperature coefficients were obtained from ^1H-^{15}N HSQC spectra collected between 5 °C and 30 °C at 5 °C intervals at 750 MHz ^1H-frequency on a 0.3 mM sample of ^{15}N-labeled Link_TSG6 containing 0.33 mM DSS, as described previously.[41] Chemical shift deviations (CSDs) were calculated and used to improve the prediction of hydrogen bonds using a cutoff line of $-2.97 - 2.19 \times$ CSD.[42] Structure calculations were carried out (as described previously) to test hydrogen bonds for compatibility with the NMR structure.[26] These data were used in combination to determine hydrogen bond permanency.[43]

2.4. Prediction of NOESY Cross-Peak Intensities and Comparison with Experimental Data. Several NMR experiments, described and performed previously,[26] were used to obtain NOE distance constraints: a 2D ^1H-NOESY (unlabeled protein) in D_2O, a 3D ^{13}C-NOESY-HSQC (^{13}C,^{15}N-labeled protein) in H_2O, a 3D ^{15}N-NOESY-HSQC, and a

3D ^{15}N-HSQC-NOESY-HSQC (^{15}N-labeled protein) both in H_2O. In order to make a comparison between the simulation and this experimental data, the raw intensities of assigned peaks were measured in each spectrum.[26] In total, there were 188, 799, 859, and 32 assigned peaks from the spectra listed above, respectively, i.e., 1878 nonduplicated intensities. The theoretical prediction of the intensity used the independent spin-pair approximation in the slow tumbling regime.[44] In this case, the theoretical intensity is $k<r^{-3}>$,[2] where k is an experimentally derived constant (dependent on each particular experimental setup), r is the distance between protons, and the angular brackets represent an average over the simulation. The value of k was calculated for each of the NOESY-based experiments, and $\log I$ (I is the experimentally measured intensity) was plotted against $\log(k<r^{-3}>^2)$ for each set of assigned peaks. In the case of degenerate methyl or methylene groups, $<r^{-3}>^2$ was calculated between all combinations of proton pairs and summed.

2.5. Prediction of ^{15}N-Relaxation and Comparison with Experimental Data. Angular autocorrelation functions[45] of N−H-vectors were calculated at each amide directly from the molecular dynamics simulation. Although their decays were exponential, the overall correlation time estimated from the simulation was ∼2.4 ns, i.e., lower than that obtained from experimental observations (∼4.5 ns), see below. It was therefore concluded that the simulations were not capable of providing a realistic description of the overall molecular tumbling. Therefore, the angular autocorrelation functions were calculated in the molecular frame (see above), which provides an approximation to their internal correlation functions, $C^i(t)$, as described in the Lipari-Szabo model-free method for describing NMR relaxation.[45] These functions have a value of 1.0 at time $t = 0$ and are theorized to decay to an asymptotic value $S^2 \le 1.0$, where S^2 is the order parameter for internal motions. The correlation functions were calculated out to 1 ns (500 points at 2 ps intervals), and the value of S^2 was taken at this endpoint. In order to calculate the ^{15}N-relaxation parameters at each residue, their internal motion autocorrelation functions were extended to 2^{22} points by setting $C^i(t) = S^2$ for $t > 1$ ns. Each correlation function was multiplied by $0.2\exp(-t/\tau_m)$ at every point, where τ_m is the overall correlation time. The resultant data sets were then converted to spectral density functions, $J(\omega)$, by performing a real fast Fourier transform. These were used to calculate $J(\omega)$ for specific values of angular frequency, ω, which were then input into the standard equations used to calculate ^{15}N-relaxation parameters, i.e., T_1-relaxation rates and steady-state NOE enhancement, η.[12,46] In these calculations, the N−H-bond length was assumed to be 0.102 nm, the chemical shift anisotropy was set to -160 ppm, and τ_m was set to 4.5 ns in accordance with the experimental data. In this manner, the values were predicted from the simulation at each nitrogen in the protein backbone. The calculations were repeated for the N−C$^\alpha$-bond with the same parameters, except for the overall correlation time, which needed to be increased to 5.5 ns to agree with the experimental data (this change may seem unusual at first sight, but it should be remembered that τ_m is part of the Lipari-Szabo model and is not a well-defined physical parameter in a dynamic

Dynamics of the Hyaluronan-Binding Link Module

J. Chem. Theory Comput., Vol. 3, No. 1, 2007 **5**

molecular system because internal motion and overall tumbling cannot be separated). The ^{15}N-T_1, ^{15}N-T_2, and ^1H-^{15}N steady-state NOE (η) data were acquired in a similar manner to that described previously[29] at a ^1H frequency of 500 MHz on 0.3 mM and 2 mM samples, respectively, of ^{15}N-labeled Link_TSG6 at pH 6.0, 25 °C, in 10% (v/v) D$_2$O. Lipari-Szabo generalized order parameters were calculated from the relaxation data using standard relaxation equations that assume a single overall correlation time and an exchange contribution.[12,45]

2.6. Prediction of Scalar Couplings and Comparison with Experimental Data. The backbone ϕ-dihedral angles were extracted from the simulation at all residues, except for proline and glycine. From these values the vicinal couplings were calculated at each point using the 'zero motion' Karplus equation: $^3J_{HNH\alpha}(\phi) = 9.5\cos^2\phi - 1.4\cos\phi + 0.3$, from which the predicted value was computed by averaging across all time frames.[47] A ^{15}N-HMQC-J spectrum was recorded on a sample of ^{15}N-Link_TSG6 (2 mM, pH 6.0, 10% v/v D$_2$O) at 500 MHz for the direct measurements of $^3J_{HNH\alpha}$ coupling constants.

3. Results

3.1. The Global Secondary Structure. The 125 lowest-energy minimized structures of Link_TSG6, derived from previous NMR measurements,[26] were averaged in the molecular frame (see Methods). Figure 1(a) shows the resulting conformation using traditional secondary structural visualization of the backbone, which is 0.1 nm RMSD from the lowest energy NMR structure. A similar average conformation, calculated from the 0.25 μs molecular dynamics simulation, is shown beside it in Figure 1(b). The Link module of TSG-6 contains two α-helices (α1, residues Y16-F25 and α2, residues Y33-I42) and two antiparallel β-sheets.[24,26] Sheet I comprises strands β1 (V2-E6), β1a (L14), β2 (H29-A31), and β6 (D89-Y93), and sheet II comprises strands β3 (A49-M52), β4 (R56-I61), and β5 (G74-D77). Comparison of the average structures, Figure 1(a,b), immediately shows that the simulation has maintained all the secondary structural elements of the molecule. Furthermore, the average backbone conformations of most of the loops are similar in the two average structures; the major exception is the β5-β6 loop (green star in Figure 1), which deviates significantly from the average NMR backbone conformation. Also, a slight deviation was noticed in the β1-α1 loop (green square in Figure 1), which is in steric contact with the β5-β6 loop in the NMR structure. Therefore, it should be expected that deviations may occur in experimental predictions from the simulation at these positions.

These findings can be explored in more detail by examining the range of conformations present during the simulation and in the NMR-derived ensemble. Figure 2(a) shows an overlay in the molecular frame of the 125 lowest-energy minimized structures derived from NMR data (of 250), in which the C-terminal tail, the β1-α1 loop, the β4-β5 loop, and the β5-β6 loop are not as well-defined as the other parts of the molecule. This can be compared with the output from the simulation, Figure 2(b), in which 125 structures (taken at 2 ns intervals) have been overlaid. It is apparent that the

simulation ensemble is a more disordered set of structures than the NMR ensemble, Figure 2(a), even in the secondary structure elements. Of particular note are the β1-α1, α2-β3, and β5-β6 loops, which are substantially more disordered in the simulation ensemble. However, the β4-β5 loop (that forms one side of the hyaluronan-binding groove) has a similar range of conformations in both the simulation and the NMR-derived ensemble.

The overall change in conformation of the protein during the simulation can be visualized by plotting the minimum RMSD between each simulated backbone conformation and the starting structure, Figure 3(a). It can be seen that the simulation started at around 0.2 nm RMSD from the starting structure (after equilibration) and relaxes over the first 50 ns of the simulation. Comparison to a similar RMSD calculation with the β5-β6 loop omitted, Figure 3(a), shows that the structural relaxation is predominantly due to movement and partial unfolding of this loop. Following this, the simulation reaches an equilibrium, being on average 0.3 nm RMSD from the NMR structure.

