Interdependence of Backbone Flexibility, Residue Conservation, and Enzyme Function: A Case Study on β1,4-Galactosyltransferase-I†

K. Gunasekaran,‡ Buyong Ma,‡ B. Ramakrishnan,§,|| Pradman K. Qasba,*,§ and Ruth Nussinov*,||,⊥

Laboratory of Experimental and Computational Biology, Structural Glycobiology Section, and Basic Research Program, SAIC-Frederick, Inc., Laboratory of Experimental and Computational Biology, NCI-Frederick, Frederick, Maryland 21702, and Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Received January 10, 2003; Revised Manuscript Received February 5, 2003

ABSTRACT: β1,4-Galactosyltransferase-I (β4Gal-T1) catalyzes the transfer of a galactose from UDP-galactose to *N*-acetylglucosamine. A recent crystal structure determination of the substrate-bound enzyme reveals a large conformational change, which creates binding sites for the oligosaccharide and α-lactalbumin, when compared to the ligand-free structure. The conformational changes take place in a 21-residue-long loop (I345−H365) and in a smaller loop containing a tryptophan residue (W314) flanked by glycines (Y311−G316; Trp loop). A series of molecular dynamics simulations carried out with an implicit solvent model and with explicit water successfully identify flexibility in the two loops and in another interacting loop. These observations are confirmed by limited proteolysis experiments that reveal an intrinsic flexibility of the long loop. The multiple simulation runs starting with the substrate-free structure show that the long loop moves toward its conformation in the ligand-bound structure; however, it gets stabilized in an intermediate position. The Trp loop moves in the opposite direction to that of the long loop, making contacts with residues in the long loop. Remarkably, when the Trp loop is restrained in its starting conformation, no large conformational change takes place in the long loop, indicating residue communication of flexibility. Sequence and structural analysis of the β4Gal-T1 family with 37 known sequences reveals that in contrast to the unconserved long loop, which undergoes a much larger conformational change, the Trp loop including the glycines is highly conserved. These observations lead us to propose a new functional mechanism that may be conserved by evolution to perform a variety of functions.

Proteins are dynamic systems. Their internal motions enable them to perform specific functions (*1*−*6*). The dynamics involved in protein−ligand and protein−protein interactions have been demonstrated to be fundamentally linked to function in several cases [examples include HIV protease (*7*), GroES (*8*), src family kinases (*9*), citrate synthase (*10*), and triosephosphate isomerase (*2*)]. Nuclear magnetic resonance (NMR)[1] studies have demonstrated that movements of chain ends and fluctuation of surface side chains and loops occur on experimental time scales (*11*). NMR ^{15}N relaxation, neutron diffraction, and hydrogen exchange are among the experimental techniques used to study internal motions. Computer simulations (molecular dynamics, MD) are often used in conjunction with experimental data to understand the important motions (*12*). Although the simulation time scale is usually shorter than many biologically interesting large-scale motions, MD can still provide valuable insight for further experimental investigation (*13*). Despite extensive research and consistency between experimental and theoretical studies over the involvement of loops in enzyme function, identification of

† The research of R.N. in Israel has been supported in part by the Magnet grant, by the Center of Excellence in Geometric Computing and its Applications funded by the Israel Science Foundation (administered by the Israel Academy of Sciences), by the Ministry of Science grant, and by the Tel Aviv University Basic Research grants. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract NO1-CO-12400. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

* Correspondence should be addressed to R.N. or P.K.Q. at NCI-Frederick, Building 469, Room 151 or 221, Frederick, MD 21702. Tel: 301-846-5579 or 301-846-1933. Fax: 301-846-5598. E-mail: ruthn@ncifcrf.gov or qasba@helix.nih.gov.

‡ Laboratory of Experimental and Computational Biology, NCI-Frederick.

§ Structural Glycobiology Section, Laboratory of Experimental and Computational Biology, NCI-Frederick.

|| Basic Research Program, SAIC-Frederick, Inc., Laboratory of Experimental and Computational Biology, NCI-Frederick.

⊥ Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University.

[1] Abbreviations: β4Gal-T1, β1,4-galactosyltransferase-I; UDP-gal, UDP-galactose; GlcNAc, *N*-acetylglucosamine; EEF1, effective energy function; ABNR, adopted basis Newton−Raphson; RMSD, root-mean-square deviation; MD, molecular dynamics; U5P, uridine 5′-monophosphate; NMR, nuclear magnetic resonance; ASA, accessible surface area.

10.1021/bi034046r CCC: $25.00
Published on Web 03/14/2003

the functional loops from the growing sequence database and, importantly, their role in function remain difficult tasks.

Flexibility plays a central role in the function of β1,4-galactosyltransferase-I (β4Gal-T1). Here, we study this enzyme through simulations, sequence and structural analyses, and limited proteolysis experiments. β4Gal-T1 is a 402-residue-long Golgi membrane enzyme that catalyzes the transfer of galactose from UDP-galactose (UDP-gal) to *N*-acetylglucosamine (GlcNAc; *14*). It also has a well-defined second function as the catalytic component of lactose synthase, which is a 1:1 complex of β4Gal-T1 and α-lactalbumin (*15*). Lactose is a major component of milk. The α-lactalbumin binding domain of this enzyme predates the rise of mammals (*16*). β4Gal-T1 belongs to a very large superfamily of glycosyltransferases (*17*, *18*). These transferases are type II membrane proteins. They synthesize the complex oligosaccharide moiety on the glycoproteins and glycolipids in the Golgi apparatus. Although there is no sequence homology among the superfamily, subfamily members exhibit a high sequence identity, particularly in their catalytic domain. Nearly 100 mammalian glycosyltransferases are known to date. However, the crystal structures of only five are currently available. Interestingly, although these proteins are known to exist in two conformational states corresponding to the absence and presence of the donor substrate (*19*), only for the case of β4Gal-T1 is the structure known in both conformations (*20*, *21*). Thus, β4Gal-T1 uniquely enables dynamics and structural studies of the functional mechanism not only of β4Gal-T1 but also of the very large glycosyltransferase superfamily which shares the same fold. Understanding the mechanism should assist in engineering catalysts for novel glycoconjugate synthesis (*22*).

The crystal structure of β4Gal-T1 was first determined by Gastinel and co-workers (referred to as "conf1"; crystallized in the absence of the substrate UDP-gal; Figure 1, shown in purple; *20*). Crystals were obtained only when the N-terminus 114-residue segment was cleaved off (*20*). The truncated protein (G115−S402) had catalytic activities identical to those of the wild type. Although the protein crystal diffracted to 2.4 Å, residues G115−S130 could not be located in the electron density map, indicating segment disorder. Gastinel et al. soaked the β4Gal-T1 crystals in a buffer containing UDP-gal to determine the binding site of the substrate. However, only the UDP group, and not the galactose, was localized in the structure. Furthermore, the GlcNAc and the α-lactalbumin interaction sites could not be inferred.

To identify the α-lactalbumin binding sites, Ramakrishnan and Qasba redetermined the crystal structure of β4Gal-T1 (sequence from S130 to S402) in complex with α-lactalbumin and UDP-gal/GlcNAc ("conf2"; substrate bound; Figure 1, shown in cyan) at 2.0 Å resolution (*21*). The substrate-bound crystal structure of β4Gal-T1 revealed a large conformational change when compared with the earlier structure (conf1). It was also noted that the conformational change was not due to the interaction with α-lactalbumin since they obtained a similar structure (conf2; complexed with UDP-gal) even in the absence of α-lactalbumin (*23*). The conformational change occurs in two regions of β4Gal-T1. The first region is a long loop comprised of residues I345− H365 (some of the residues move as far as 20 Å). The second region is a short loop containing a tryptophan

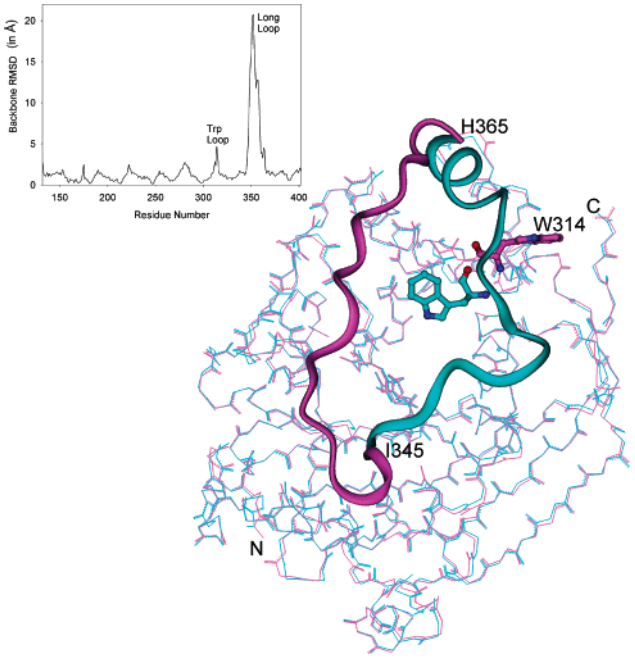

FIGURE 1: C^{α} Superposition of crystal structures of β4Gal-T1 crystallized without (purple, conf1) and with (cyan, conf2) the ligand UDP-gal. A significant structural change is seen in a long loop (I345 to H365, shown in ribbon representation) and in a short loop (Trp loop, Y311 to G316) containing a tryptophan residue (W314, represented in ball-and-stick model). Only the backbone atoms are shown for the rest of the molecule, for clarity. The inset shows the residue-wise root-mean-square deviation (RMSD) for the backbone atoms, calculated from the C^{α} superposed coordinates.

(W314) residue flanked by glycines (Y311−G316; Trp loop). Whereas the Trp loop is buried under the long loop and W314 is hydrogen bonded to the ligand UDP-gal in conf2, it is surface exposed and far from the long loop in conf1. Earlier spectroscopic experiments have shown that a tryptophan residue must interact directly with the negative charge of the UDP-gal and it must undergo conformational change upon ligand binding (*24*). Further, cross-linking experiments have shown that β4Gal-T1 does not interact with α-lactalbumin in the absence of UDP-gal or GlcNAc (*21*, *25*).

