J|A|C|S
ARTICLES

Published on Web 06/11/2010

In Silico Prediction of the 3D Structure of Trimeric Asialoglycoprotein Receptor Bound to Triantennary Oligosaccharide

Sai Kumar Ramadugu,[†] Ying-Hua Chung,[†] Ernesto J. Fuentes,[‡] Kevin G. Rice,[§] and Claudio J. Margulis*,[†]

Department of Chemistry, Department of Biochemistry, and Division of Medicinal and Natural Products Chemistry, University of Iowa, Iowa City, Iowa 52241

Received March 15, 2010; E-mail: claudio-margulis@uiowa.edu

Abstract: In this study, we present a general-purpose methodology for deriving the three-dimensional (3D) arrangement of multivalent transmembrane complexes in the presence of their ligands. Specifically, we predict the most likely families of structures of the experimentally intractable trimeric asialoglycoprotein receptor (ASGP-R), which consists of human hepatic subunits (two subunits of H1 and one subunit of H2), bound to a triantennary oligosaccharide (TA). Because of the complex nature of this multivalent type-II transmembrane hetero-oligomeric receptor, structural studies have to date been unable to provide the 3D arrangement of these subunits. Our approach is based on using the three-pronged ligand of ASGP-R as a computational probe to derive the 3D conformation of the complex and then using this information to predict the relative arrangement of the protein subunits on the cell surface. Because of interprotein subunit clashes, only a few families of TA conformers are compatible with the trimeric structure of ASGP-R. We find that TA displays significant flexibility, matching that detected previously in FRET experiments, and that the predicted complexes derived from the viable TA structures are asymmetric. Significant variation exists with respect to TA presentation to the receptor complex. In summary, this study provides detailed information about TA−ASGP-R interactions and the symmetry of the complex.

1. Introduction

The asialoglycoprotein receptor (ASGP-R), identified by Ashwell and co-workers, functions as a transport receptor on the surface of hepatocytes. It binds to desialylated serum glycoproteins with exposed terminal galactose (Gal) and *N*-acetylgalactosamine (GalNAc) residues, eliciting receptor-mediated endocytosis operated via clathrine-coated pits.[1−4] The receptor consists of two polypeptide subunits, H1 and H2 (in human hepatic cells) or RHL1 and RHL2/3 (in rat hepatocytes), arranged as a hetero-oligomer.[5−8] Several groups have proposed that the minimum ratio of the two polypeptides is 2:1, although some experiments have indicated H1/H2 ratios that are in the range 2−5:1−2.[9,10] It is thought that the presence of a hetero-oligomeric system confers high affinity to the receptor.[11] This receptor is known to be highly specific for the Gal or GalNAc residues. ASGP-Rs bind with high affinity to Gal or GalNAc in the form of tri- and tetraantennary oligosaccharides, whereas they bind with lower affinity to Gal or GalNAc in the form of biantennary oligosaccharides. The K_d values for binding of tri- and tetraantennary oligosaccharides to ASGP-R were found to be in the nanomolar range, while that for biantennary ones was found to be in the micromolar range.[12−14] In addition, when the terminal linkage is changed from β 1→4 to β 1→3, the binding affinity decreases significantly. This shows that ASGP-Rs are very sensitive to the geometry of their binding ligands.[15−18]

Many of the aforementioned binding studies were completed before the crystal structure of any carbohydrate recognition domain (CRD) of ASGP-R had been determined. Most C-type lectins such as ASGP-R show significant sequence similarity

[†] Department of Chemistry.
[‡] Department of Biochemistry.
[§] Division of Medicinal and Natural Products Chemistry.