3.2. Intramolecular Hydrogen Bonds. The decrease in the exchange rate incurred by folding from the free polypeptide chain and intramolecular hydrogen bonding can be described by an amide-hydrogen protection factor, P. Therefore, the NMR and simulation ensembles were used to estimate the value of $\ln(P)$ at each amino acid, as plotted on Figure 3(b). These values can be compared against the experimentally observed amide-hydrogen persistence for a protonated protein placed into D$_2$O.[26] However, it should be pointed out that the protection factor does not take into account differences in intrinsic exchange rate between amino acids, and the comparison will remain qualitative. At this level of analysis, a key difference between the predictions made by the NMR and simulation ensembles is in the region of Y16. The lower protections factors predicted by the simulation in this region (as compared with the alpha-helical regions) are in better qualitative agreement with the experimental data than the NMR ensemble, which predicts relatively large protection factors. Furthermore, within the majority of the secondary structural elements (i.e., 30 amides) the agreement between experiment and theoretical prediction from the simulation is good. Notable exceptions are Y3H^N, T32H^N, A53H^N, Y59H^N, I61H^N, I76H^N, Y78H^N, and Y91H^N, which are denoted with arrows on Figure 3(b). Several of those discrepancies can be accounted for by the NMR experimental parameters, in which data sets could only be collected every 105 min, and with limited resolution in the indirect dimension. For example, although Y3, T32, I76, Y59, and Y78 HN-resonances could not been seen in the first spectrum of the free protein (>105 min after reconstitution with D$_2$O), they were seen in the same experiment on the hyaluronan-bound protein, where global exchange is ~10 times slower.[26] However, the involvement of these amides in hydrogen bonds is also consistent with the observed NOE pattern in the free protein. Furthermore, temperature coefficient data are consistent with all of these discrepancies (see the Supporting Information). In particular, analysis suggests that several hydrogen bonds are present in the region A53-E86 that were not determined by amide exchange experi-
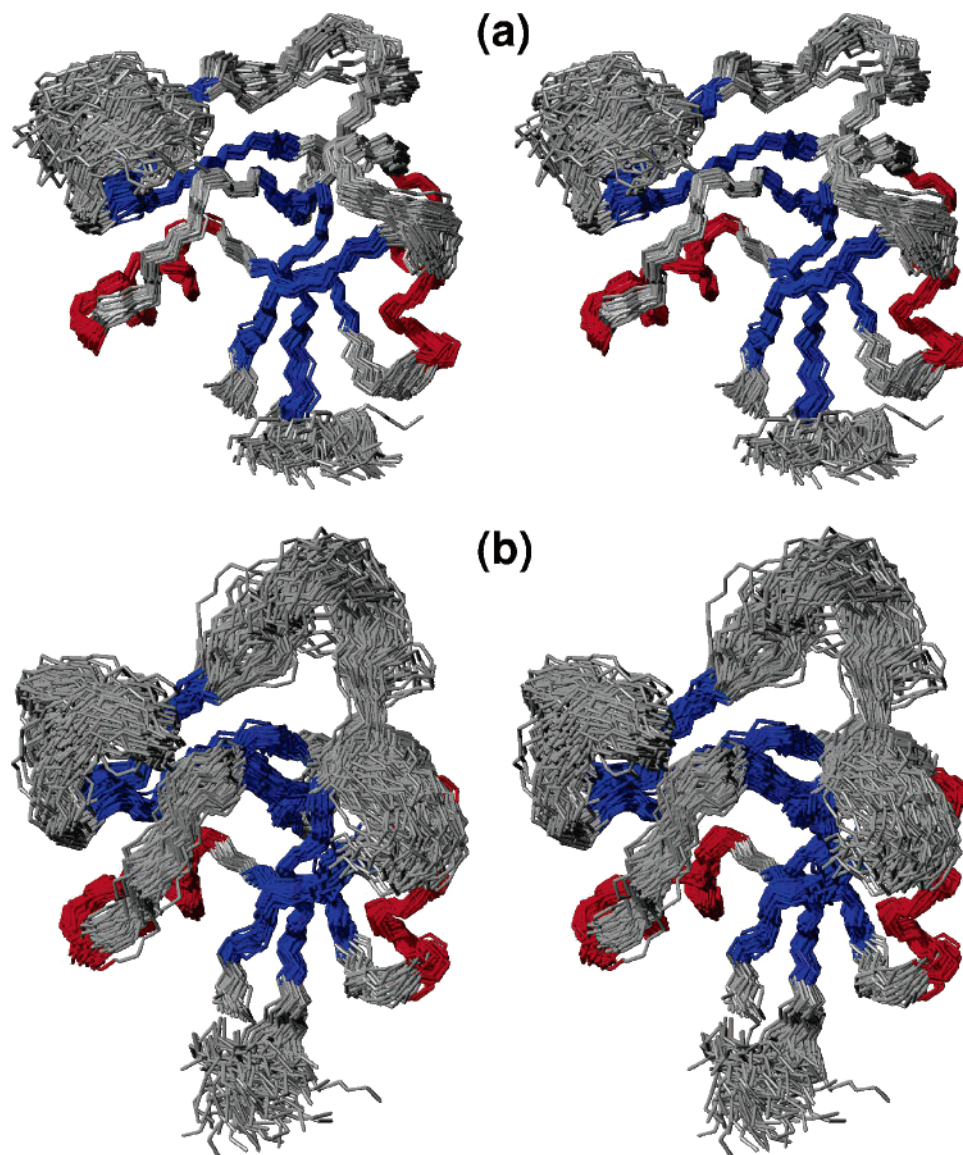
Figure 2. (a) Stereo overlays of 125 (of 250) minimized structures derived from simulated annealing and (b) from equally spaced (2 ns) conformations extracted from a 0.25 μs simulation of Link_TSG6. Both structures are colored according to secondary structure (red α-helices and blue β-strands).

ments, which includes the $\beta4$, $\beta5$ strands and the $\beta4$-$\beta5$ loop. The other discrepancies (I61HN, A53HN, and Y91HN) can be accounted for by assignment problems or overlap with other resonances. A single slowly exchanging hydroxyl hydrogen, on T32, was also observed in NMR spectra of the free protein,[26] and the NMR structures do not provide any evidence for a hydrogen bond in spite of the hydroxyl group being very well defined. Such slowly exchanging hydrogen atoms have been seen in less than 5% of reported threonine assignments (according to BioMagResBank). The simulations suggested that this hydroxyl group was in close contact with solvent and not involved in any substantial intramolecular hydrogen bonding interactions, so also does not provide an explanation.

Hydrogen bonds were characterized according to their temporal occupancy during the 0.25 μs simulation using criteria based on angle and distance (see Methods). Those hydrogen bonds in the backbone with a fractional occupancy greater than 50% (considered to be strong hydrogen bonds)

are detailed in Table 1. As expected, high occupancy hydrogen bonds were restricted to the secondary structure elements, Figure 4. The hydrogen bonds found in the simulations are marked with dotted lines and labeled with their percentage occupancy in Figure 4.

Within both α-helices (Figure 4) the hydrogen bond pattern predicted by the simulation was as expected, with the characteristic $i + 4 \rightarrow i$ H^N...O interactions (from hereafter hydrogen bonds will be represented by "...") and fractional occupancies in excess of 70%. In both helices, it was found that hydrogen-bonding interactions from $i+3 \rightarrow i$ H^N...O (as found in 3_{10}-helices) could coexist with those normally associated with α-helices but with a lower occupancy ($\sim$25%). At the N-terminal end of both α-helices the backbone hydrogen bonds were consistent with those inferred from experimental data,[26] see Figure 4. The two hydrogen bonds forming the α1-helix cap (not shown in Figure 4) were observed in the simulation: T15H^N...Q18O$^{\epsilon 1}$ (26%) and Q18H^N...T15O$^{\gamma 1}$ (25%). However, they were found to have

Dynamics of the Hyaluronan-Binding Link Module

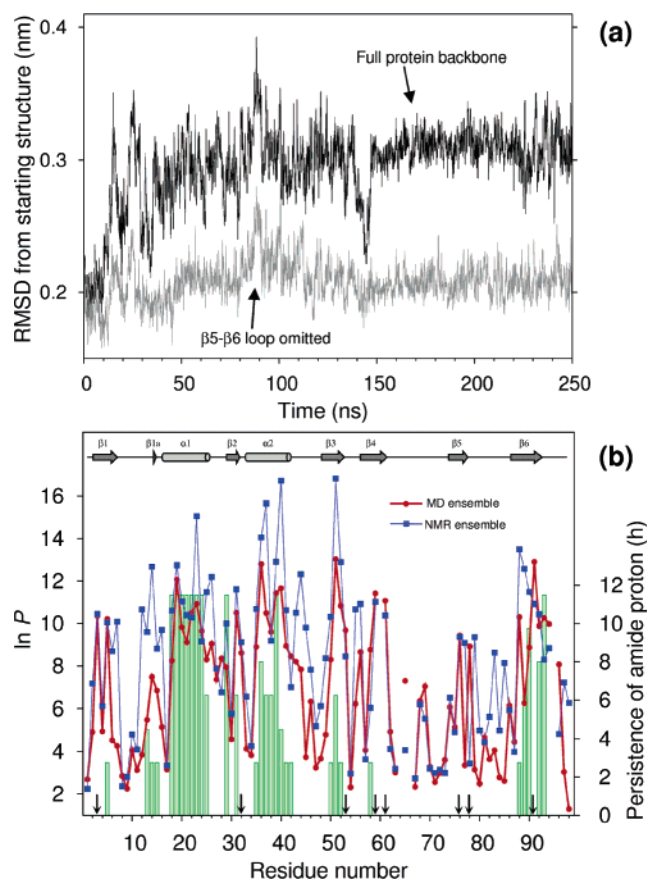*J. Chem. Theory Comput., Vol. 3, No. 1, 2007* **7**

Figure 3. (a) RMSD of the backbone between the simulation and the NMR-derived starting structure (as a function of time) with and without consideration of the $\beta5$-$\beta6$ loop in the sums. (b) Calculated values for the amide protection factors in Link_TSG6 derived from a 0.25 μs simulation (red line and circles) and from the NMR ensemble (blue line and squares). The green bars indicate the latest time at which each amide could be observed in a H_2O/D_2O exchange experiment. Arrows denote obvious inconsistencies between the theoretically predicted protection factors and experimental measurement.

a relatively low persistence in the simulation and another hydrogen bond was identified, T15H$^{\gamma1}$...Q18O$^{\epsilon1}$ (26%), which has not been confirmed experimentally. At the N-terminus of the $\alpha2$-helix T32H^N...Q35O$^\epsilon$ (81%) and Q35H^N...T32O$^\gamma$ (62%) were found to cap it, with a rather large persistence. At their C-termini the α-helices were predicted to have more hydrogen bonds than determined previously by NMR.[26] First, G27 and G43 were inferred to be involved in hydrogen bonds (to C23 and A39, respectively) in the simulation, and while a slowly exchanging amide has not been identified experimentally here, amide temperature coefficients support their existence (see supplementary Table 1). Also, at the C-terminus of both helices a capping hydrogen bond was found from $i + 5 \rightarrow i$ (as found in π-helices), again not seen experimentally. This was of high frequency in $\alpha1$ (G28H^N...C23O) at 92% and lower in $\alpha2$ (F44H^N...A39O) at 50% (see Figure 4). Again, temperature coefficient data indicate that they are likely to be present in the free protein.