A comparison of the two crystal structures also reveals a difference in the quaternary interactions, raising the question of whether the conformational change is induced by the substrate or by this quaternary difference. β4Gal-T1 forms a dimer in both conf1 and conf2 crystals. In conf1, the long loop is at the interface between the two β4Gal-T1 subunits. The C-terminal end of the long loop beginning with R359 (specifically, R359, D361, and I363) makes several contacts with the second subunit. This suggests that in conf1 the long loop could be stabilized by the subunit−subunit interactions. The subunit arrangement in conf2 is different, with the long loop being at the interface of β4Gal-T1 and α-lactalbumin. There are also differences in the methods of protein expression, purification, and crystallization between conf1 and conf2. In conf1, β4Gal-T1 is a glycoprotein obtained from mammalian cells. The sequence initiates with residue G115. The short N-terminal segment from G115 to S130 was found to be disordered in the crystal. In conf2, β4Gal-T1 was expressed and purified as inclusion bodies from *Escherichia coli*, and the sequence starts from residue S130. Despite many attempts, this protein could not be crystallized in the conf1 conformation (i.e., without ligand). However, it had all the

functional activities and could be crystallized with a ligand (for example, with UDP-gal). Hence, two questions arise: First, is it possible to engineer the conf2 sequence (S130−S402) so that it could be crystallized in the absence of a ligand (i.e., in conf1)? Second, is there an energy barrier for the transition between conf1 and conf2?

Here, we first identify the flexible regions in β4Gal-T1 through MD simulations. We compare the results with the flexible regions identified by limited proteolysis experiments. We then investigate the origin of the observed flexibility and how it correlates with sequence conservation in the β4Gal-T1 family. Our results demonstrate that the intrinsic flexibility associated with the conserved Trp loop is communicated to the long loop through hydrophobic and hydrogen-bonding interactions. This communication lowers the energy barrier for the transition between conf1 and conf2. We propose that such a mechanism may effectively be used by evolution to perform a variety of functions.

MATERIALS AND METHODS

Molecular Dynamics Simulations. A series of MD simulations in implicit solvent and explicit water were carried out in order to explore the dynamic behavior of β4Gal-T1 (Table 1). In the implicit solvent simulations, the protein is represented in atomic detail but the solvent is represented only in terms of its bulk properties. The advantage of using the implicit solvent model is that it allows a greater sampling of the conformational space by increasing the speed of the calculations. As a result of the absence of solvent caging and frictional forces, the effective time span of the implicit solvent simulation is longer than the actual duration of the simulation. For the implicit simulations, we used the effective energy function (EEF1) combined with the CHARMM19 polar hydrogen potential energy function (*26*, *27*). MD simulations performed in house and elsewhere using the EEF1 implicit solvent model have shown that native proteins are stable with the EEF1, and the observed deviations agree with the simulations in explicit water (*27*−*30*). Further, the EEF1 has been shown to discriminate between the native and the misfolded proteins (*31*). Most of the simulations discussed here are performed using the EEF1 force field, but we have also carried out simulations in explicit water to confirm our conclusions.

To compare the simulations starting with conf1 and conf2, we considered the segment comprising residues T132−S402 for which coordinates are available in both conformations (PDB code: conf1, 1FGX A; conf2, 1J8X B). We also mutated I158 in the conf1 crystal structure to V158 as in conf2 using InsightII (Accelrys Inc., San Diego). For the implicit simulations with the EEF1 force field, the crystal structures were first subjected to 300 steps of minimization. The adopted basis Newton−Raphson (ABNR) routine, which performs energy minimization using a Newton−Raphson algorithm, was used for this purpose (CHARMM package version c27b1; *26*). In the initial phase of heating, the simulation temperature was raised slowly stepwise, with the system preserving the crystallographic structure. At the final required temperature, the system was equilibrated for 50−100 ps with a time step of 2 fs. A shorter period of equilibration time is used in all our simulations, since we are interested in conformational changes that are taking place during the simulation, but at the same time we also need to eliminate any artifacts. Longer duration productive MD simulations with a time step of 2 fs followed the equilibration period. Bond lengths between hydrogen and heavy atoms were constrained using the SHAKE algorithm (*32*). The conformers were saved every 2 ps. To remove the center of mass motions, the conformers were superimposed on the minimized starting structure using InsightII. The structures were superimposed by considering either all backbone atoms or all C^{α} atoms as needed.

For the explicit simulations, in addition to the crystal structure, we also considered the associated crystal waters. A total of 81 crystal waters were included for conf1 and 96 waters for conf2. The systems were soaked in a cubic water box with edge length of 65 Å. Care was taken to ensure that the effective water density remained close to 1.0 g/cm^3 and an equal number of water molecules were present in the conf1 and conf2 systems (a total of 7608 water molecules were included). As in the case of the implicit simulations, the systems were subjected to minimization, followed by slow heating and equilibration before the productive runs. Periodic boundary condition was imposed during the simulations so that a constant number of water molecules would be maintained within the cubic box. The nonbonded interactions were truncated at 12.0 Å. A time step of 1 fs was used for the simulations, and the conformers were saved every 1 ps.

We also carried out simulations on β4Gal-T1 dimers. β4Gal-T1 exists as a dimer in the asymmetric unit with different quaternary arrangements. A set of 10 ns implicit solvent simulations starting with dimeric forms of conf1 and conf2 was carried out. Apart from the three identified flexible regions (D278−Y286, Y311−G316, and H347−E368), the rest of the protein chain behaved in a manner similar to that of the implicit simulations on monomers. The effects of subunit−subunit interactions on the three flexible loops are discussed later.

The more recent crystal structure of β4Gal-T1 in complex with UDP-gal (1KYB; *23*) is conformationally similar to the one reported earlier (1J8X, complexed with UDP-gal and α-lactalbumin; *21*). Superposition of these two structures (1J8X B and 1KYB A) resulted in the root-mean-square deviation (RMSD) of 0.5 Å for all of the backbone atoms. However, the coordinates for the ligand UDP-gal (in particular, for the galactose group) were available only in this crystal structure. Therefore, for simulations starting with the ligand-bound form of conf2, we made use of 1KYB A coordinates. We also modeled the ligand position in conf1 (1FGX A) on the basis of its location in conf2 (1KYB A) using InsightII. MD simulations in explicit water were carried out by starting with the UDP-gal-bound form of conf1 and conf2. A total of 7599 water molecules were present in the cubic box with an edge length of 65 Å. The simulations were performed following the procedure described above. Since the CHARMM package did not have parameters for UDP-gal, we created the parameter file by making use of available parameters in CHARMM for nucleic acids and sugars. The parameters for the linkage (phosphate) between these two groups, as well as the charges, were taken from the published work of Imberty and co-workers (*33*).

When multiple simulations were required, simulation parameters such as the random seed and equilibration period were modified in order to generate different trajectories. Multiple trajectories help to identify recurring features and to avoid artifacts arising from the simulation procedure (*34*). The simulations were carried out using eight R10K 250 MHz processors on the SGI Origin2000 of the NCI-Frederick Advanced Biomedical Computing Center as well as using the NIH Biowulf cluster of eight processors with CPU speed varying between 450 and 866 MHz (1 ns simulation in explicit water took about 18 and 21 days on the Biowulf cluster and Origin2000, respectively; 10 ns simulation of dimer in implicit solvent took about 8 days on the Biowulf cluster).

Limited Proteolysis Experiments. The recombinant catalytic domain (d129−402) of bovine β4Gal-T1 was used in this study. β4Gal-T1 (200 μg) in 1 mL was partially digested with 10 units of Glu-C protease (Sigma Chemicals) without any buffer or salt for 1 h. The partially digested protein was applied to a GlcNAc− or UDP−agarose column (Sigma Chemicals).

RESULTS AND DISCUSSION

Quantitative Comparison of Crystal Structures. When conf1 and conf2 structures are superimposed on the basis of their C^{α} positions, a root-mean-square deviation (RMSD) of 3.4 Å is obtained (Figure 1). A plot of residue-wise backbone RMSD shows that there are only two regions with a significant structural change: (i) the long loop (I345−H365) and (ii) the short Trp loop (Y311−G316; Figure 1 inset). The C^{α} RMSDs for the long and the Trp loop segments are 11.5 and 3.2 Å, respectively. The C^{α} of K351 and K352 in the long loop move by as much as 20 Å. When both loop segments are excluded, an RMSD of 1.2 Å is obtained, indicating that the rest of the structure does not undergo large structural changes. Further, part of the long loop (P357−R362) in conf1 adopts a helical conformation in conf2. A slightly different scenario is noticed for the short Trp loop, where an approximately 180° flip of the peptide unit linking W312 and G313 takes place in addition to changes in the backbone dihedral angles for the other neighboring residues (ψ of Y311 and W314, ϕ of W312). Thus, the transition between conf1 and conf2 cannot be brought about by changes in a single residue on either side of the loops unlike in many known cases of loop movements (for example, triosphosphate isomerase).

The refined β4Gal-T1 conf1 crystal structure shows a single predominant conformation for the long loop with high thermal parameters (*B*-factor). The loop has low thermal parameters in the conf2 complex with α-lactalbumin. Further, H347 is coordinated to the Mn^{2+} ion, but no hydrogen bond exists between the long loop and UDP-gal. There are six hydrogen bonds between the long loop and the rest of the protein in conf1, whereas in conf2 there are only three. However, several internal hydrogen bonds occur within the long loop in conf2 (two in conf1 versus ten in conf2). This suggests that the long loop in conf2 is stabilized by intraloop hydrogen bonds and contacts with the metal ion.