(1) Ashwell, G.; Harford, J. *Annu. Rev. Biochem.* **1982**, *51*, 531–554.
(2) Spiess, M. *Biochemistry* **1990**, *29*, 10009–10018.
(3) Drickamer, K.; Taylor, M. E. *Annu. Rev. Cell Biol.* **1993**, *9*, 237–264.
(4) Breitfeld, P. P.; Simmons, C. F., Jr.; Strous, G. J.; Geuze, H. J.; Schwartz, A. L. *Int. Rev. Cytol.* **1985**, *97*, 47–95.
(5) McPhaul, M.; Berg, P. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8863–8867.
(6) Shia, M. A.; Lodish, H. F. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1158–1162.
(7) Bischoff, J.; Libresco, S.; Shia, M. A.; Lodish, H. F. *J. Cell Biol.* **1988**, *106*, 1067–1074.
(8) Bischoff, J.; Lodish, H. F. *J. Biol. Chem.* **1987**, *262*, 11825–11832.
(9) Lodish, H. F. *Trends Biochem. Sci.* **1991**, *16*, 374–377.
(10) Bider, M. D.; Wahlberg, J. M.; Kammerer, R. A.; Spiess, M. *J. Biol. Chem.* **1996**, *271*, 31996–32001.
(11) Bider, M. D.; Spiess, M. *FEBS Lett.* **1998**, *434*, 37–41.
(12) Baenziger, J. U.; Fiete, D. *Cell* **1980**, *22*, 611–620.
(13) Lee, Y. C.; Townsend, R. R.; Hardy, M. R.; Lonngren, J.; Arnarp, J.; Haraldsson, M.; Lonn, H. *J. Biol. Chem.* **1983**, *258*, 199–202.
(14) Townsend, R. R.; Hardy, M. R.; Wong, T. C.; Lee, Y. C. *Biochemistry* **1986**, *25*, 5716–5725.
(15) Rice, K. G.; Wu, P.; Brand, L.; Lee, Y. C. *Biochemistry* **1991**, *30*, 6646–6655.
(16) Rice, K. G.; Lee, Y. C. *J. Biol. Chem.* **1990**, *265*, 18423–18428.
(17) Wu, P. G.; Rice, K. G.; Brand, L.; Lee, Y. C. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9355–9359.
(18) Rice, K. G.; Weisz, O. A.; Barthel, T.; Lee, R. T.; Lee, Y. C. *J. Biol. Chem.* **1990**, *265*, 18429–18434.

in their CRDs. An important analysis of such CRDs was provided by Kolatkar and Weis.[19] In that study, the authors mutated the active site in the CRD of mannose binding protein A (MBP A, another C-type lectin) to that of ASGP-R and solved the structure of the protein bound to Gal. The modified protein was termed QPDWG because of the mutations $Glu^{185} \rightarrow Gln^{185}$, $Asn^{187} \rightarrow Asp^{187}$, and $His^{189} \rightarrow Trp^{189}$ along with the insertion of a Gly-rich loop. Gal and GalNAc were able to bind to the CRD of QPDWG, but mannose was not able to do so. The structure of the H1 CRD of ASGP-R was later determined by Meier et al.[20] However, it was not cocrystallized with Gal or GalNAc. Since the overlay of the structures of the CRDs of the QPDWG mutant and H1 resulted in a very low root-mean-square deviation (rmsd), Meier and co-workers reasonably concluded that the conformation of Gal-bound QPDWG found in the X-ray structure of Kolatkar and Weis should be analogous to that of Gal-bound H1. Although H2 remains structurally uncharacterized, H1 and H2 share 68% sequence homology in the CRD and an overall 58% sequence similarity.[21,22]

Many studies have now provided important information on the binding affinity of TA and the H1/H2 subunit ratio on the cell surface; however, none has addressed the 3D arrangement of these subunits when they are bound to TA. In the study presented here, we have used computational techniques to predict the 3D protein subunit arrangement on the cell membrane as well as its conformational variability. Our novel approach takes advantage of the recently developed fast sugar structure prediction software (FSPS)[23−26] and the replica-exchange molecular dynamics (REMD) method[27−30] to identify all of the conformations of TA that do not conflict with binding to ASGP-R. Our computational procedure also identifies the 3D arrangements of the H1 and H2 protein subunits that are compatible with TA binding.

2. Methods

In order to achieve the goal of predicting the 3D arrangement of a multivalent hetero-oligomeric transmembrane protein in a ligand-bound conformation, we used several distinct and powerful computational techniques. Our in-silico protocol was based on using TA, a complex but computationally tractable molecule, to probe the 3D arrangement of the much more complex protein receptor. In order to do this, several obstacles had to be overcome. First, a complete ensemble of TA conformational structures was obtained. Second, a reliable approximation to the Gal−H1/H2 binding geometry was obtained. Third, a homology model for H2 was constructed. Finally, because the predicted complexes should be stable in solution, explicit solvent simulations had to be run.