The simulations reproduced the majority of the hydrogen bonds in β-sheets, as shown in Figure 4, in the sense that

hydrogen bonds had a high-percentage occupancy where slowly exchanging amides were observed experimentally.[26] The two obvious exceptions to this agreement were D89H^N...A49O and L14H^N...W88O, which were identified as slowly exchanging amides, but have persistencies of only 11% and 28%, respectively, in the simulations. Furthermore, the simulation indicated the presence of more persistent hydrogen bonds across the strands than had been reported previously.[26] In particular, the $\beta1$-$\beta6$, $\beta3$-$\beta6$, $\beta3$-$\beta4$, $\beta4$-$\beta3$, and $\beta4$-$\beta5$ strand pairs were all suggested to have one or two persistent hydrogen bonds at positions that had not been deduced previously by amide exchange data (Figure 4 and Table 1), e.g., Y78H^N...V57O and G65H^N...G69O. Such extensions are consistent with recent temperature coefficient measurements and with structure building exercises (see the Supporting Information, Table 1). Also, of particular note is a tertiary structural hydrogen bond between A53H^N and A31O (99%), which has not been proposed previously (not shown in Figure 4). Temperature coefficient measurements were supportive of this observation.

The strongest salt-bridges (present for greater than 80% of the time) were found between R56...D77 (end of $\beta4$ to $\beta5$), R81...E86 (within the $\beta5$-$\beta6$ loop), and R5...E26 (end of $\beta1$ to end of $\alpha1$). These were present in the starting structure and have been maintained during simulation. Weaker salt-bridges (15−50%) were predicted between K13...E18, K20...E24, K34...E37, R8...E26, R87...E18, and H29...E24. The residue K11 formed a bifurcated salt-bridge to both E6 and D89.

3.3. Comparison with Experimental NOESY Data. As noted above, the simulations compared favorably with previous characterizations of secondary structure outside of the $\beta5$-$\beta6$ loop. Comparison with experimental NMR intensity data was therefore used to further test the simulations. Figure 5(a) shows the log$I$ values plotted against the predictions, log(kr^{-6}), for the NMR-derived starting structure. Figure 5(b) shows (using the same axes ranges) the same data averaged over the simulation; in this case the predictions are log($k<r^{-3}>^2$). It can be seen that, overall, the correlation between the theoretical predictions and experimental measurements is good in both cases. However, the simulation produced some outliers, Figure 5(c), which were not present in the original NMR-derived structure. These data are shown again, relative to the protein sequence, Figure 5(d), where the absolute difference between the experimental and theoretical log values have been plotted. The regions where disagreement is present are shown using bars and localize to residues 10−18, 50−58, and 80−86. These NOEs are between residues in the $\beta1$-$\alpha1$, $\beta3$-$\beta4$, and $\beta5$-$\beta6$ loops. It is inferred that this is principally due to large motions of the $\beta5$-$\beta6$ loop in the simulation (which has experimental NOE cross-peaks to protons in the other two loops). The RMS deviation between log$I$ and log(kr^{-6}) is 0.584 for the lowest-energy NMR-derived structure. For the simulation if all data are taken into consideration, the deviation is 0.831, and if the 42 outliers (out of 1878) are taken out, then the deviation becomes 0.588, which is not significantly different from the starting structure. In other words, the dynamic representation of the protein given by the simulation is as

Table 1. Hydrogen Bonds between Backbone Amide Hydrogen (H^N) and Carbonyl (O) Present for More than 50% of Frames in the Simulation

H^N donor	O acceptor	% occupancy	secondary structural element	H^N donor	O acceptor	% occupancy	secondary structural element
G1	N94	82.9	β1-β6 strand	I42[a]	A38	72.2	α2 helix
Y3[a]	C92	99.7	β1-β6 strand	G43	A39	98.2	α2 helix
R5[a]	A90	98.9	β1-β6 strand	F44	A39	51.0	α2 helix
A19[a]	T15	99.8	α1 helix	G50[a]	G58	74.8	β3-β4 strand
K20[a]	Y16	94.5	α1 helix	W51[a]	D89	99.9	β3-β6 strand
A21[a]	A17	78.3	α1 helix	M52[a]	R56	98.9	β3-β4 strand
V22[a]	E18	96.1	α1 helix	A53	A31	99.2	tertiary
C23[a]	A19	82.5	α1 helix	R56	M52	83.0	β3-β4 strand
E24[a]	K20	92.0	α1 helix	G58[a]	G50	83.6	β3-β4 strand
F25[a]	A21	90.1	α1 helix	Y59[a]	I76	97.5	β4-β5 strand
E26	V22	82.7	α1 helix	I61	G74	88.0	β4-β5 strand
G27	C23	77.1	α1 helix	G65	G69	80.0	β4-β5 loop
G28	C23	92.3	α1 helix	G69	G65	62.3	β4-β5 loop
H29[a]	Y93	75.8	β2-b6 strand	I76[a]	Y59	98.6	β4-β5 strand
A31[a]	Y91	99.2	β2-b6 strand	Y78	V57	83.2	β4-β5 strand
L36[a]	T32	92.9	α2 helix	W88[a]	L14	80.2	β1a-β6 strand
E37[a]	Y33	98.7	α2 helix	A90[a]	R5	80.4	β1-β6 strand
A38[a]	K34	93.3	α2 helix	Y91	Y51	83.9	β3-β6 strand
A39[a]	Q35	96.0	α2 helix	C92[a]	Y3	94.6	β1-β6 strand
R40[a]	L36	97.9	α2 helix	Y93[a]	H29	99.9	β2-β6 strand
K41[a]	E37	76.9	α2 helix	N94	G1	98.8	β1-β6 strand

[a] Experimentally determined slowly exchanging amide hydrogen.

good a fit to the experimental data as the static representation from simulated annealing (except in the β5-β6 loop). Furthermore, recent work has shown that a physically correct ensemble also correctly predicts the nonobserved NOEs.[48] Such analysis is possible using the aromatic region (F, H, Y, and W residues) of the 2D-NOESY spectrum recorded in D_2O, which is the least ambiguous part of any of the NOESY spectra collected. In this region, 36 NOEs were predicted by both the lowest-energy NMR structure and the simulation ensemble (<5 Å) that were not assigned experimentally (it should be stressed that some of these 'missing assignments' may be due to artefacts or overlap). However, a further 68 NOEs were predicted by the NMR structure alone and 56 by the simulation alone that were not assigned experimentally. Therefore, in fact, the simulation predicts fewer nonassigned/nonobserved NOEs than the static lowest-energy NMR structure.

For direct comparison with the NMR spectra, the average distance ($<r^3>^{-1/3}$) was calculated from the simulation for all assigned NOEs in the 2D ^1H-NOESY in D_2O. It was assumed that if an NOE could be observed between protons in an NMR spectra, their average distance must be less than 0.5 nm (as is typically done). Using chemical shift tables, cross-peaks were labeled directly onto the raw NMR spectrum. A section of the 2D NOESY in the aromatic-fingerprint region of the spectra is shown in Figure 6. Predictions that are within 0.5 nm are labeled in blue, while those in which the prediction was greater than 0.5 nm are labeled in red; this part of the spectrum was chosen to show both the agreement with the experimental data and also some disagreement. Out of 216 peaks assigned in the 2D NOESY,

the simulation predicted 185 to be within 0.5 nm (86%). The static starting structure placed 195 within the same distance (90%).

3.4. Comparison with ^{15}N-Relaxation Data. While both a static structure and a dynamic simulation can predict the average NOEs reasonably well, only the simulation is capable of predicting parameters that are highly dependent on dynamics. One such set of observables can be obtained from ^{15}N-relaxation NMR experiments. Theoretically, NMR dipole—dipole relaxation (as discussed here) is related to angular correlation functions calculated from vectors joining the relaxing nuclei.[45,49] First, autocorrelation of the N—H-vectors on long timescales directly from the aqueous simulation resulted in an overall rotational correlation time of 2.4 ns that was too rapid for a molecule of this size.[50] The Link_TSG6 molecule diffused with axial symmetry, but the diffusion tensor was found to have a slight anisotropy ($D_{||/⊥}$) of 1.13.