Comparison of Sequences of the β4Gal-T1 Family. We further compared sequences of β4Gal-T1 from different sources in order to examine the conservation of residues in the hinge and loop regions. A total of 37 sequences belonging to the β4Gal-T1 family were obtained from the PFAM database (*35*). The percentage of sequence identity with respect to the bovine sequence for which the crystal structure is available varied from 33% to 90%. On the basis of an identity matrix, a total of 25 residues were found to be conserved in all sequences (100%; Figure 2). As expected, many of the conserved positions were located near the functional site. For the long loop, the residues in neither the hinge region nor the loop were conserved significantly apart from H347. Interestingly, the Trp loop showed very high sequence conservation. In particular, the glycines flanking W314 were highly conserved [G313 (97% conserved), G315 (100%), and G316 (84%)]. Although this loop is surface exposed in conf1, it shows very high sequence conservation compared to the rest of the protein in the β4Gal-T1 family.

Identification of Flexible Regions. We first carried out three 10 ns simulations at 300 K starting with the conf1 crystal structure and using the EEF1 implicit solvent model force field (Table 1, simulation 1). Conformers were saved every 2 ps during the 10 ns production runs. We calculated the per residue average RMSD for backbone atoms from the superimposed conformers. On the basis of the plot of mean RMSD versus residue number we identified three regions, apart from the N-terminus, to be highly flexible: (i) D278−Y286, (ii) Y311−G316, and (iii) H347−E368 (Figure 3). The second loop (Y311−G316) is the Trp loop, which shows a conformational change when the crystal structures (conf1 and conf2) are compared. The third loop (H347−E368) is essentially the same as the long loop with an extension of three additional C-terminal residues (T366, K367, and E368). The first loop (D278−Y286) does not show any conformational change when conf1 and conf2 crystal structures are compared. However, this loop is found to be highly flexible in all three MD simulation runs. This loop forms contacts with the flexible third loop in conf1 and is hydrogen bonded to α-lactalbumin in the conf2 crystal structure. It is very likely that the loop is flexible in solution in the absence of any intermolecular interaction.

During all three 10 ns simulations, the third loop (or the long loop) moves by as much as 11 Å toward its conf2 position. The loop begins to move within a few hundreds of picoseconds in a unidirectional manner and reaches a stable conformation at the end of 2 ns (Figure 4). During the remaining 8 ns simulation the loop does not move any further toward conf2. Interestingly, the second loop (Trp loop) moves in the opposite direction to that of the third loop. Residues within the loop, W314 as well as the adjacent glycines (G313, G315, and G316), make several contacts with side chains positioned in the third loop (P357, Q358, R359, and F360; Figure 5a). W314 makes hydrophobic as well as hydrogen-bonding interactions with the third loop residues. Some of the hydrogen-bonding interactions are between the (i) W314 side chain indole group and Q358 side chain carboxamide group and (ii) W314 backbone carboxamide group and R362 side chain guanidino group (Figure 5b). This raises the possibility that these contacts are responsible for the large conformational change of the third loop. To identify the hinge residues, we examined the backbone dihedral angle changes for the loop residues along the trajectory. However, no significant dihedral angle change

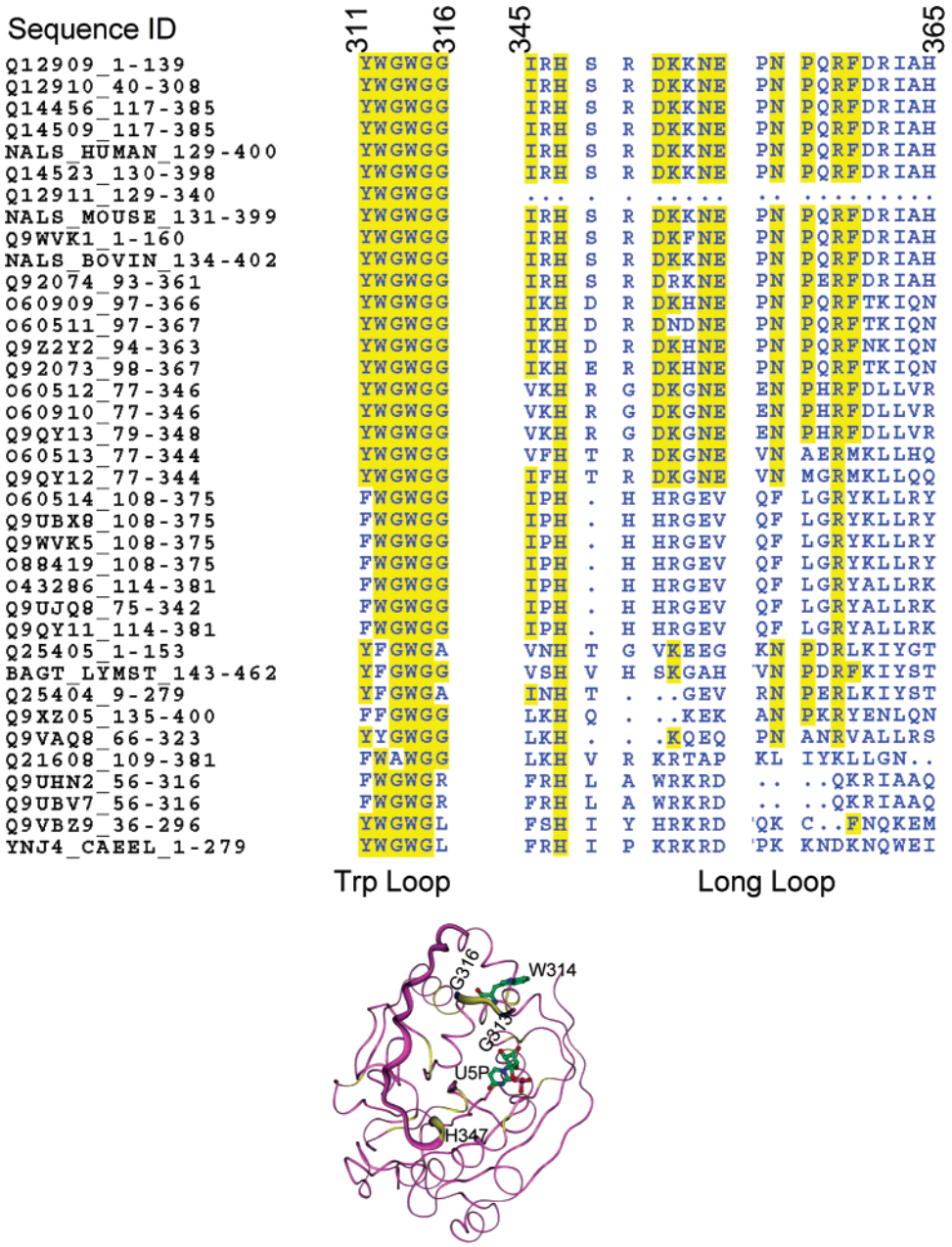

FIGURE 2: Sequence conservation for the Trp and long loops from the alignment of 37 known β4Gal-T1 sequences. The long loop has high sequence variation compared to the Trp loop. In particular, the glycines flanking W314 are highly conserved [G313 (97% conserved), G315 (100%), and G316(84%)]. The positions that are conserved, based on identity in more than 50% of the sequences, are marked in yellow. Only the nondeletion positions with respect to the bovine sequence (NALS_BOVIN) are shown here. The aligned sequences were obtained from the PFAM database (*35*). It may be noted here that this data set is not unique; the sequence from the same organism may be represented more than once but with a different identifier. A ribbon representation of the β4Gal-T1 conf1 crystal structure is shown at the bottom of the alignment. The fully conserved positions (100%) are marked in yellow. The G313 and G316 positions are shown in blue. The ligand uridine 5′-monophosphate (U5P) is shown in ball-and-stick model.

takes place for any single residue. Nevertheless, H347, S348, R362, and I363 show somewhat larger variations in their backbone dihedral angles. Starting with the conf1 crystal structure, we observed flexibility in the regions that show conformational variations when the conf1 and conf2 crystal structures are compared. Nevertheless, we did not observe the second and third loops to reach their equivalent positions as in conf2. Hence, complete transition from conf1 to conf2 is not observed during the simulations at 300 K, indicating a possible presence of an energy barrier for the transition between the two conformational states. Alternatively, the absence of complete conf1 to conf2 transition could be due to the lack of inclusion of explicit solvent molecules in our simulations.

We carried out two MD simulations at 300 K in explicit water starting with the conf1 crystal structure (Table 1, simulation 14). Analysis of the 3 ns trajectories revealed flexibility in the three loops. One of the simulation runs revealed a larger movement of the third loop toward its conf2 position (by as much as 14 Å; Figure 6a). The superposition of the starting and average conformation resulted in a RMSD of 2.9 Å for the C^{α} atoms. As can be seen from Figures 4 and 6a, the observed movement is consistent with that of the implicit simulation. To see the effect of a longer