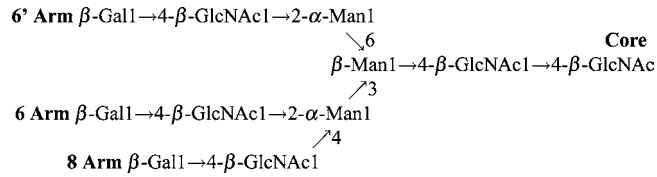

Figure 1. Sequence of a triantennary oligosaccharide (TA). The 1→6-linked arm is conventionally termed the 6′ Arm, the 1→2-linked arm the 6 Arm, and the 1→4-linked arm the 8 Arm.

Table 1. Dihedral Angle Increments for Linkages in TA[a]

linkage	belongs to	scanning increment
α-Man1→6β-Man	6′ Arm	45°
β-GlcNAc1→2α-Man	6 Arm	45°
β-GlcNAc1→4α-Man	8 Arm	45°
α-Man1→3β-Man	Core	60°
β-Gal1→4β-GlcNAc	6′, 6, and 8 Arms	180°
β-GlcNAc1→4β-GlcNAc	Core	120°
β-Man1→4β-GlcNAc	Core	120°
β-GlcNAc1→2α-Man	6′ Arm	120°

[a] See Figure 1 for definitions of the Core and the 6′, 6, and 8 Arms.

The TA ligand consists of three branches known as antennae, each of which has a terminal β-Gal residue that can bind to CRDs, such as those of H1 and H2 of ASGP-R. The sequence of TA is shown in Figure 1. Assuring full conformational sampling of a complex branched oligosaccharide is always challenging because most computational techniques are prone to fail when ergodicity problems are present. These ergodicity problems arise because of the coupled nature of dihedral rotations at branching points or crowded linkages. In order to guarantee that our sampling was reasonably complete, we used two independent techniques: our in-house-developed FSPS[23−26] and the computationally expensive REMD method.[27−30]

2.1. FSPS Studies. In the case of TA, the conformational flexibility is mainly due to glycosidic angle variability at the locations of branching. In order to better sample the conformations of TA, we used FSPS to conduct a coarse-grained search in glycosidic space.[23−26] Briefly, the protocol followed in a typical FSPS calculation consists of the following steps:

- An exhaustive linked coarse-grained search in $\phi-\psi$ space for all linkages, for use in discarding sterically disallowed conformations.
- Energy minimizations of all of the allowed conformers in the gas phase or in an implicit solvent.
- Pooling of the structures into "unique families" defined on the basis of their energetic and angular similarities. For example, in this work we required all of the members of a given family to have ϕ and ψ angular differences of <20° and energy differences of <5 kcal/mol.
- Calculation of NMR observables (nuclear Overhauser effects, residual dipolar couplings, J couplings) for a representative member of each family in order to rank its likelihood in comparison with experimental observables.

The initial 3D model of TA was built using the YASARA software, version 9.6.14.[31] In this work, we were not interested in deriving the most likely structure of TA in solution but instead were concerned with its conformational structure when bound to ASGP-R. Therefore, the last two steps in the protocol were not needed. In order to balance accuracy and computational cost, we used the angular scanning increments shown in Table 1. A typical Ramachandran space basin of attraction has a radius of ∼50°. We chose finer increments when close to branching points, since these dihedral angles provide TA with most of its conformational variability. This coarse graining resulted in 167 792 allowed conformations without steric clashes. These conformations were