Therefore, rather than computing N—H-bond-vector autocorrection functions for the freely rotating protein, they were calculated in the molecular frame (removing overall tumbling). In this frame the correlation functions did not decay to zero but to a constant value that describes the Lipari-Szabo order parameter (S^2), see Figure 7(a). Comparison with the experimentally derived values gave a correlation coefficient of 0.38 and an RMSD of 0.14 (see the Supporting Information). The internal correlation functions were multiplied by an exponential representing overall tumbling, and the resultant data set was Fourier-transformed to produce a spectral density function. From this, the ^1H-^{15}N heteronuclear steady-state NOE enhancement (η) and T_1-values were calculated at residue specific positions. These were compared

Dynamics of the Hyaluronan-Binding Link Module

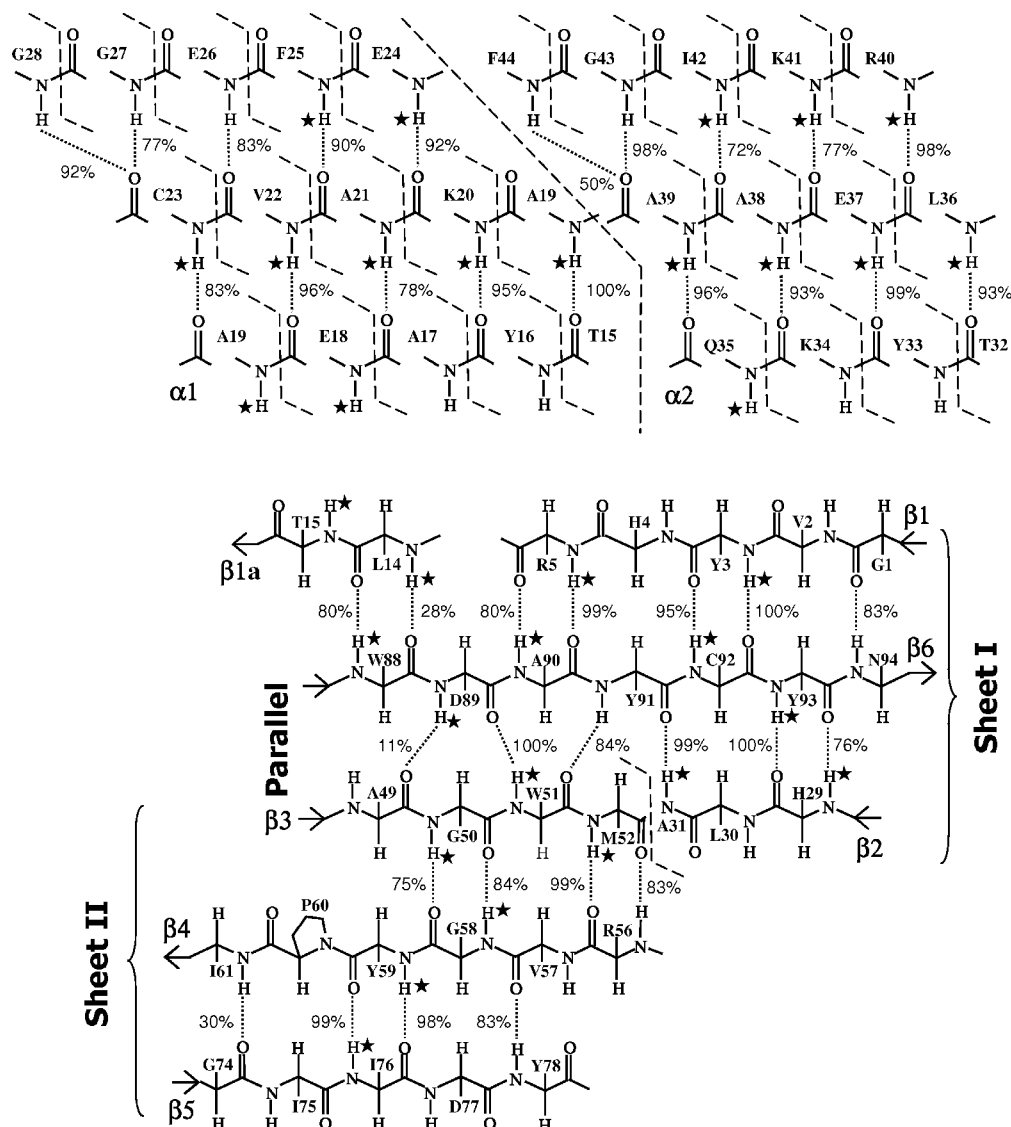*J. Chem. Theory Comput., Vol. 3, No. 1, 2007* **9**

Figure 4. Hydrogen bonds within the secondary structure elements of Link_TSG6. Experimentally determined slowly exchanging amides[26] are marked with stars. Hydrogen bonds found in the simulation (with a fractional occupancy greater than 50%) are marked by a dotted line and labeled with their percentage occupancy.

against actual NMR measurements made on ^{15}N-labeled protein. The best fit to the overall correlation time was 4.5 ns, leading to the comparison shown in Figure 7(b,c). For the η- and T_1-values the respective correlation coefficients to experimental data were 0.74 and 0.15, and the respective RMSDs were 0.16 and 0.44 s^{-1} (see the Supporting Information).

As found previously in a simulation of ubiquitin,[15] some of the residues had T_1-values that were significantly larger and η-values that were significantly lower than the experimental data. This was primarily due to rapid local motion (on the picosecond time scale), rather than long time scale flexibility, perhaps due to the libration of the peptide plane. To investigate this, the calculation was repeated but this time using the N–C$^\alpha$-bond rather than the N–H-bond, i.e., a bond located along the axis of motion of the peptide plane. Due to the increased order parameters resultant from such a calculation, Figure 7(a), the overall correlation time that fitted the experimental data had to be increased to 5.5 ns. The correlation coefficient to experimental data increased to 0.57,

and the RMSD decreased to 0.11 (see the Supporting Information). Calculations using these values are shown using thick dotted lines in Figure 7(b,c). It can be seen that the fit is much closer to the experimental data than calculations using the N–H-bond. For the η- and T_1-values the respective correlation coefficients to experimental data were 0.74 and 0.28, and the respective RMSDs were 0.09 and 0.26 s^{-1} (see the Supporting Information). A disagreement between theory and experiment is still present around residue 80, corresponding to the $\beta5$-$\beta6$ loop where most of the NOE violations were found.

3.5. Comparison with Scalar Coupling Data. The average conformation of the backbone can be assessed using vicinal scalar-couplings measured from NMR spectra. One coupling that can be measured routinely at most residues is the vicinal coupling ($^3J_{HNH\alpha}$) between H^N and H$^\alpha$, which has a high value in β-sheets ($\sim$10 Hz) and a low value in α-helices ($\sim$4 Hz). Figure 8(a) shows that the couplings predicted from the simulation ensemble are in good agreement with experimental measurements (from an HMQC-*J*
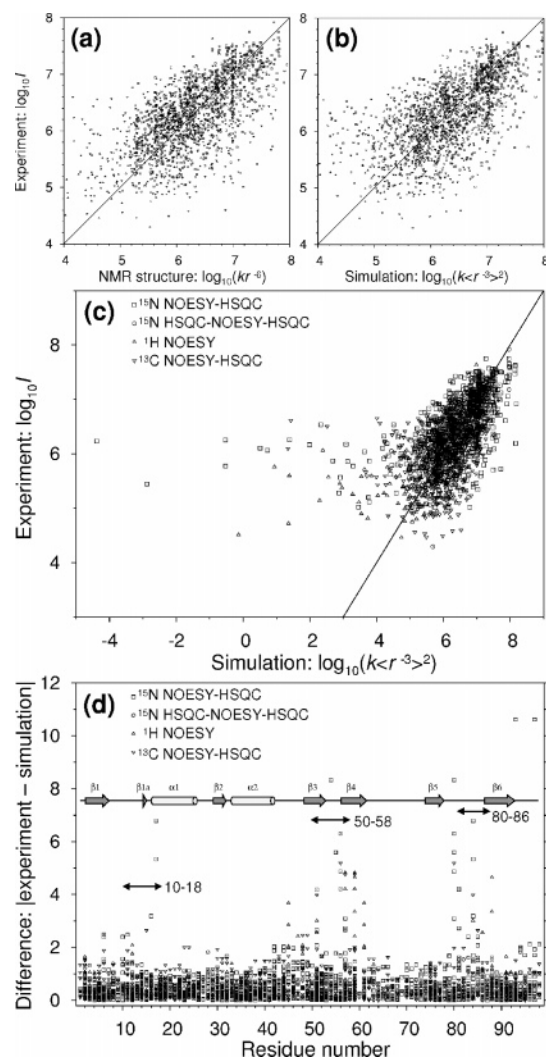
Figure 5. Correlation between the calibrated experimental NOESY measurements (log intensity, *I*) from four NMR data sets[26] and (a) the original lowest energy NMR structure and (b) the 0.25 μs simulation performed here. (c) Shows this correlation to the simulation for all of the predicted NOEs and (d) shows the difference between the log values along the backbone of the protein. Regions of disagreement are marked with double arrows.

experiment). In fact, the agreement with the simulation is better than the lowest energy NMR structure. The simulation ensemble and NMR structure have correlation coefficients to experimental data of 0.70 and 0.60, respectively, while the RMSDs to experimental data are 1.48 and 1.91, respectively (see the Supporting Information).

Pairs of (ϕ, ψ) values were extracted from the simulation, and scatter plots were constructed for every peptide linkage in the protein (see the Supporting Information). It was found that many of the linkages have multiple regions of exploration and, in particular, those N-terminal to glycine residues. For those linkages that had a single region of exploration (monomodal), an average and standard deviation can be meaningfully expressed. Most of the averages are restricted to the right-handed α-helix and β-region of the Ramachandran plot as expected (see the Supporting Information).