Table 1: Molecular Dynamics Simulations on β4Gal-T1

simulation no.	starting structure	temp (K)/ time (ns)	MD runs	purpose	observations
			Using Implicit Solvent Model (EEF1)		
1	conf1	300/10	3	to identify flexible regions	three loop regions identified to be flexible; long loop moves toward its conf2 position
2	conf2	300/10	3	to identify flexible regions	three loop regions identified to be flexible
3	conf1/conf2	325, 350, 375, 400/2	2 (for each)	to examine flexibility and transition at higher temperature	conf1 and conf2 merge at 400 K
4	conformer at end of equilibration run of conf1 at 400 K	300/2	3	to examine communication of flexibility between the long and Trp loops	long loop does not move toward conf2, unlike simulation 1
5	conformer from equilibration run of conf1 at 400 K	300/2	1	to eliminate artifact in simulation 4 due to starting structure	long loop in this case also does not show large movement
6	modeled; Trp loop conformation in conf1 changed to its conf2	300/2	2	to eliminate artifact in simulations 4 and 5 due to starting structure	similar observation with some movement in the N-terminus of the long loop
7	conf1, with Trp loop restrained	300/2	2	to confirm communication of flexibility between the long and Trp loops	the long loop does not move toward its conf2 position
8	conf1 mutant, residues G313−G316 mutated to alanine	300/4	1	to examine if reduction of flexibility with the Trp loop affects the long loop	the long loop does not move, stable in its starting position
9	same as in simulation 4	400/2	1	to see if W314 flips out of the binding cavity at higher temperature	the Trp loop (with W314) flips out
10	conf1, with the long loop restrained	300/2	1	to examine if the Trp loop flexibility is independent of the long loop	the Trp loop is flexible
11	conf1 truncated (K352−S402 cleaved off)	300/10	1	to examine if the Trp loop flexibility is independent of the long loop	the Trp loop is flexible
12	conf1 dimer	300/10	2	to examine the effect of subunit−subunit interactions on the long loop	the long loop is flexible but stabilized closer to its starting position
13	conf2 dimer	300/10	1	to examine the effect of subunit−subunit interactions on the long loop	no effect on the flexibility, since the quaternary structure is different from conf1
			In Explicit Solvent with Periodic Boundary Condition		
14	conf1	300/3	2	to identify flexible regions and confirm the observation from simulation 1	the long loop moves toward conf2; all three loops are flexible
15	conf1	400/2	2	to check flexibility and the long loop movement at higher temperature	the long loop moves toward conf2
16	conf2	300/3	1	to examine flexibility and unfolding of the helical conformation of the long loop	the long loop is flexible, but the helix is stable
17	conf2	400/2	2	to examine the unfolding of the helical conformation of the long loop	the helix unfolds first, and the long loop begins to move toward its conf1 position
18	conf1 + UDP-gal	300/3	1	to examine the effect of ligand on the flexibility of the long loop	the long loop is flexible and moves toward its position in conf2
19	conf1 + UDP-gal	400/2	1	to examine the effect on the flexibility of the long loop at higher temperature	the long loop is flexible and moves toward its position in conf2
20	conf2 + UDP-gal	300/3	1	to examine the effect of ligand on the flexibility of the long loop	no effect on the flexibility; the helix is stable as in the case of simulation 16

simulation period, we increased the temperature to 400 K and carried out two 2 ns simulations in explicit water (Table 1, simulation 15). As we expected, at 400 K, the third loop moved by as much as 13 Å toward its conf2 position (Figure 6b). The second loop also showed a tendency to be buried under the third loop. These observations revealed that conf1 as in the crystal structure is not a stable conformation in the absence of quaternary interactions. The intrinsically flexible third loop is stabilized in an intermediate position between conf1 and conf2. To examine the effect of the ligand on the observed flexibility of the third loop, we carried out simulations starting with conf1 in the presence of UDP-gal. We modeled the ligand coordinates in conf1 on the basis of its position in conf2. The simulations were carried out in explicit water at 300 K as well as at 400 K starting with the ligand-bound form of conf1. The simulations revealed that the ligand had no significant effect on the flexibility of the third loop (Table 1, simulations 18 and 19). The third loop

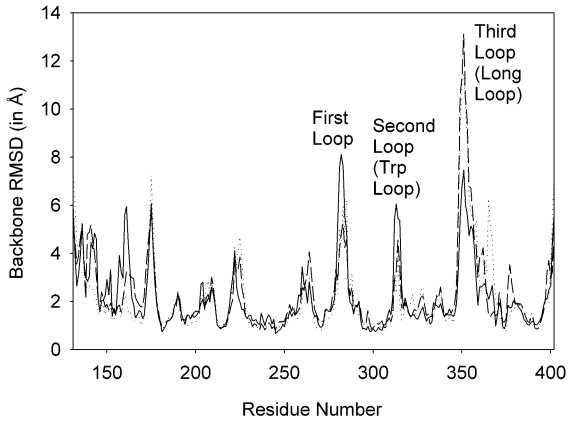

FIGURE 3: Plot of the average residue-wise root-mean-square deviation (RMSD) for the backbone atoms, calculated from the C^{α} superposed coordinates of conformers obtained during three 10 ns implicit simulations (represented as continuous, dashed, and dotted lines) at 300 K starting with conf1. Apart from the N-terminus, three internal regions of β4Gal-T1 show high flexibility. The second and third flexible loops are essentially the same as the Trp and long loops, respectively.

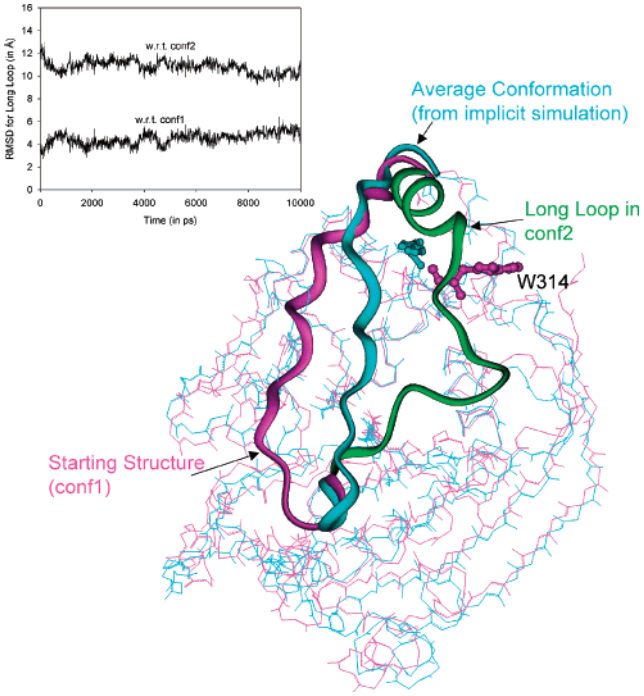

FIGURE 4: C^{α} superposition of the starting structure (conf1, purple) and the average conformation (cyan) obtained from a 10 ns implicit simulation at 300 K. The long loop (third loop) moves by as much as 11 Å toward its conformation in conf2. For comparison, the long loop conformation in conf2 is shown in ribbon representation (green). The Trp loop (second loop) moves in the opposite direction to that of the third loop. The inset shows the C^{α} root-mean-square deviation (RMSD) for the long loop alone calculated with respect to its conformation in conf1 and conf2, during the 10 ns simulation period. As can be seen, the RMSD for the loop increases with respect to conf1 whereas it decreases with respect to conf2, indicating the movement of the loop toward its conf2 conformation.

still moves toward its position in conf2, as in the case of the simulations without the ligand. Interestingly, during the simulation at 400 K, residue W314 flips inside the binding cavity and makes contact with the ligand (Figure 6c).

Implicit MD simulations starting with the conf2 crystal structure did not reveal a large conformational change, unlike in conf1 (Table 1, simulation 2). The third loop remained

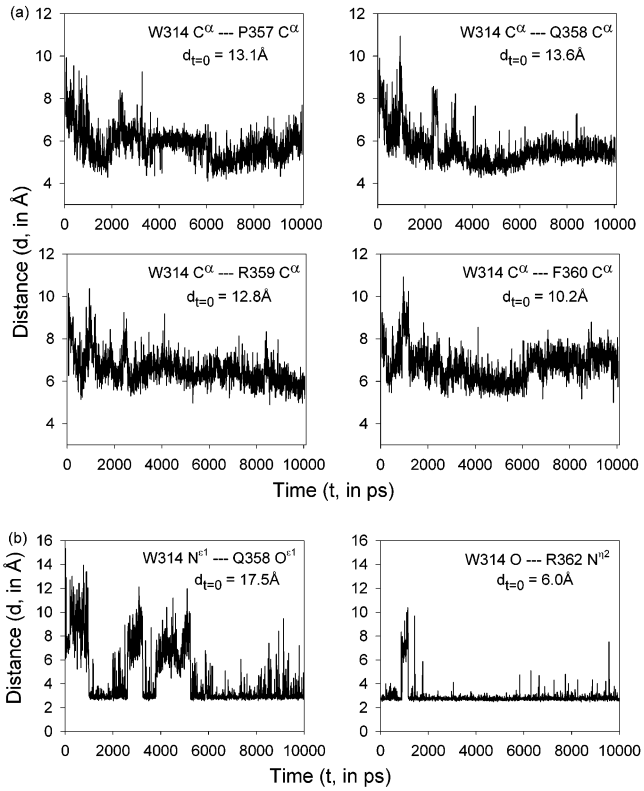

FIGURE 5: (a) C^{α}—C^{α} distance between W314 and residues positioned in the third loop. The trajectories show that the second loop moves closer to the third loop and makes contact (not all of the contacting residues are shown here). (b) Hydrogen-bonding interaction between W314 and Q358 + R362. The residue W314 is involved in hydrophobic as well as hydrogen-bonding interactions with the third loop residues, during the course of a 10 ns implicit simulation at 300 K starting with conf1.

stable in its starting conformation, and in particular, the second loop was much less flexible. However, the first loop showed a larger flexibility compared to the conf1 case. The first loop is in contact with the third loop in conf1 whereas they are well separated in the case of conf2. The absence of intramolecular contacts might have led to the observed larger flexibility. It appears that the third loop in conf2 is more stable than in conf1. This might be because part of the third loop is in a helical conformation in the conf2 crystal structure (ten intraloop hydrogen bonds in conf2 versus two in conf1). To verify the observations from the implicit simulations, we carried out MD simulations in explicit water starting with the conf2 crystal structure at 300 and at 400 K (Table 1, simulations 16 and 17). Indeed, a 2 ns MD run at 400 K in explicit water did reveal that the helical region of the third loop unfolds first (Figure 7). This indicated that the helical conformation of the third loop is relatively less stable and might undergo a transition to a loop structure in solution. However, we caution that the viscosity of water at this temperature is reduced and therefore the difference between the 300 and 400 K simulations may be related to factors other than the energy barrier for the transition. We also carried out simulation at 300 K starting with the UDP-gal-bound form of conf2 (Table 1, simulation 20), in which also the third loop was stabilized in its initial conformation. The observation that conf2 was relatively more stable compared to conf1 is consistent with the experimental fact that conf2 is the active form (*21*). The stable population of conf2 might