(19) Kolatkar, A. R.; Weis, W. I. *J. Biol. Chem.* **1996**, *271*, 6679–6685.
(20) Meier, M.; Bider, M. D.; Malashkevich, V. N.; Spiess, M.; Burkhard, P. *J. Mol. Biol.* **2000**, *300*, 857–865.
(21) Spiess, M.; Lodish, H. F. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6465–6469.
(22) Fuhrer, C.; Geffen, I.; Huggel, K.; Spiess, M. *J. Biol. Chem.* **1994**, *269*, 3277–3282.
(23) Xia, J.; Daly, R. P.; Chuang, F.-C.; Parker, L.; Jensen, J. H.; Margulis, C. J. *J. Chem. Theory Comput.* **2007**, *3*, 1629–1643.
(24) Xia, J.; Daly, R. P.; Chuang, F.-C.; Parker, L.; Jensen, J. H.; Margulis, C. J. *J. Chem. Theory Comput.* **2007**, *3*, 1620–1628.
(25) Xia, J.; Margulis, C. *J. Biomol. NMR* **2008**, *42*, 241–256.
(26) Xia, J.; Margulis, C. *Biomacromolecules* **2009**, *10*, 3081–3088.
(27) Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. *Chem. Phys. Lett.* **2001**, *335*, 435–439.
(28) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
(29) Seibert, M. M.; Patriksson, A.; Hess, B.; Spoel, D. V. D. *J. Mol. Biol.* **2005**, *354*, 173–183.
(30) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(31) Krieger, E.; Koraimann, G.; Vriend, G. *Proteins* **2002**, *47*, 393–402.

subjected to gas-phase minimizations using the GROMACS 4.0 software[32] and the OPLS-AA force field.[33] Results derived using this method were compared against much more expensive REMD simulations in explicit solvent.

2.2. REMD Simulations. The REMD technique is based on parallel simulations of identical system replicas (M_1 to M_n) at different temperatures (T_1 to T_n).[27−30] Snapshots of the system throughout the simulation are exchanged on the basis of a Monte Carlo acceptance scheme.[34] At higher temperature, the system is able to cross barriers that are inaccessible at lower temperatures. The protocol for exchange guarantees correct Boltzmann sampling at each of the studied temperatures for which free energies can be derived if all of the ergodicity problems are surmounted. REMD simulations have been applied to the problem of protein folding (e.g., see ref 35) as well as studies of many other systems, such as DNA−carbon nanotube structures[36] and DNA−RNA dynamics.[37]

As an initial equilibration step, we minimized TA in a cubic water box consisting of 4055 simple point charge (SPC) water molecules[38] using the steepest-descent algorithm. Subsequently, we carried out an *NPT* (constant number of particles *N*, pressure *P*, and temperature *T*) MD equilibration of the system for 1 ns at 300 K. All of the simulations were carried out using the GROMACS 4.0 software[32] and the OPLS-AA force field.[33] In order to properly describe the electrostatic interactions in a periodically replicated system, we used the particle-mesh Ewald (PME)[39] method with a real-space cutoff of 9 Å. The final structure from this equilibration run was used as the initial condition for the REMD simulations. Our REMD simulations consisted of 48 MD replicas at temperatures ranging from 300 to 396 K in 3 K increments. The choice of the number of replicas and the temperatures was generated using the protocol of Van der Spoel[40] (also see http://folding.bmc.uu.se/remd/index.php). Our explicit-solvent REMD simulations were 35 ns in duration. This time scale was deemed appropriate on the basis of the observed convergence of the Ramachandran plots for the crowded linkage points at the branchings of the 6′ Arm, 6 Arm, and 8 Arm, which were almost unchanged after 15 ns. The integration time steps for the MD and REMD simulations were set at 1 and 2 fs, respectively. Exchange of replicas was attempted every 1 ps during the 35 ns simulation. We used the Nose−Hoover coupling scheme[41,42] with a coupling τ value of 0.05 ps for temperature control and the Parinello−Rahman coupling scheme[43] with a coupling parameter of 1.0 ps for pressure control. In all of the simulations, the cutoff value for van der Waals interactions was set to 0.9 nm. This simulation resulted in 17 500 models of TA saved from the lowest-temperature replica at a frequency of 2.5 ps.

2.3. Structural Alignment Studies. Our structural alignment studies consisted of three steps, which are described in sections 2.3.1 to 2.3.3. All of the alignments were performed using the PyMOL Software (DeLano Scientific, Palo Alto, CA).[44]

2.3.1. Identification of the Binding Site for H1. The first predictive step involved deriving the orientation of Gal in the H1 binding site. Since the structure of the H1 subunit with its binding saccharide had not been solved,[20] we used the X-ray structure of Gal bound to the QPDWG mutant[19] as a template to orient Gal with respect to H1. This approach was expected to result in only negligible error, given the almost identical nature of the two binding sites (see below). To limit the complexity of the predictive problem, we did not attempt to predict variations of the well-established binding mode for the interaction between the terminal Gal of TA and H1 or H2.