3.6. The Dynamic Structure of the Backbone and Side Chains. Having shown that the simulation predicts the
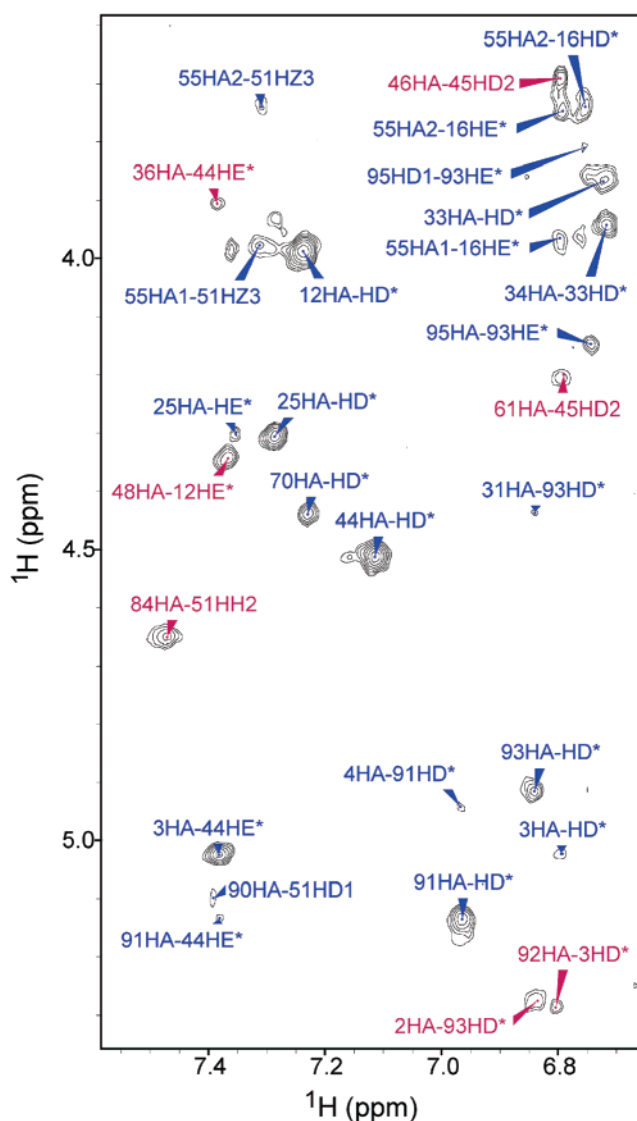

Figure 6. Part of the aromatic fingerprint region extracted from the ^{1}H-^{1}H NOESY experiment performed in D$_2$O on Link_TSG6, showing the NOEs that are, on average, predicted to be within 0.5 nm in the simulation (blue) and those that are not (red). Each peak is labeled with its NMR assignment. This region was purposely selected as one with more disagreements than average.

diverse range of experimental data as well as the static ensemble of NMR-structures (except at the β5-β6 loop) it can be used to inform us about nanosecond time scale dynamics in the protein backbone and side chains. Previously, it was shown that upon hyaluronan-binding, the β4-β5 loop opens under conformational rearrangements involving the C47$-$C68 disulfide bridge. It is therefore pertinent to examine further whether the free protein momentarily adopts such a conformation (which would substantially increase the protein surface accessible area). Figure 8(b) shows that the initial protein surface area (60 nm^2) decreases momentarily (to 52 nm^2) before rising again. Following this, the protein undergoes an oscillatory motion with the surface area varying between 53 nm^2 and 62 nm^2, over a period of 150 ns. Recalculation of the accessible surface area with the β4-β5 loop omitted indicated that this oscillatory motion did

Dynamics of the Hyaluronan-Binding Link Module

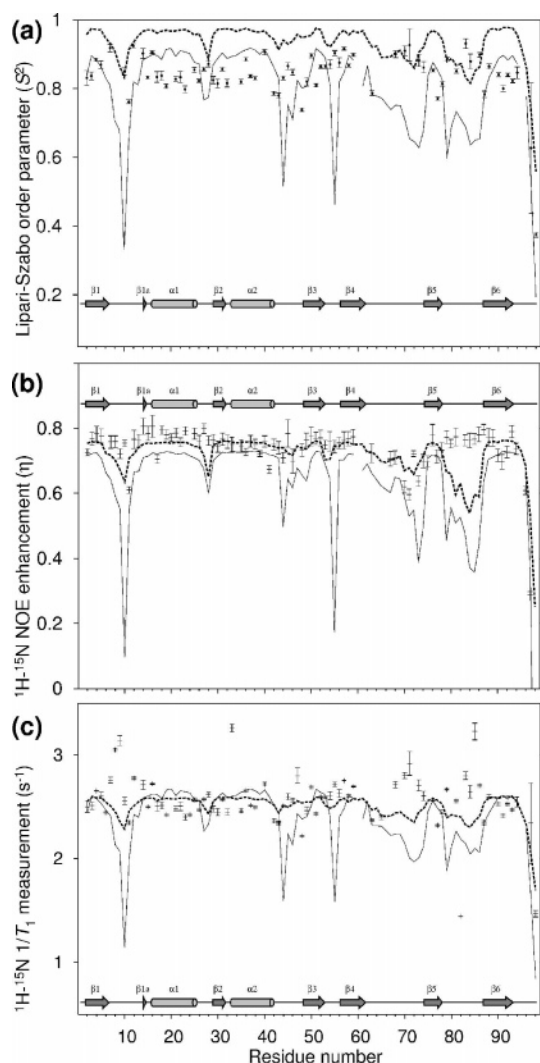*J. Chem. Theory Comput., Vol. 3, No. 1, 2007* **11**

Figure 7. (a) Calculated Lipari-Szabo order parameters from the simulation of Link_TSG6. The thin, continuous line represents order parameters calculated using the NH-vector and the thick, dotted line from the N−Cα-vector. (b,c) ^{15}N-relaxation parameters calculated from correlation functions calculated in the molecular frame and an overall tumbling time of 4.5 ns and 5.5 ns for the NH− and N−Cα-bonds, respectively. The experimental data are also plotted, using error bars.

not come from mobility of this loop and the momentary exposure of a binding groove. Further, the overall trend in the radius of gyration (calculated without mass weighting), shown in Figure 8(c), was similar to that seen in accessible surface area. Therefore, it is suggested (since the radius of gyration provides a view of global protein reorganization) that this trend in surface area simply comes from a slow overall breathing motion of the protein on the nanosecond time scale.

The flexibility of various parts of the protein simulation was quantified by calculating the RMSF (root-mean-square fluctuation) for N−H-groups in the molecular frame, Figure 9(a). In the majority of cases, the amide hydrogen atoms have more flexibility than their neighboring nitrogen atoms during the simulation because the peptide planes undergo local librational motion; this may account for the discrepancies between the calculated ^{15}N-relaxation measurements.
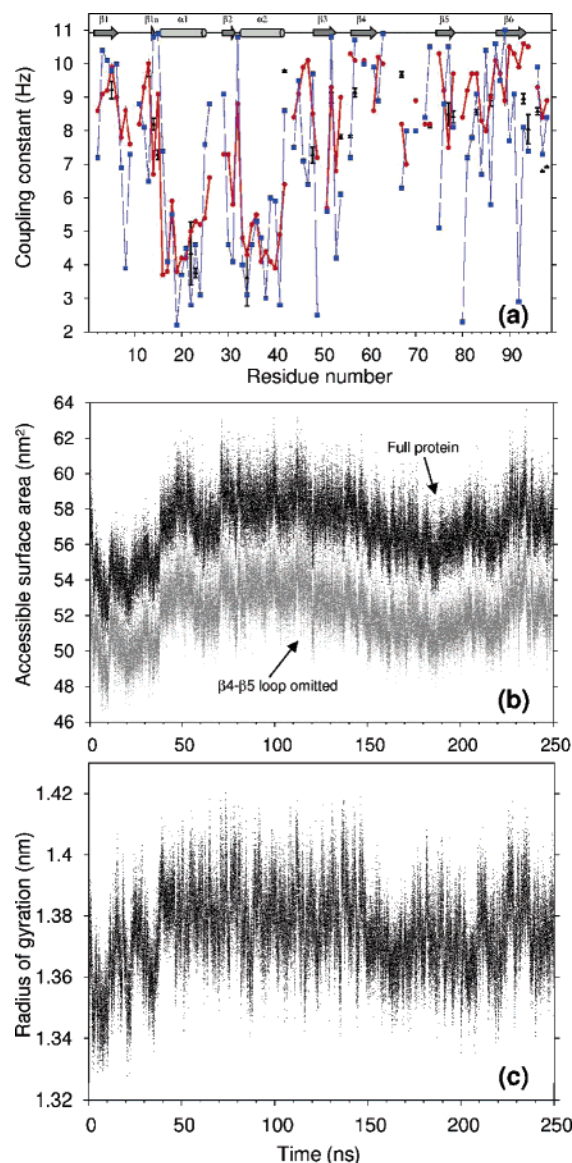

Figure 8. (a) Comparison between experimentally determined $^3J_{H\alpha HN}$ (using error bars to one standard deviation) and the values calculated from the simulation (blue squares and thin lines) and from the lowest energy NMR-derived structure (red circles and thick lines). (b) Calculated surface area for the whole protein as a function of simulation time (black) and for the protein with the $\beta4$-$\beta5$ loop omitted (gray). (c) Radius of gyration as a function of time for the whole protein (without mass weighting).