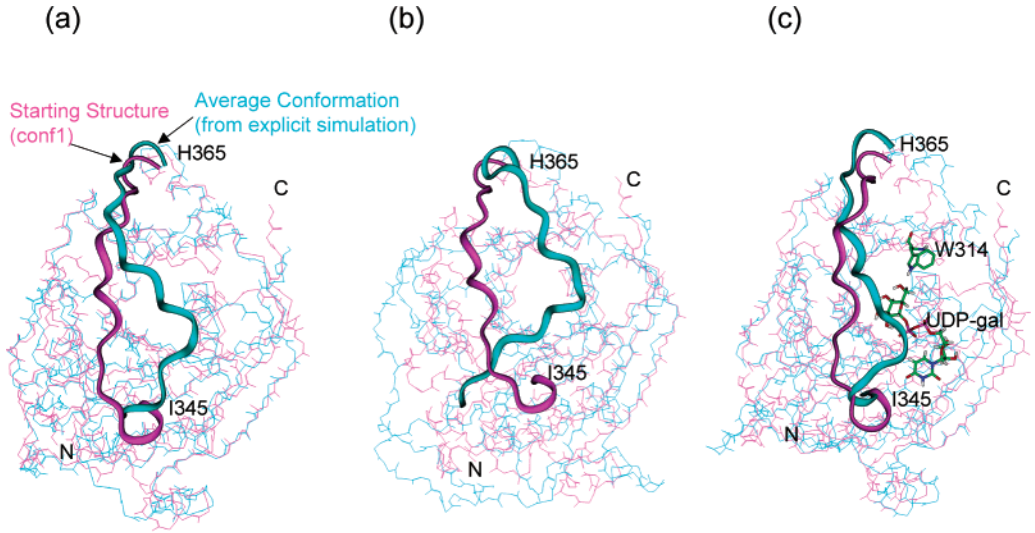

FIGURE 6: (a) C^{α} superposition of the starting structure (conf1, purple) and the average conformation (cyan) observed during a 3 ns simulation at 300 K in explicit water. As can be seen, consistent with the implicit simulation (Figure 4), the third loop has moved toward its conf2 position. (b) C^{α} superposition of the starting structure (conf1, purple) and a conformer at 2 ns (cyan) during a 2 ns simulation at 400 K in explicit water. A large structural change is seen in the third loop and as well as in the first 23 residues of the N-terminus. (c) C^{α} superposition of the starting structure (conf1 + UDP-gal, purple) and a conformer at 300 ps (cyan) during a 2 ns simulation at 400 K in explicit water. The ligand UDP-gal is shown in stick model. The residue W314 flips inside the binding pocket and makes contact with the ligand, and the third loop begins to move toward its conf2 position.

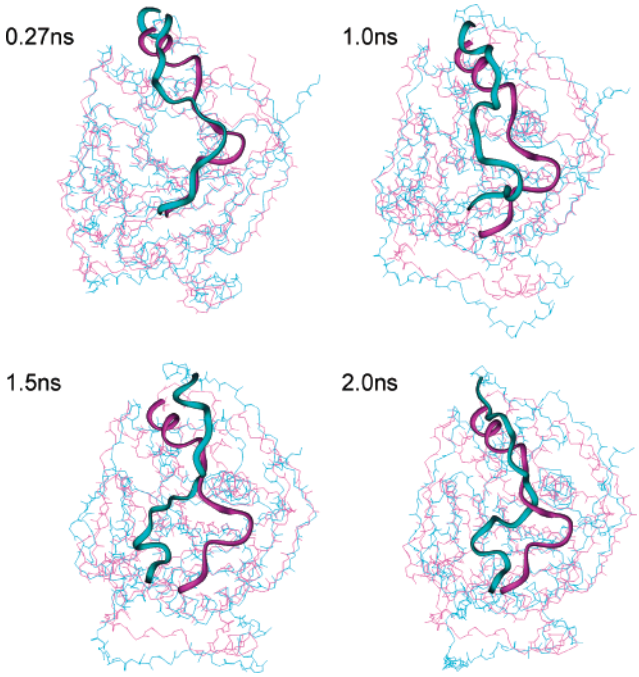

FIGURE 7: C^{α} superposition of the starting structure (conf2, purple) and conformers observed (cyan) during a 2 ns simulation at 400 K in explicit water. The helical conformation of the third loop region unfolds first, indicating its relatively lower stability.

be necessary so that interaction with α-lactalbumin could be made for lactose synthesis.

A comparison of the average potential energy observed during the explicit water simulations at 300 K revealed that conf2 (−77035.3 kcal/mol) has a lower energy as compared to conf1 (−77016.8 kcal/mol). This was also the case when explicit water simulations at 300 K were carried out on the ligand-bound form of conf1 and conf2 (conf1, −76986.0 kcal/mol; conf2, −77076.1 kcal/mol), since the UDP-gal interacts favorably with conf2. Also, implicit simulations at 300 K revealed equal energy for both conformations (conf1/conf2, −6224.6 kcal/mol; Table 2). We note, however, that the entropy contribution is not taken into account in the energy calculations. Nevertheless, this comparison further indicates that the stability of the third loop conformation as observed in the conf1 crystal may arise from the subunit−subunit interactions. To validate this proposition, we carried out 10 ns simulations at 300 K starting with conf1 and conf2 dimers (Table 1, simulations 12 and 13). The simulation starting with the conf1 dimer revealed that the third loop is stabilized closer to its initial conformation (Figure 8). It may be noted here that the third loop is in contact with the other subunit in the conf1 crystal structure. The subunit−subunit interactions also played a role in stabilizing the first loop to some extent. As expected, in the case of the conf2 dimer simulation, no effect on the third loop was observed since the quaternary arrangement is different from conf1.

Energy Barrier? MD Simulations at Higher Temperature. Since MD simulations starting with conf1 at 300 K did not reveal the complete transition from conf1 to conf2 and instead remained in an intermediate stable conformation, we suspected the presence of an energy barrier. MD simulations at higher temperature can be informative with regard to conformational changes that may occur on longer time scales and can overcome energy barriers. We carried out a total of sixteen 2 ns MD runs starting with conf1 and conf2 at temperatures 325, 350, 375, and 400 K (two runs for each; Table 1, simulation 3). When the trajectories were analyzed, as expected, we found that the flexibility in all three loops increased as we raised the temperature (Figure 9). The simulation at 400 K starting with conf1 showed much larger fluctuations in the third loop. Examination of the average conformation observed during the simulations showed that at this temperature conf1 essentially merges with conf2 but with significant structural differences in the two regions of the molecule (Figure 10). First, the helical conformation in part of the third loop in conf2 did not form when starting with conf1. This is understandable since helix formation is

Table 2: Average Energies from Simulations at 300 K on β4Gal-T1[a]

simulation no.[b]	starting structure	average energies (kcal/mol)[c] total	internal	external	solvation
1	conf1	−6224.6	2661.9	−8886.5	−2345.5
2	conf2	−6224.6	2650.7	−8875.3	−2328.7
4	conformer at end of equilibration run of conf1 at 400 K[d]	−6163.6	2660.8	−8824.4	−2353.7
5	conformer from equilibration run of conf1 at 400 K[d]	−6196.4	2667.8	−8864.1	−2354.1
6	modeled; Trp loop conformation in conf1 changed to its conf2[d]	−6189.4	2673.5	−8862.9	−2356.4

[a] The average energies were calculated only for unrestrained simulations at 300 K with the wild-type sequence. [b] The simulation number corresponds to that in Table 1. [c] The average energies were calculated from the last 2 ns simulation period for simulations 1 and 2 and from the last 0.5 ns period for simulations 4, 5, and 6. Total = internal + external energies; internal = bond + angle + dihedral energies; external = van der Waals + electrostatic + solvation energies. [d] In these cases, the starting structure had the long loop (third loop) in the conf1 conformation and the Trp loop (second loop) in the conf2 conformation. Subsequently, the total energies are higher (28−61 kcal/mol) compared to those starting with the crystal structures (conf1 and conf2).

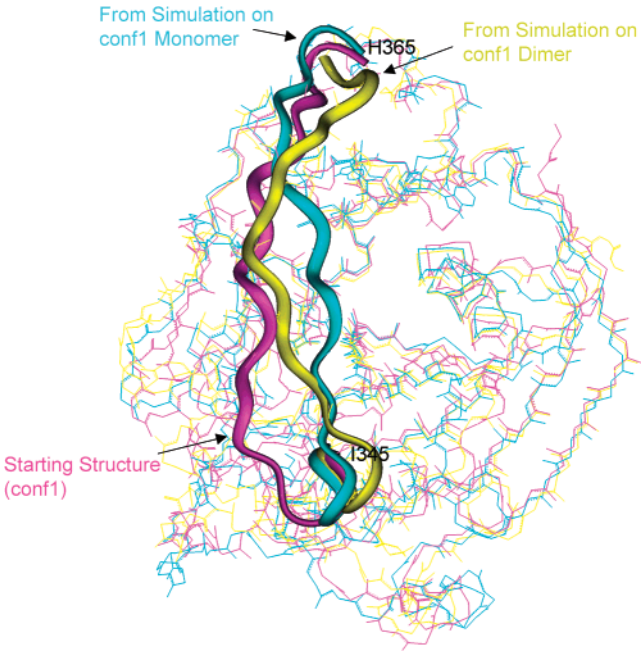

FIGURE 8: C^{α} superposition of the starting structure (purple) and the average conformations observed during the 10 ns simulations in implicit solvent starting with conf1 monomer (cyan) and dimer (yellow; only one subunit is shown). The third loop is stabilized closer to its initial conformation due to the subunit−subunit interactions, during the simulation on conf1 dimer.

not expected during the simulations at 400 K. The second major difference is in the first loop, which moves toward the binding pocket to a larger extent in conf2 as compared to conf1. The absence of contacts between the first and third loops in conf2 might have lead to this larger movement. Apart from these two differences, the conf1 structure essentially merges with conf2 when the rest of the structures are compared. More significantly, during the 400 K simulations starting with conf1, the second loop which was surface exposed has flipped into the binding cavity as in conf2. This flipping of the second loop is seen during the equilibration run (within 100 ps) when the third loop had just begun to move toward conf2. This observation is in agreement with an earlier suggestion based on structural studies that, after the ligand enters, W314 must flip inside the binding pocket first so that the third loop can be placed over the second loop (*21*). Thus, the simulation successfully established the sequence of events that may take place during ligand binding. Further, once the second loop flips inside the binding cavity as in conf2, it is much less flexible than when it is outside surface exposed as in conf1. A similar observation was also made with the simulations at 400 K starting with conf2. The loss of flexibility of the second loop when it is buried inside the binding pocket might be due to the hydrophobic nature of the W314.