2.3.2. Alignment of the Trimeric H1 of ASGP-R with TA. The second predictive step involved deriving all of the feasible 3D arrangements of the ASGP-R protein subunits in the complex. The structure of H2 has not been determined to date. Since the H1 and H2 subunits are highly similar from a sequence and functional perspective, the structure of H2 is expected to be similar to that of H1. Therefore, we first used three H1 subunits instead of two H1 subunits and one H2 subunit. We later replaced the H1 subunit connected at the 8 Arm with our homology model of H2 described in section 2.3.3. In order to derive all of the feasible 3D arrangements of the three H1 subunits, we attached an ASGP-R subunit to each of the three Gal-terminated TA antennae using the orientation derived in section 2.3.1. In other words, each TA conformer derived from our REMD simulation was associated with a unique TA−ASGP-R 3D arrangement, which could in principle have steric clashes. Structures with steric clashes were subsequently discarded.

The following protocol was followed for all of the structural alignments:

- The terminal Gal residue of the 6′ Arm of TA was aligned with the Gal binding site of H1 derived from our comparison with QPDWG using the pairfit command in the PyMOL software. The coordinates of one H1 subunit and TA were saved to a pdb file for further alignment of the other two subunits.
- The terminal Gal binding site of the 6 Arm of TA in the monomeric complex was aligned with the Gal binding site of a second H1, and the dimeric complex was saved to a new pdb file.
- The terminal Gal residue of the 8 Arm of TA in the dimeric complex was aligned with a third H1 subunit, and the final structure was saved in pdb format.

2.3.3. Homology Modeling of CRD of the H2 Subunit. The final predictive step involved deriving a homology model of the structurally uncharacterized H2 subunit. An attempt to derive the structure of H2 and H1 from homology to the QPDWG mutant was published[45] before the crystal structure of H1 was resolved. In contrast to this early study,[45] our homology model used the known structure of H1 to derive that of H2. The residues coordinating Ca^{2+} in the binding pocket of H1 and H2 subunits are highly conserved.[21] Furthermore, the amino acid sequence alignments of ASGP-Rs from rat, mouse, and humans show that the residues that ligate with Ca^{2+} in each of these structures are identical. Thus, it is highly probable that the H1 and H2 subunits interact with Gal in a very similar manner. These findings provide an excellent starting point for the theoretical derivation of a reasonable structure for the H2 subunit. Homology modeling of the H2 subunit was carried out using the MODELLER software, version 9v6.[46,47]

(32) Hess, B.; Kutzner, C.; Spoel, D. V. D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(33) Damm, W.; Frontera, A.; TiradoRives, J.; Jorgensen, W. *J. Comput. Chem.* **1997**, *18*, 1955–1970.
(34) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96–123.
(35) García, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.
(36) Johnson, R. R.; Kohlmeyer, A.; Johnson, A. T. C.; Klein, M. L. *Nano Lett* **2009**, *9*, 537–541.
(37) Curuksu, J.; Zacharias, M. *J. Chem. Phys.* **2009**, *130*, 104110.
(38) Jorgensen, W. L.; Chandrashekar, J.; Madura, J.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(39) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(40) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073–2077.
(41) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
(42) Hoover, W. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(43) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
(44) DeLano, W. L. *The PyMOL User's Manual*; DeLano Scientific: Palo Alto, CA, 2002.
(45) Bianucci, A. M.; Chiellini, F. *J. Biomol. Struct. Dyn.* **2000**, *18*, 435–451.
(46) Eswar, N.; Eramian, D.; Webb, B.; Shen, M.-Y.; Sali, A. *Methods Mol. Biol.* **2008**, *426*, 145–159.