Amide groups not conforming to this general relationship occur exclusively around R81 and localize to regions of the protein simulation that disagreed with the experimental NOE measurements. The simulation predicts that the major points of flexibility of the amide-hydrogen atoms are K98, G10, N67, and F25 (RMSFs ~0.4 nm). Residues G10 and K98 have low ^{1}H-^{15}N NOE values, Figure 7(a), indicative of flexibility, and the side chain of N67 has previously been shown to be flexible.[51] The simulations also suggest that this flexibility is present in the adjacent nitrogen atoms, except at F25, which is one of the least flexible points of the backbone. Therefore, as expected, there is a good correlation between amide hydrogen and nitrogen RMSFs in the
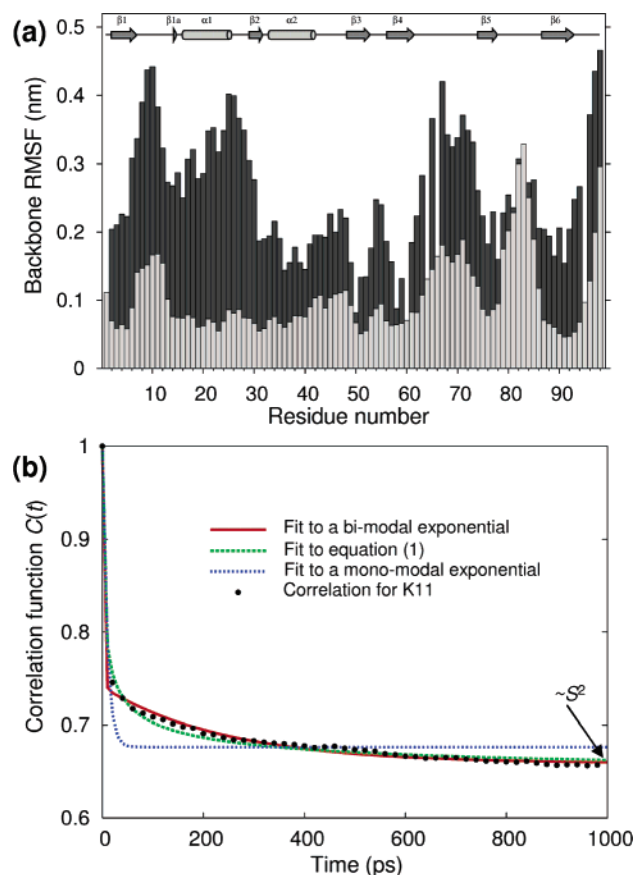
Figure 9. (a) Histogram of the root-mean-square fluctuation (RMSF) of the backbone H^N (black bars) and N (gray bars) atoms at specific residue positions. (b) Autocorrelation function for the N−H vector of K11 fitted to mono- and bimodal exponential and the equation described in eq 1 of the text.

simulation. From Figure 9(a), G50 and G58 have the most ordered amide hydrogen atoms. These residues are located in the core of the protein and form hydrogen bonds to each others' carbonyl oxygen atoms across the β3- and β4-strands, Figure 4. However, in the calculations of molecular frame order parameters (S^2) described above, Figure 7(a), these residues had S^2-values that were indistinguishable from residues in the other secondary structural elements. Furthermore, K98, G10, G55, and F44 had low calculated S^2 values, but the RMSFs of K98 and G10 are high ($\sim$0.4 nm), whereas, for G55 and F44, the RMSFs are lower ($\sim$0.2 nm). Therefore, there does not appear to be a relationship between the molecular frame order parameter (S^2) and amide-hydrogen RMSF.

The angular autocorrelation functions for N−H-vectors contain information on the time scale of motion of the backbone, i.e., the faster the decay, the faster the motion. It was found that calculated correlation functions ($C^{(i)}$) in the molecular frame did not fit very well to an exponential form as postulated originally by Lipari and Szabo.[45] Rather, interpretation of the correlation functions as an exponential resulted in ambiguities in establishing the decay constant; this may be due to the fact that it is theoretically impossible to define a consistent molecular frame for a dynamic molecule. However, the empirical eq 1 was found to fit all of the decay curves well, and a meaningful decay constant
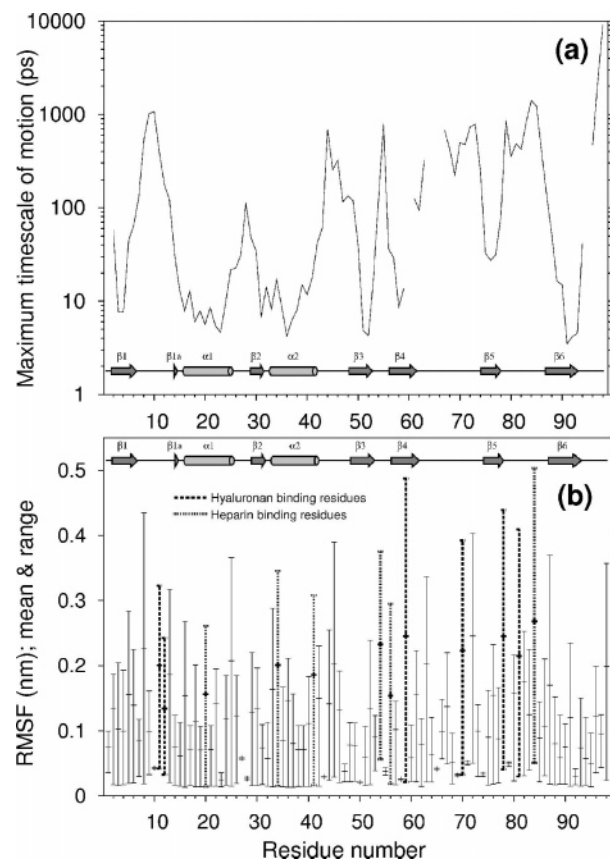

Figure 10. (a) The maximum time scale of internal motions at residue-specific positions calculated by fitting an empirical equation (see text) to the molecular frame correlation functions. (b) The minimum, maximum, and average RMSF of side chains are shown with key hyaluronan and heparin binding residues indicated. Note: K20, K34, K41, and K54 have been shown experimentally to bind heparin; R56 and R84 were predicted to be involved, but folded mutants could not be obtained for these residues.

(k) could be extracted to estimate the time scale of motion, Figure 9(b).

$$C^{(i)}(t) = S^2 + \frac{1 - S^2}{1 + (t/k)^{1/2}} \qquad (1)$$

For time $t = 361k$ the correlation function has decayed almost fully to $0.05(1 + 19S^2) \approx S^2$. Hence, this is a time parameter that represents the longest time scale correlations found at each amide hydrogen, and these are plotted for each residue in Figure 10(a). It is interesting to note that the backbone amide groups within the secondary structure elements are predicted to undergo rather rapid motion (picosecond time scale), whereas those in loops undergo motion on the nanosecond time scale. The tail is predicted to undergo motion on an even longer time scale.

3.7. Side-Chain Dynamics of Key Binding Residues. Figure 10(b) shows the average, maximum, and minimum RMSF found within each side chain in Link_TSG6; key residues that bind to hyaluronan and heparin are indicated. Residues involved in binding to glycosaminoglycans are particularly flexible in the free protein and particularly so at their ends, where they would make contact with ligand. One
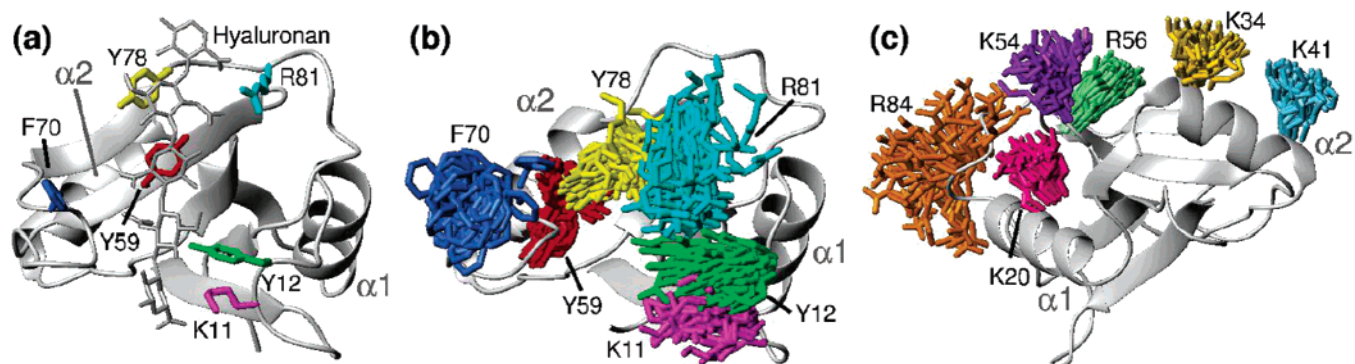
Dynamics of the Hyaluronan-Binding Link Module

J. Chem. Theory Comput., Vol. 3, No. 1, 2007 **13**

Figure 11. (a) The hyaluronan-bound NMR-structure of Link_TSG6 with a hyaluronan oligosaccharide modeled into the binding groove, reproduced from Blundell et al., with key residue indicated. (b,c) Overlay of 50 structures from the simulation of Link_TSG6 showing (b) the hyaluronan-binding residues, aligned with the model in (a) and (c) the heparin-binding residues with the protein viewed from the side.

of the most flexible side chains in the whole molecule is Y59, which is found at the center of the hyaluronan-binding groove in the structure of the complexed protein. However, this residue appears well defined in the family of NMR structures.[26] A possible explanation for this is that a large number of NOE cross-peaks are often observed to tyrosine residues, which may result in simulated annealing overconstraining them on the surface. Furthermore, the region K54–R84, while the simulation suggests that it contains the highest density of dynamic side chains in the molecule and many hyaluronan and potential heparin binding residues, has relatively slow backbone dynamics (nanosecond time scale), which may have an important part to play in ligand specificity and binding.