In summary, the simulations at 400 K represent the movement of the third loop from conf1 to conf2 better than the simulations at 300 K. This observation together with the W314 flip only at higher temperature (400 K) suggests that there may be an energy barrier associated with the transition. Formation of the helix in the third loop (as in conf2) is not observed during the conf1 400 K simulations. The transition from conf2 to conf1 does not take place when starting with conf2, even at 400 K. This could be since part of the third loop is in helical conformation and the unfolding of this helical structure may require explicit water molecules, as demonstrated by the simulations at 400 K in explicit water starting with conf2.

The Role of the Second Loop: Communication of Flexibility. Could the energy barrier arise due to the position of W314 since it must flip inside the binding pocket first so that the third loop could move all the way from its conf1 to the conf2 position? The second loop (Trp loop; Y311−G316) is surface exposed and away from the third loop (long loop; I345−H365) in conf1, whereas in conf2 the second loop is buried under the third loop (P355 C^{α}---W314 C^{α} distance is 16.8 and 4.6 Å in conf1 and conf2, respectively). The accessible surface area (ASA) calculation, using the Lee and Richards method (*36*), shows that the largest change is at the tip of the second loop [conf1, G313 (53% exposed), W314 (63%), G314 (32%); conf2, G313 (8%), W314 (45%), G314 (6%)]. During the conf1 simulations at 300 K, the second loop moves toward the third loop, and W314 makes hydrophobic contacts and is involved in hydrogen bonding with the third loop side chains. To avoid such a situation we selected a conformer (at the end of 50 ps equilibration period) from a conf1 400 K simulation where the second loop had flipped inside the binding pocket, like in conf2, but the third loop was essentially in conf1. In this conformation, W314 is less likely to make many contacts as it could when it was outside (P355 C^{α}---W314 C^{α} distance is 19.3 Å), unless it moves out of the binding pocket. We carried out three 2 ns simulations at 300 K starting with this hypothetical initial structure (third loop in conf1, second loop in conf2; Table 1, simulation 4). We expected that during the simulations the third loop would move to conf2 quickly and more easily now that the second loop was out of the

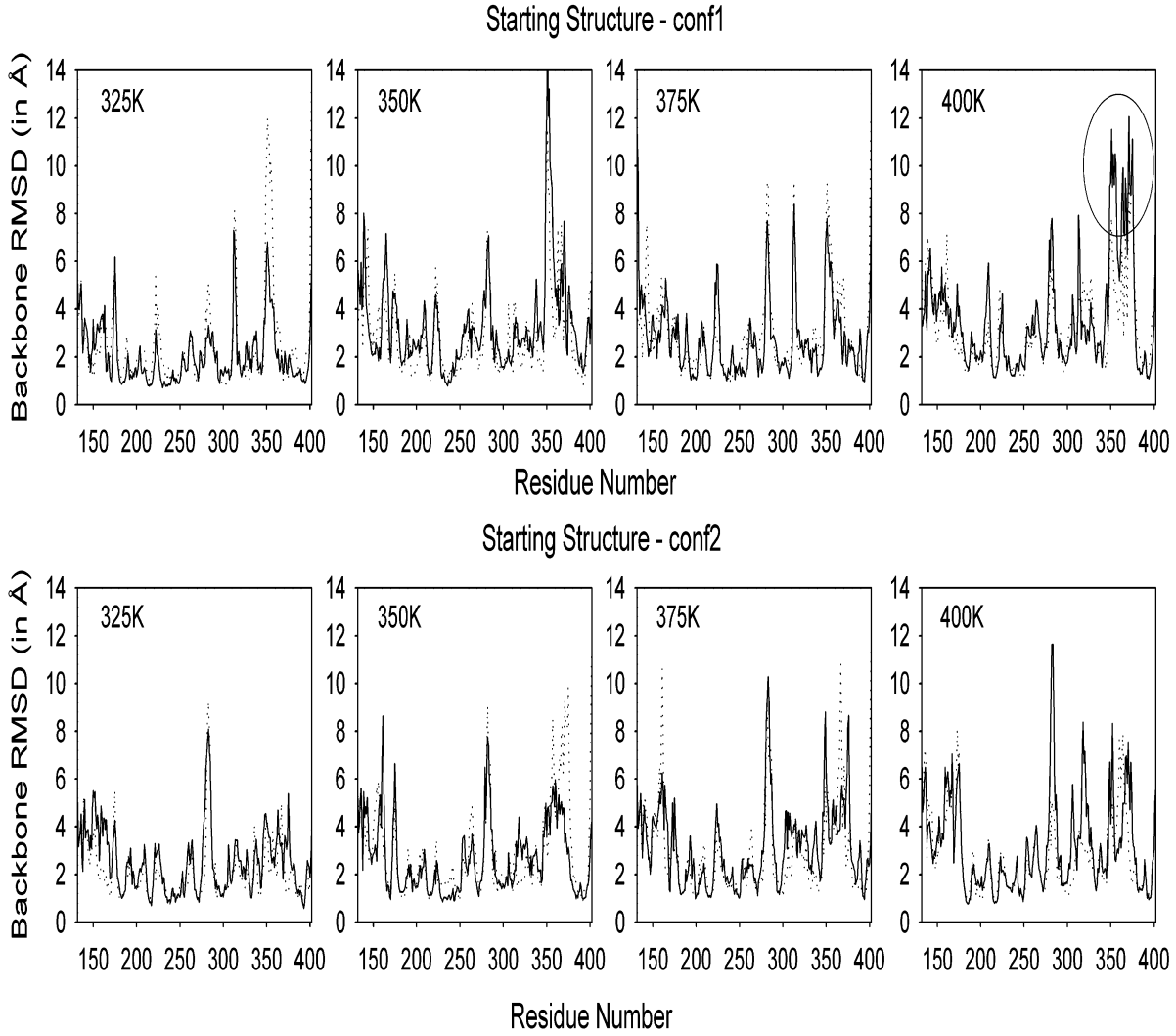

FIGURE 9: Plot of the average residue-wise root-mean-square deviation (RMSD) for the backbone atoms obtained from simulations starting with conf1 and conf2 at various temperatures (two MD runs for each, represented as continuous and dotted lines). A large conformational change takes place in the third loop at 400 K starting with conf1.

way. Contrary to our expectation, the third loop remained close to its starting position although it was flexible (Figure 11a,b). But, this was not the case when the second loop was outside, as in conf1 in which case the third loop moved by as much as 11 Å toward its conf2 position. This observation indicated that the flexibility of the smaller second loop (6 residues) is communicated to the longer third loop (21 residues). Further, another 2 ns simulation starting with a conformer selected from the equilibration run (just after W314 flipped inside) also led to a similar observation (Table 1, simulation 5). To further rule out an artifact with our starting structure, we modeled the second loop of conf2 into the conf1 crystal structure and carried out two 2 ns simulations at 300 K (Table 1, simulation 6). This also led to a similar observation with the exception of the N-terminal part of the third loop (I345−P355) moving toward conf2. Examination of the contacts showed that, when modeled, W314 is closer to the third loop than in our hypothetical structures adopted from the simulation (P355 C^{α}---W314 C^{α} distance is 11.2 Å).

To further examine the hypothesis that flexibility in the second loop is communicated to the third loop, we restrained the second loop in its conf1 position by applying a restraint force [30 kcal/(mol Å^2)], knocking out its flexibility, and carried out two 2 ns simulations (Table 1, simulation 7). In both MD runs, the third loop was stabilized in its starting conformation (conf1) and did not move toward conf2 (Figure 11c). These observations suggest that the second loop acts as a lubricant for the third loop so that its flexibility is increased, facilitating the transition from conf1 to conf2. The mobility imparted to the third loop enables it to cross the energy barrier. In the absence of the contacts between the second and third loops, the energy barrier for the transition from conf1 to conf2 and vice versa would be higher (Figure 12). On the basis of simulations 4, 5, and 6, an approximate estimate of the increase in the energy barrier would be between 28 and 61 kcal/mol (Table 2). Recently, we have been able to crystallize the W314A mutant in the absence of any ligands (R. Velavan, B. Ramakrishnan, and P. K. Qasba, unpublished results). Earlier, despite many attempts neither the wild type (S130−S402) nor any other mutants that were made could be crystallized without UDP-gal or GlcNAc/α-lactalbumin. This provides experimental support for our proposition. The W314A mutant also significantly lost its catalytic activity when compared to the wild type. When W314 is mutated to Ala, it is very likely that there would be a significant loss of contacts between the second and third loops.