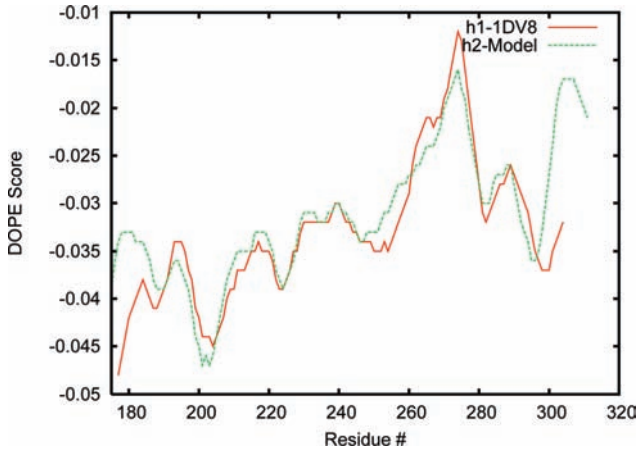

Figure 2. Discrete optimized potential energy (DOPE) score per residue for the template (H1 subunit of ASGP-R, PDB entry 1DV8) and the target (H2 subunit of ASGP-R). The target (H2 subunit) is shown in green and the template (H1 subunit) in red.

The protocol consisted of three steps:

- The MODELLER software has an internal database of related protein sequences with greater than 95% similarity. The sequence of H2 was used to query this database. Fourteen templates with sequence identity greater than 30% were selected for use in the second step in the protocol.
- The second step involved the use of a more accurate 2D alignment program within MODELLER that takes into account structural information. All 14 template structures were considered. A sequence identity table and a clustering tree (dendrogram) expressing differences among closely related templates were generated. For the next step, three aspects were important: the resolution of the crystal structure, sequence similarity data (from the previous step), and dendrogram data from the present step. On the basis of these three criteria, the best choice for the next step (3D model building) was the H1 subunit (which had a resolution of 2.0 Å, 68% sequence similarity, and a close relation to H2 through dendrogram data).
- The structure of H1 (PDB entry 1DV8) and the sequence of H2 were used to build a 3D model for H2. Five candidates were proposed by the software. To evaluate these models, we used a routine approach based on the objective score function and the discrete optimized protein energy (DOPE) score per residue. The DOPE score is a statistical potential based on pair distribution functions for all of the atoms in the native structure.[48] A DOPE score value of 0.00 and above corresponds to higher energy, and negative values correspond to lower energy. This evaluative method is described in the MODELLER manual. The DOPE score for the template and best target are overlaid in Figure 2. The similarity between the scores for the template and the target indicates high spatial homology, except for moderate deviations at the H2 protein termini.

2.4. Molecular Dynamics Studies of the Predicted Complex. All 17 500 trimeric H1−TA complex candidates were checked for intersubunit steric clashes. Out of these 17 500 candidate structures, only 22 had no intersubunit clashes. With the core of the sugar oriented toward the $+Z$ axis and the viewpoint chosen to be along this axis, about half of the 22 structures had the H1 subunits connected to the 6′, 6, and 8 Arms arranged in a clockwise fashion. The remaining structures were arranged in a counterclockwise fashion (the definitions of clockwise and counterclockwise are based on the arrangement of the arms; see section 3.2). In order to further refine these two families of structures, the H1 subunit connected to the 8 Arm was replaced by our homology model of H2. The choice of the 8 Arm was based on experimental evidence from Rice et al.[18] showing that the 8 Arm Gal binds specifically to the RHL2/3 subunit (analogous to H2). Four of the 22 candidate complexes, two belonging to each orientation (clockwise and counterclockwise) of the three subunits, were subjected to simulated-annealing molecular dynamics (SAMD). These simulations were carried out in explicit solvent at constant volume using a temperature range of 500−300 K sampled every 10 K for a total of 10 ns. All of the simulations were performed using the GROMACS 4.0 software[32] and the OPLS-AA force field.[33] To prevent sugar detachment at the initial nonphysiological high temperature of the SAMD simulations, constraints on the distances between Ca^{2+}, the binding-site atoms in each of the protein subunits, and the 3′-OH and 4′-OH of the terminal Gal residues of the oligosaccharide were imposed. These 24 distance constraints were derived from the model predicted by our structural alignment studies (see above). In each case, the total number of atoms periodically replicated was ∼100 000. Upon conclusion of the cooling temperature ramp, all of the bond constraints were removed, and the dynamics of the system was further followed for 2.5 ns in the *NPT* ensemble at 300 K. Independently, for further comparison, one of the complexes was studied for 12 ns at 300 K with and without replacement of H1 by H2 at the 8 Arm position.