Figure 11(a) shows the lowest-energy NMR-structure of hyaluronan-bound Link_TSG6,[26,32] while in Figure 11(b), the 50 conformations of the same residues from the simulation of the free protein are given for comparison. All of the residues can be seen to be particularly dynamic, and their orientation is largely projecting into solution away from the protein surface. It can also be seen clearly that residue Y78 moves closer to Y59, due to unfolding of the $\beta5$-$\beta6$ loop. In Figure 11(c), the heparin binding side chains (which are on another face of the protein) are shown. All these basic side chains are again projecting away from the protein surface, but in this case are relatively more ordered (except for R84, which is probably partly due to movement of the unfolded $\beta5$-$\beta6$ loop).

4. Discussion

The 0.25 μs simulation of Link_TSG6 performed here, while as good a fit to the experimental NOE data as static NMR-structures outside of the $\beta5$-$\beta6$ loop, has the ability to provide additional new insight into protein dynamics. Deviations from the experimental data were almost exclusively localized to the $\beta5$-$\beta6$ loop, which lost its original structure and became disorganized within the first 50 ns of simulation time, Figure 3(a). In the rest of the protein, the RMSF of backbone nitrogen atoms was less than 0.1 nm for the majority of the backbone, Figure 9(a); notable exceptions were found in the backbone around G10, R81, and N67 and at the protein C-terminus (K98). Similar observations were made in two

simulations of ubiquitin, started from either NMR and XRC structures,[15] where both trajectories were found to be stable with an average RMSF for the backbone atoms of around 0.1 nm. Comparison with NMR NOESY cross-peaks from ubiquitin showed that only four restraints were violated by more than 0.05 nm in the NMR-trajectory, while the XRC trajectory had slightly more, with 13 violations (out of 985 unambiguous restraints); this is a total of 1.3% violations in ubiquitin compared to 2.2% in the simulation presented here. As we have found for the $\beta5$-$\beta6$ loop of Link_TSG6, all the ubiquitin violations were restricted to one section of the protein.

The simulation of Link_TSG6 correctly predicted high-occupancy intramolecular hydrogen bonds in secondary structure where they had been experimentally determined on the basis of amide exchange rates and NOE patterns but were predicted to extend further out from the ends of β-strands than previously reported.[26] The most occupied hydrogen bonds were localized to α-helices and toward the middle of β-sheets. A highly occupied tertiary structure hydrogen bond was predicted linking A53H^N (in the $\beta3$-$\beta4$ loop) to A31O (in strand $\beta2$) that has not been documented previously; this could account for the high-resolution obtained for the A53 methyl side chain during the NMR-structure determination, which was surprising given that A53 is the central residue in a β-turn. The simulations also predicted that the single slowly exchanging hydroxyl group of T32, found by NMR measurements,[26] was solvent exposed and not involved in extensive hydrogen bonds. The reason for this slow exchange, therefore, still remains a mystery, although it may indicate an ion-binding site. In a previous 3 ns simulation of bacterial cytochrome c,[17] intramolecular hydrogen bonds were used as a basis for calculating amide protection factors. While amide hydrogen bonds were found throughout the protein, their fractional occupancy was highest in the helices, giving them the highest amide protection factors. However, bacterial cytochrome c does not possess β-sheets, so they could not be considered. In the present simulation, it was found that high amide hydrogen protection can be found in both helices and sheets. Within β-sheets, the most central strands ($\beta3$, $\beta6$) have the highest protection factors, as might be expected.

The aqueous simulation of Link_TSG6 predicted a rotational tumbling time of 2.4 ns, which is lower than that calculated (4.5−5.5 ns) from ^{15}N-relaxation measurements. This compares favorably with a previous aqueous simulation of ubiquitin,[15] which had an even greater disagreement (0.74 ns as opposed to an experimental value of 4.03 ns), suggesting that the current simulation is making reasonably accurate predictions on the nanosecond time scale. Both SPC (used in the ubiquitin simulations) and TIP3P (used here) display good thermodynamic properties, but their kinetic properties differ from the experimental values, and, in particular, they severely overestimate the diffusion coefficient of water while underestimating bulk water viscosity.[52−54] These disagreements underline the fact that molecular mechanics simulations of this type, i.e., using simple models of water, cannot predict the overall tumbling rates for proteins. In spite of these disagreements, the diffusional anisotropy ($D_{||⊥}$) calculated for ubiquitin was close to the experimental value of 1.17.[15] The diffusional anisotropy calculated for Link_TSG6 is similar to ubiquitin and hence may be a good predictor of the experimental value (not yet experimentally determined).

The simulation of Link_TSG6 appears to contain too much fast time scale (subnanosecond) fluctuation at the amide hydrogens, which results in a rapid initial decay of the angular correlation function and order parameters that are lower than those required to predict the ^{15}N-relaxation parameters. The flexibility of the amide hydrogen around F25, in particular, is not reflected in the RMSD of the neighboring nitrogen backbone, Figure 9(a), nor is reflected in the ^{15}N-heteronuclear NOE measurements, Figure 7(b). The apparent disagreement between the heteronuclear NMR data and back-calculations from simulations is therefore likely to be due to rapid local librations of the peptide planes. Simulations of ubiquitin exhibited similar phenomena, and the calculated order parameters deviated from the experimentally derived values. Interestingly, however, it was found that internal motions in the 2−6 ns range are not well represented by heteronuclear relaxation data because they are obscured by relaxation due to the overall tumbling time of the molecule, which is of the same order. Therefore, ^{15}N-relaxation experiments that are traditionally used to infer dynamic information in proteins may be largely insensitive to important nanosecond time scale dynamics and vastly more sensitive to picosecond time scale dynamics. Simulations therefore currently play an essential role in investigating motions on these longer nanosecond timescales that are not accessible to NMR, but it is crucial that each simulation is tested against experimental data to assess its general validity before conclusions are drawn.

Residues Y78 to W88 in the $\beta5$-$\beta6$ loop are seen to have a high flexibility in the simulation, and the mean structure deviates from the NMR-derived conformation. Comparison with experimental data suggest that this motion is not realistic. This $\beta5$-$\beta6$ loop of Link_TSG6 is surrounded by several acidic and basic residues (D77, R81, R84, E86, R87, and D89), which suggests that there are a complex set of electrostatic interactions. Since the original structure calculations may not have placed the side chains in ideal positions,

this could have led to some unfavorable electrostatic interactions that destabilized the loop. The unfolding of the loop leads to Y78 moving to be in the proximity of R81, as shown in Figure 11(b). However, it should be noted that this loop is the longest in the protein (twice as long as any other) and has no secondary structure, disulfide bridges, hydrogen bonds, or salt-bridges to stabilize it. Furthermore, it is also the location of a sizable insertion in the second Link module of the hyaluronan-binding domain of the Type C members of the Link module superfamily.[32]

Furthermore, it is known that CHARMM22 does not model the α-L region of (ϕ,ψ) potential energy surface (including glycines) by significantly overestimating the relative energy, which may explain some of these observed anomalies within the loops of TSG-6.[55,56] Simulations performed using the CMAP correction to the CHARMM22 force-field may be able to more accurately model the protein and alleviate some of these problems.[35]

The hyaluronan-binding $\beta4$-$\beta5$ loop was difficult to define by NMR and simulated annealing because fewer NOEs per residue than average could be obtained[26] since it contains two proline residues (P64 and P66) and three glycine residues (G65, G69, and G71). Due to this, the $\beta4$-$\beta5$ loop had a similar flexibility in both the calculated NMR ensemble and the simulation, which is largely fortuitous. The simulation also indicated that other loops in the protein are more flexible than the NMR ensemble suggests, probably because simulated annealing results in overconstraining when there is a large amount of NOE data, i.e., it purposely uses experimental data to generate the impression of a static structure when no such structure exists in solution. This overconstraining may be due to the fact that there are often more NOE restraints than rotatable degrees of freedom in the molecule.

The simulations, while long (0.25 μs), provided no evidence for the opening/closing motion of the hyaluronan-binding $\beta4$-$\beta5$ loop on the nanosecond time scale. This suggests that the bound conformation of this loop is not sampled in unbound protein, further indicating an induced-fit mechanism for binding of hyaluronan.[26] A second set of NMR-assignments was found for the $\beta4$-$\beta5$ loop at I61, V62, and K63, which was hypothesized to be due to cis−trans isomerization of P64 or P66.[26] It should be noted that the current simulation is not capable of investigating such long-time scale phenomena.

It was also found that the backbone of loops and the C-terminal tail were involved in slower time scale motions (nanosecond) than secondary structural elements, which were involved in faster time scale motions (picosecond). In particular, the backbone within elements α1, α2, $\beta3$, and $\beta6$ have the fastest time scale motions (less than 10 ps), which interestingly coincide with the regions in the protein that have the highest amide hydrogen exchange protection factors. In fact, it could be hypothesized that there is an inverse relationship between time scale of local dynamic motion and amide hydrogen protection factors (EX2-regime) according to this calculation.

Surface side chains are generally not well characterized in NMR structure ensembles since there are relatively few

Dynamics of the Hyaluronan-Binding Link Module

J. Chem. Theory Comput., Vol. 3, No. 1, 2007 **15**

NOE constraints to side chain nuclei. A consequence of this is that NMR ensembles generally show them in a variety of conformations, some flat against the protein surface, simply due to lack of experimental data. The most favorable configuration for many of these side chains is to be pointing away from the protein surface and highly solvent exposed, as shown in Figure 11(c). This is particularly important for understanding how heparin binds to this Link-module, which is likely to occur via the positive charge at the end of the long and flexible side chains of lysine and arginine residues. However, it should be pointed out that side chains are the most difficult part of the protein to characterize experimentally, and the predictions made for side chains by the simulation are largely untested at this stage.