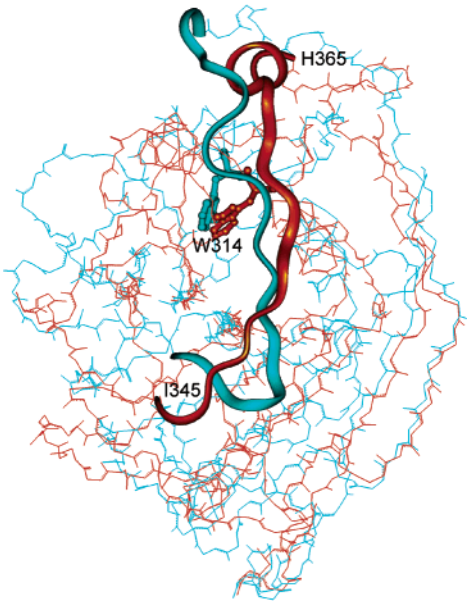

FIGURE 10: C^{α} superposition of the average conformations obtained during the 2 ns implicit simulations at 400 K starting with conf1 (cyan) and conf2 (brown). The structures essentially merge at this temperature. In particular, residue W314 in conf1 is buried under the third loop, as in the case of conf2 (Figure 1).

This hypothesis is further supported by the fact that W314 is flanked by glycines (one preceding and two succeeding positions; G313-W-G-G316). Examination of 37 sequences belonging to the β4Gal-T1 family showed that these glycines are highly conserved (Figure 2). Further, our examination showed that when these glycines are replaced by alanines (using InsightII), the C^{β} atoms of the alanines pointed outward toward the surface. This observation indicates that the presence of alanines would not have severely affected the size of the binding pocket. Another 2 ns simulation starting with conf1 with only the third loop restrained confirmed that the second loop was highly flexible (Table 1, simulation 10). Therefore, the flexibility of the second loop is independent of the third loop. A rotary motion involving residues G313 and W314 for the second loop was suggested previously (*21*).

The role of glycines in flexibility and function has already been observed in many other known cases. The HIV protease flap region, where a hydrophobic residue (isoleucine) is flanked by highly conserved glycines (two preceding and two succeeding) in the turn region of a β-hairpin, provides another such glycine example. The flap region acts as a gate to the ligand binding site. As in the case of β4Gal-T1, the flap region undergoes a large conformational change (peptide bond flip) (*7*). This suggests evolutionary pressure for flexibility that is required for function. Recently, the role of conserved glycines in the catalytic loop of fructose-1,6-bisphosphate aldolase has been explored through site-directed mutagenesis (where each glycine in the loop has been mutated to alanine) and kinetic studies (*37*). It was observed that the replacement of glycine residues has only marginal effects on substrate recognition and binding but significantly affects catalysis, indicating the importance of flexibility for function. Thus, in the β4Gal-T1, HIV protease, and aldolase cases, clearly there is a correlation between conservation, flexibility, and function.

Flexible Loops: Comparison with Experiments. Although two of the three flexible loops identified by MD simulations are found to undergo conformational change in the crystal structures, further experiments are needed to confirm the flexibility of the loops. X-ray crystallography provides only a snapshot of the molecular conformation in solution. It has been shown that sites of limited proteolysis in globular proteins are characterized by chain flexibility devoid of regular secondary structure (*38*, *39*). Using limited proteolysis, the structural and dynamic features of folded and partly folded proteins have been investigated (*40*, *41*). This approach relies on the fact that unfolded proteins are degraded much faster than native rigid protein structures. In native proteins, limited proteolysis occurs at sites of enhanced segmental mobility or in flexible loops. To experimentally verify the observations from the simulations, we carried out limited proteolysis experiments using the Glu-C protease

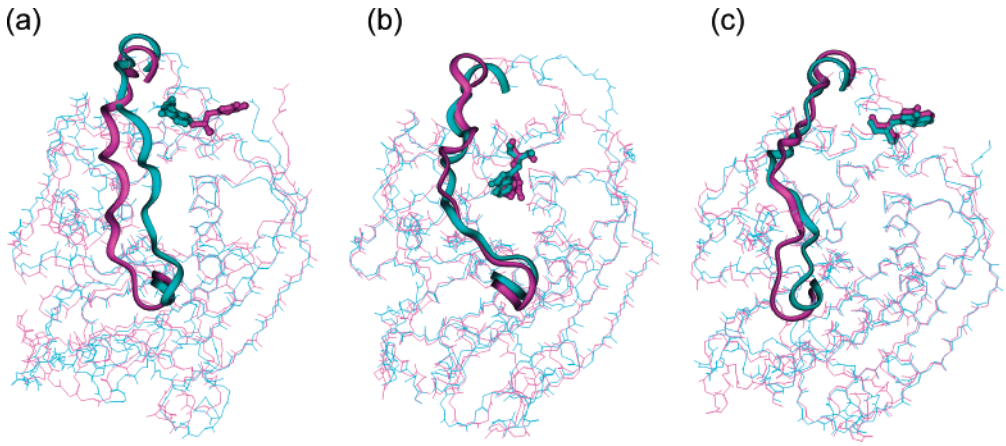

FIGURE 11: A comparison of three cases that shows the second loop position (with W314 shown in ball-and-stick model) affecting the movement of the third loop in 2 ns implicit simulations at 300 K. (a) Superposition of the starting structure conf1 (purple) and the average conformation (cyan) obtained during the simulation. (b) Superposition of a hypothetical starting structure (purple; conformer at the end of the equilibration run of conf1 at 400 K), in which the second loop is in conf2 and the third loop is in conf1 conformation, and the average structure (cyan) obtained during the simulation. (c) Superposition of the starting structure conf1 (purple) and the average conformation (cyan) obtained during the simulation, in which a restraint force [30 kcal/(mol $Å^2$)] was applied to the second loop. The third loop is retained near its starting conformation in the (b) and (c) cases, unlike in (a). These observations suggest communication of flexibility between the second and third loops.

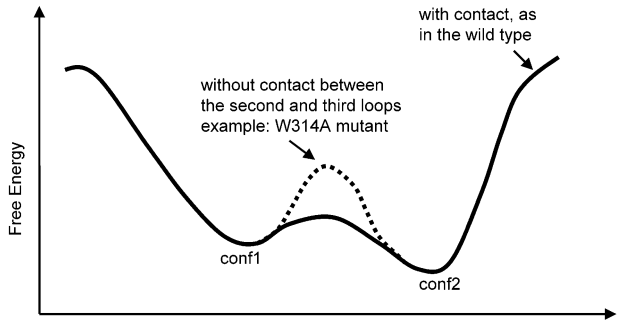

FIGURE 12: Schematic energy diagram indicating the energy barrier arising from the absence of contacts between the second and third loops. For example, in the W314A mutant, there would be significant loss of contacts between the second and third loops, which increases the energy barrier for the transition between conf1 and conf2.

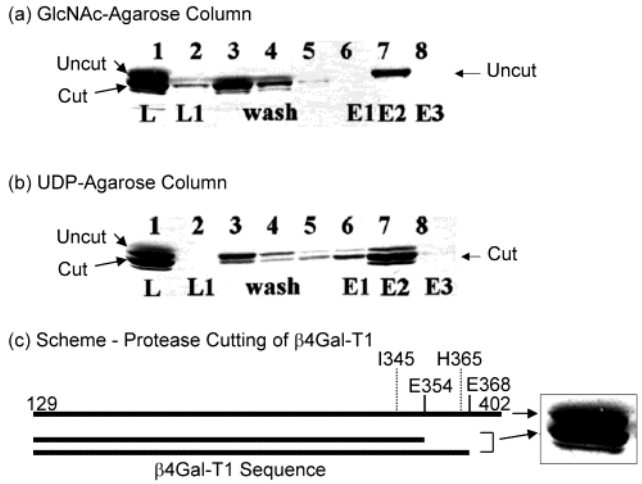

FIGURE 13: Experimental confirmation of flexibility through limited proteolysis experiments. Partially digested β4Gal-T1 with Glu-C protease bound to (a) GlcNAc−agarose and (b) UDP−agarose columns. SDS gel analysis of a 20 μL sample of each column fraction: lane 1, sample prior to the loading (L); lane 2, pass-through after loading on the column (L1); lanes 3−5, washing of the column with the load buffer (wash); lanes 6−8, elution of the bound protein from the column (E1, E2, E3). The upper band corresponds to the uncut protein (129−402), the lower band (second major band) corresponds to the C-terminal cut protein (129−368), and the lower most (third minor band) corresponds to further cut protein (129−354). (c) Schematic representation of the experimental observations.

(Figure 13). Limited digestion of Gal-T1 results in a relatively small polypeptide released from the C-terminus. This is confirmed by the N-terminal sequencing of the cleaved protein. From the size of the cleaved protein, it appears that Glu-C with E354 or E368 being the potential cleavage sites has cut off 30−40 amino acids from β4Gal-T1. Both residues are in the third loop (I345−H365; or long loop), shown to be flexible in the simulations. Further, binding studies indicate that the cleaved protein folds into a compact tertiary structure even in the absence of the C-terminus that includes the third loop. Experimental studies indicate that the truncated β4Gal-T1 (without the third loop) is able to bind UDP-gal but not GlcNAc, since residues important for UDP-gal binding were not cleaved off. This is further supported by structural data on the wild type that the UDP-gal binding site is compact and is independent of the third loop. It is likely that the W314 interacts with the UDP-gal as in the second crystal structure of β4Gal-T1 (conf2), even without the third loop. A difference spectroscopic experiment on the β4Gal-T1 has shown that a tryptophan must interact directly with the ligand molecule and it must undergo conformational change when the ligand binds (*24*). Our simulations indicate that the flexibility of the second loop is independent of the third loop (Table 1, simulations 10 and 11).

CONCLUSIONS

Enzyme activity relates to conformational flexibility (*3*). Loops frequently position the catalytically important residues at the functional site. Consequently, it is essential to explore the conformational changes observed between the bound and unbound crystal structures of an enzyme to understand whether the loops involved are intrinsically flexible, the extent of their flexibility, and the residues responsible for the flexibility. Comparison of the crystal structures of β4Gal-T1 crystallized with (conf2) and without (conf1) the substrates (UDP-gal or higher concentration of GlcNAc with α-lactalbumin) reveals a large conformational change taking place in a long loop (I345−H365) and in a smaller loop containing a tryptophan residue flanked by glycines (Y311−G316; Trp loop).