3. Results and Discussion

3.1. FSPS and REMD Studies of Free TA. A comparison of the fully coupled Ramachandran plots for the 1→6 linkage of the 6′ Arm, the 1→2 linkage of the 6 Arm, and the 1→4 linkage of the 8 Arm based on the FSPS and REMD simulations are shown in Figure 3. These linkages are the most important for determining the overall structure of TA. We see from Figure 3 that both methods describe almost the same dihedral space, except for fairly small deviations arising from the fact that REMD was carried out in explicit solvent whereas FSPS is a gas-phase procedure. The overlap detected gave us confidence that we were indeed sampling all of the possible conformations of TA in solution. An important difference between the two approaches is that FSPS required only 3 days and 10 processors to obtain these structures, whereas REMD required 3 months and 48 processors. Examples of the different conformations of TA derived from the REMD simulations are displayed in Figure 4. In this plot, all of the conformers are displayed with the 6 and 8 Arms pointing in the same direction. Our plots of the explicit-solvent REMD distance distributions at 300 K and the corresponding free energies derived from them are shown in Figure 5. The distributions are of the distances from the C1 atoms in the terminal Gal's of the arms to the C1 atom in the terminal GlcNAc in the core region. From the width of its probability distribution, it is clear that the 6′ Arm displays the largest range of possible distances. Flexibility is a well-known characteristic of 1→6 linkages[49] and is due to the presence of ω, the third dihedral angle. The probability distribution for the 6′ Arm displays a major peak at 16 Å along with other less-prominent peaks at 9 and 12 Å at ∼0.6 and ∼1.7 kcal/mol, respectively, above the free-energy minimum. The probability distribution for the 6 Arm displays two peaks centered around 14 and 21 Å, with corresponding free-energy differences of 1.8 kcal/mol. The 8 Arm shows a single peak centered around 22 Å. On the basis of these free-energy profiles, transitions from

(47) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. *Current Protocols in Bioinformatics*; Wiley-Interscience: New York, 2006; Chapter 5, Unit 5.6.
(48) Shen, M.-Y.; Sali, A. *Protein Sci.* **2006**, *15*, 2507–2524.
(49) Kirschner, K.; Woods, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–10545.