In the NMR-determined structure of Link_TSG6 complexed with hyaluronan,[26] Figure 11(a), Y59 and Y78 lie flat against the protein surface and are highly ordered, even unusually displaying distinct H^{ϵ}/H^{δ} chemical shifts on each side of the ring. In contrast, in the free protein they project outward and are highly mobile, having only one chemical shift for the H^{δ} and H^{ϵ} ring protons. This is reflected in the simulation, which predicts that amino acids known to be involved in hyaluronan, bikunin, and heparin binding are among the most mobile in the protein. Therefore, the specific, average side-chain orientations determined from the few observed NOEs is a gross simplification arising from simulated annealing, which does not account for side-chain mobility. However, it may be possible to gain further insight into the specific dynamics of side chains by other methods and simultaneous use of NOEs and residual dipolar couplings[57] or by improving the experimental data set by using specific isotopic labeling.[58]

This new representation of the Link-module of TSG-6 provides information about dynamics on the nanosecond time scale while being as good a fit to experimental data over the majority of the structure as a static ensemble derived by simulated annealing. It will be invaluable for understanding the dynamic molecular basis of its interactions with ligands and the role and conservation of individual amino acids in the Link-module superfamily. Ultimately, these advances will underpin a dynamic understanding of extracellular matrix assembly, remodeling, and the changes that occur in physiological and disease states.

Abbreviations: XRC − X-ray crystallography; NMR − nuclear magnetic resonance; MD − molecular dynamics; NOE − nuclear Overhauser effect; NOESY − NOE spectroscopy; HSQC − heteronuclear single quantum coherence; HMQC − heteronuclear multiple quantum coherence; CSD − chemical shift deviation; TSG-6 − tumor necrosis factor stimulated gene-6; Link_TSG6 − recombinant link-module domain from human TSG-6; RMS − root-mean-square; RMSD − root-mean-square deviation; RMSF − root-mean-square fluctuation.

Supporting Information Available: Temperature coefficients and chemical shift deviations for Link_TSG6, experimental order parameters, scatter plots for each of the (φ,ψ)-angles from the molecular dynamics simulation of Link_TSG6, average and standard deviation for these (φ,ψ)-angles, and graphs of the simulation predictions (order parameters, relaxation measurements. and scalar couplings) plotted against experimental measurements with correlation coefficients and RMSDs. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) D'Aquino, J. A.; Freire, E.; Amzel, L. M. *Proteins* **2000**, *S4*, 93−107.

(2) Monecke, P. *Biophys. J.* **2006**, *90*, 841−850.

(3) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11287−11292.

(4) Kimura, S. R.; Brower, R. C.; Vajda, S.; Camacho, C. J. *Biophys. J.* **2001**, *80*, 635−642.

(5) Wider, G. *Biotechniques* **2000**, *29*, 1278−1290.

(6) Dauter, Z. *Acta Crystallogr. D* **2006**, *62*, 1−11.

(7) Furnham, N.; Blundell, T. L.; DePristo, M. A.; Terwilliger, T. C. *Nat. Struct. Mol. Biol.* **2006**, *13*, 184−185.

(8) Derewenda, Z. S.; Vekilov, P. G. *Acta Crystallogr. D* **2006**, 62, 116−124.

(9) Eyal, E.; Gerzon, S.; Potapov, V.; Edelman, M.; Sobolev, V. *J. Mol. Biol.* **2005**, *351*, 431−442.

(10) Schwieters, C. D.; Clore, G. M. *J. Biomol. NMR* **2002**, *23*, 221−225.

(11) Schröder, G. F.; Alexiev, U.; Grubmüller, H. *Biophys. J.* **2005**, *89*, 3757−3770.

(12) Farrow, N. A.; Muhandiram, D. R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984−6003.

(13) Malliavin, T. E. *Curr. Org. Chem.* **2006**, *10*, 555−568.

(14) Clore, G. M.; Schwieters, C. D. *J. Mol. Biol.* **2006**, *355*, 879−886.

(15) Nederveen, A. J.; Bonvin, A. *J. Chem. Theor. Comp.* **2005**, *1*, 363−374.

(16) Camacho, C. J. *Proteins: Struct., Funct., Bioinform.* **2005**, *60*, 245−251.

(17) Kieseritzky, G.; Morra, G.; Knapp, E. W. *J. Biol. Inorg. Chem.* **2006**, *11*, 26−40.

(18) Tammi, M. I.; Day, A. J.; Turley, E. A. *J. Biol. Chem.* **2002**, *277*, 4581−4584.

(19) Day, A. J.; Prestwich, G. D. *J. Biol. Chem.* **2002**, *277*, 4585−4588.

(20) Day, A. J.; Sheehan, J. K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 617−622.

(21) Day, A. J.; de la Motte, C. A. *Trends Immunol.* **2005**, *26*, 637−643.

(22) Milner, C. M.; Higman, V. A.; Day, A. J. *Biochem. Soc. Trans.* **2006**, *34*, 446−50.

(23) Milner, C. M.; Day, A. J. *J. Cell. Sci.* **2003**, *116*, 1863−73.

(24) Kohda, D.; Morton, C. J.; Parkar, A. A.; Hatanaka, H.; Inagaki, F. M.; Campbell, I. D.; Day, A. J. *Cell* **1996**, *86*, 767−75.

(25) Mahoney, D. J.; Blundell, C. D.; Day, A. J. *J. Biol. Chem.* **2001**, *276*, 22764−71.

(26) Blundell, C. D.; Mahoney, D. J.; Almond, A.; DeAngelis, P. L.; Kahmann, J. D.; Teriete, P.; Pickford, A. R.; Campbell, I. D.; Day, A. J. *J. Biol. Chem.* **2003**, *278*, 49261−70.

(27) Mahoney, D. J.; Mulloy, B.; Forster, M. J.; Blundell, C. D.; Fries, E.; Milner, C. M.; Day, A. J. *J. Biol. Chem.* **2005**, *280*, 27044−27055.

(28) Almond, A.; Brass, A.; Sheehan, J. K. *J. Mol. Biol.* **1998**, *284*, 1425−1437.

(29) Almond, A.; DeAngelis, P. L.; Blundell, C. D. *J. Am. Chem. Soc.* **2005**, *127*, 1086−1087.

(30) Mulloy, B.; Forster, M. J. *Glycobiology* **2000**, *10*, 1147−56.

(31) Verli, H.; Guimaraes, J. A. *Carbohydr. Res.* **2004**, *339*, 281−90.

(32) Blundell, C. D.; Almond, A.; Mahoney, D. J.; DeAngelis, P. L.; Campbell, I. D.; Day, A. J. *J. Biol. Chem.* **2005**, *280*, 18189−201.

(33) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(34) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(35) Buck, M.; Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Jr. *Biophys. J.* **2006**, *90*, L36−L38.

(36) Jorgensen, W. L.; Chandasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(37) Tsodikov, O. V.; Record, M. T., Jr.; Sergeev, Y. V. *J. Comput. Chem.* **2002**, *23*, 600−609.

(38) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *J. Am. Chem. Soc.* **2003**, *125*, 15686−7.

(39) Best, R. B.; Vendruscolo, M. *Structure* **2006**, *14*, 97−106.

(40) Englander, S. W.; Sosnick, T. R.; Englander, J. J.; Mayne, L. *Curr. Opin. Struct. Biol.* **1996**, *6*, 18−23.

(41) Blundell, C. D.; Deangelis, P. L.; Almond, A. *Biochem. J.* **2006**, *396*, 487−98.

(42) Andersen, N. H.; Neidigh, J. W.; Harris, S. M.; Lee, G. M.; Liu, Z. H.; Tong, H. *J. Am. Chem. Soc.* **1997**, *119*, 8547−8561.

(43) Cierpicki, T.; Otlewski, J. *J. Mol. Biol.* **2000**, *302*, 1179−1192.

(44) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035−6043.

(45) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

(46) Brüschweiler, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175−83.

(47) Bruschweiler, R.; Case, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 11199−11200.

(48) Zagrovic, B.; van Gunsteren, W. F. *Proteins: Struct., Funct., Bioinform.* **2006**, *63*, 210−218.

(49) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1997**, *36*, 2390−402.

(50) Krishnan, V. V.; Cosman, M. *J. Biomol. NMR* **1998**, *12*, 177−182.

(51) Kahmann, J. D.; O'Brien, R.; Werner, J. M.; Heinegard, D.; Ladbury, J. E.; Campbell, I. D.; Day, A. J. *Struct. Fold. Des.* **2000**, *8*, 763−774.

(52) Guo, G. J.; Zhang, Y. G. *Mol. Phys.* **2001**, *99*, 283−289.

(53) Smith, P. E.; Vangunsteren, W. F. *Chem. Phys. Lett.* **1993**, *215*, 315−318.

(54) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. *Chem. Phys. Lett.* **2006**, *418*, 245−249.

(55) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III *J. Am. Chem. Soc.* **2004**, *126*, 698−9.

(56) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(57) Peti, W.; Meiler, J.; Bruschweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822−33.

(58) Rosen, M.; Gardner, K.; Willis, R.; Parris, W.; Pawson, T.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 627−636.