Here we have explored the intrinsic flexibility of the loops involved in the function of β4Gal-T1 and identified the residues responsible for the observed flexibility. Our MD simulations demonstrate the flexibility of the two loops, the long loop and the short Trp loop. Limited proteolysis experiments further confirm the intrinsic flexibility of the long loop. Examination of the average conformation observed during the implicit as well as explicit simulations in water at 300 K starting with conf1 shows that the long loop moves toward its conf2 position, stabilizing in an intermediate conformation. The Trp loop moves toward the long loop, making contacts with several residues positioned in the long loop. This observation indicates *communication of flexibility* between the two loops. Further, changing the spatial positioning of the Trp loop or restraining the Trp loop or mutating the glycines in the Trp loop all had a dramatic effect on the movement of the long loop.

Analysis of the sequences belonging to the β4Gal-T1 family shows that the glycines flanking W314 in the Trp loop are highly conserved. However, the long loop that undergoes a large conformational change exhibits very low sequence conservation. Our results suggest that the glycines are conserved for the flexibility of the Trp loop that is *communicated* to the long loop. In the absence of contacts between these two loops, there is a higher energy barrier between the two conformational states. This proposition is further supported by the fact that we are able to crystallize the W314A mutant in conf1 conformation, whereas the wild type could not be crystallized without UDP-gal or GlcNAc/α-lactalbumin (R. Velavan, B. Ramakrishnan, and P. K. Qasba, unpublished results). Thus, in the case of β4Gal-T1 there appears to be an evolutionary pressure for flexibility that is required for function. The role of glycines in flexibility and function has already been observed in many other known cases including fructose-1,6-bisphosphate aldolase and HIV protease. We are currently carrying out further analyses of the sequence and structural databases, followed by long-time molecular dynamics simulations, in order to explore how

common this mechanism is in nature, a mechanism which involves communication of flexibility between the loops and the conservation of glycines for flexibility. Remarkably, we have found that this type of mechanism is conserved across different folds and proteins performing different functions. The examples include enolase and lipase. Hence, β4Gal-T1 serves as a model system for a mechanism that is effectively and easily made use of by other proteins to perform a variety of functions.

ACKNOWLEDGMENT

We thank Drs. Sandeep Kumar, Chung-jung Tsai, David Zanuy, Velavan Ramasamy, Elizabeth Boeggeman, and in particular Jacob V. Maizel for numerous helpful discussions. We also thank the NCI-Frederick Advanced Biomedical Computing Center for time and assistance. This study utilized the high-performance computational capabilities of the Biowulf/LoBoS3 cluster at the National Institutes of Health, Bethesda, MD.

REFERENCES

1. Bennett, W. S., and Huber, R. (1984) Structural and functional aspects of domain motions in proteins, *CRC Crit. Rev. Biochem. 15*, 291−384.
2. Joseph, D., Petsko, G. A., and Karplus, M. (1990) Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop, *Science 249*, 1425−1428.
3. Yon, J. M., Perahia, D., and Ghelis, C. (1998) Conformational dynamics and enzyme activity, *Biochimie 80*, 33−42.
4. Todd, A. E., Orengo, C. A., and Thornton, J. M. (1999) Evolution of protein function, from a structural perspective, *Curr. Opin. Chem. Biol. 3*, 548−556.
5. Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000) Folding and binding cascades: dynamic landscapes and population shifts, *Protein Sci. 9*, 10−19.
6. Tsai, C. J., Ma, B., Sham, Y. Y., Kumar, S., and Nussinov, R. (2001) Structured disorder and conformational selection, *Proteins: Struct., Funct., Genet. 44*, 418−427.
7. Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Lam, P. Y., Jadhav, P. K., Hodge, C. N., Domaille, P. J., and Chang, C.-H. (1995) Flexibility and function in HIV-1 protease, *Nat. Struct. Biol. 2*, 274−280.
8. Shewmaker, F., Maskos, K., Simmerling, C., and Landry, S. J. (2001) The disordered mobile loop of GroES folds into a defined β-hairpin upon binding GroEL, *J. Biol. Chem. 276*, 31257−31264.
9. Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001) Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation, *Cell 105*, 115−126.
10. Wiegand, G., and Remington, S. J. (1986) Citrate synthase: structure, control, and mechanism, *Annu. Rev. Biophys. Biophys. Chem. 15*, 97−117.
11. Wand, A. J. (2001) Dynamic activation of protein function: A view emerging from NMR spectroscopy, *Nat. Struct. Biol. 8*, 926−931.
12. Doniach, S., and Eastman, P. (1999) Protein dynamics simulations from nanoseconds to microseconds, *Curr. Opin. Struct. Biol. 9*, 157−163.
13. Fersht, A. R., and Daggett, V. (2002) Protein folding and unfolding at atomic resolution, *Cell 108*, 573−582.
14. Beyer, T. A., and Hill, R. L. (1968) Glycosylation pathway in the biosynthesis of nonreducing terminal sequences in oligosaccharides of glycoproteins, in *The Glycoconjugates* (Horowitz, M., Ed.) Vol. III, pp 25−45, Academic Press, New York.
15. Brew, K., Vanaman, T. C., and Hill, R. L. (1968) The role of α-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction, *Proc. Natl. Acad. Sci. U.S.A. 59*, 491−497.
16. Lo, N. W., Shaper, J. H., Pevsner, J., and Shaper, N. L. (1998) The expanding β4-galactosyltransferase gene family: messages from the databanks, *Glycobiology 8*, 517−526.
17. Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities, *Biochem. J. 326*, 929−939.
18. Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat B. (1998) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities, *Biochem. J. 329*, 719.
19. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J. M. (2000) X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily, *EMBO J. 19*, 5269−5280.
20. Gastinel, L. N., Cambillau, C., and Bourne, Y. (1999) Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose, *EMBO J. 18*, 3546−3557.
21. Ramakrishnan, B., and Qasba, P. K. (2001) Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β1,4-galactosyltransferase-I, *J. Mol. Biol. 310*, 205−218.
22. Watt, G. M., Lowden, P. A., and Flitsch, S. L. (1997) Enzyme-catalyzed formation of glycosidic linkages, *Curr. Opin. Struct. Biol. 7*, 652−660.
23. Ramakrishnan, B., Balaji, P. V., and Qasba, P. K. (2002) Crystal structure of β1,4-galactosyltransferase complex with UDP-gal reveals an oligosaccharide acceptor binding site, *J. Mol. Biol. 318*, 491−502.
24. Takase, K., and Ebner, K. E. (1981) Interactions of substrates and α-lactalbumin with galactosyltransferase as measured by difference spectroscopy, *J. Biol. Chem. 256*, 7269−7276.
25. Brew, K., Shaper, J. H., Olsen, K. W., Trayer, I. P., and Hill, R. L. (1975) Cross-linking of the components of lactose synthetase with dimethylpimelimidate, *J. Biol. Chem. 250*, 1434−1444.
26. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy minimization and dynamic calculations, *J. Comput. Chem. 4*, 187−217.
27. Lazaridis, T., and Karplus, M. (1999) Effective energy function for proteins in solution, *Proteins: Struct., Funct., Genet. 35*, 133−152.
28. Inuzuka, Y., and Lazaridis, T. (2000) On the unfolding of α-lytic protease and the role of the pro region, *Proteins: Struct., Funct., Genet. 41*, 21−32.
29. Sham, Y. Y., Ma, B., Tsai, C. J., and Nussinov, R. (2002) Thermal unfolding molecular dynamics simulation of *Escherichia coli* dihydrofolate reductase: thermal stability of protein domains and unfolding pathway, *Proteins: Struct., Funct., Genet., 46*, 308−320.
30. Hassan, S. A., and Mehler, E. L. (2002) A critical analysis of continuum electrostatics: the screened Coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins, *Proteins: Struct., Funct., Genet. 47*, 45−61.
31. Lazaridis, T., and Karplus, M. (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation, *J. Mol. Biol. 288*, 477−487.
32. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, *J. Comput. Phys. 23*, 327−341.
33. Petrova, P., Koca, J., and Imberty, A. (1999) Potential energy hypersurfaces of nucleotide sugars: ab initio calculations, force-field parameterization, and exploration of the flexibility, *J. Am. Chem. Soc. 121*, 5535−5547.
34. Tsai, J., Levitt, M., and Baker, D. (1999) Hierarchy of structure loss in MD simulations of src SH3 domain unfolding, *J. Mol. Biol. 291*, 215−225.
35. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000) The pfam protein families database, *Nucleic Acids Res. 28*, 263−266.
36. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol. 55*, 379−400.

37. Zgiby, S., Plater, A. R., Bates, M. A., Thomson, G. J., and Berry, A. (2002) A functional role for a flexible loop containing Glu182 in the class II fructose-1,6-bisphosphate aldolase from *Escherichia coli*, *J. Mol. Biol. 315*, 131−140.

38. Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin, *Biochemistry 25*, 1847−1851.

39. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1999) Limited proteolysis in the study of protein conformation, in *Proteolytic enzymes: Tools and targets* (Sterchi, E. E., and Stocker, W., Eds.) pp 253−280, Springer, Heidelberg.

40. Polverino de Laureto, P., De Filippis, V., Di Bello, M., Zambonin, M., and Fontana, A. (1995) Probing the molten globule state of α-lactalbumin by limited proteolysis, *Biochemistry 34*, 12596−12604.

41. Polverino de Laureto, P., Scaramella, E., Frigo, M., Wondrich, F. G., De Filippis, V., Zambonin, M., and Fontana, A. (1999) Limited proteolysis of bovine α-lactalbumin: isolation and characterization of protein domains, *Protein Sci. 8*, 2290−2303.

BI034046R