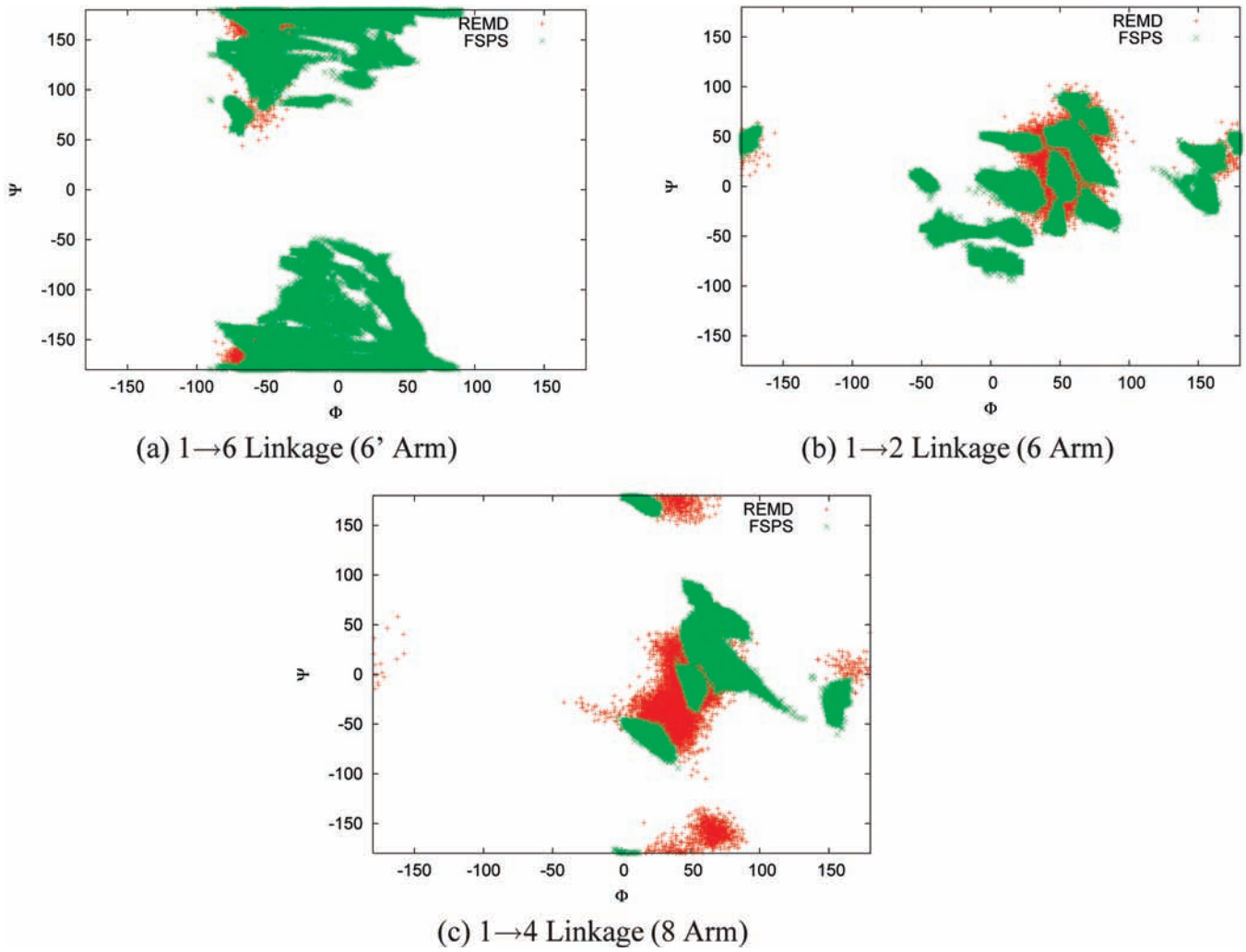

Figure 3. Dihedral angle maps for the (a) 6′, (b) 6, and (c) 8 Arms of TA. The results from our REMD simulations are shown in red and those from the FSPS in green.

the higher-free-energy conformational states to those at lower free energies appear to be almost barrierless. These results can be compared with those obtained in fluorescence resonance energy transfer (FRET) experiments performed by Rice et al.,[15] in which a two-peaked distribution (corresponding to our peaks at 9 and 16 Å) was observed for the 6′ Arm. These represent folded and extended populations, consistent with previous gas-phase molecular dynamics studies of TA by Balaji and co-workers.[50] The intermediate feature at 12 Å in our probability distribution also appears to be important. It turned out that the four independent SAMD simulations of the full ASGP-R−TA complex in explicit solvent gave rise to sugar structures matching the extended, folded, and intermediate 6′ Arm configurations (see below). In agreement with our simulations, the FRET study[15] also identified a folded and extended set of populations for the 6 Arm and only a single population in the case of the 8 Arm. These results, along with those of the biochemical experiments described in refs 51 and 52, appear to indicate that the inherent flexibility of TA is important for the high-affinity binding of TA to ASGP-R.

3.2. Structural Alignment Studies. Probing the structure of the TA−ASGP-R complex experimentally has proven to be a daunting problem, and to date the system has been refractory to high-resolution structural analysis. Thus, a number of important questions remain. What is the symmetry of the complex? What is the global 3D arrangement of the protein subunits? Is there more than one viable arrangement? Our goal was to predict the broad features of the complex, or set of possible complexes, computationally. Focusing on specific details of interprotein subunit amino acid interactions would have been unwise in this case, since we derived the structure of the H2 subunit from homology. By taking advantage of experimentally derived structural data on the mode whereby Gal binds to the CRD of QPDWG and having derived a complete pool of TA conformers, our alignment studies were able to

(50) Balaji, P. V.; Qasba, P. K.; Rao, V. S. *Biochemistry* **1993**, *32*, 12599–12611.

(51) Andre, S.; Kojima, S.; Gundel, G.; Russwurm, R.; Schratt, X.; Unverzagt, C.; Gabius, H. *Biochim. Biophys. Acta* **2006**, *1760*, 768–782.

(52) Rice, K. G.; Lee, Y. C. *Adv. Enzymol.* **1993**, *66*, 41–83.

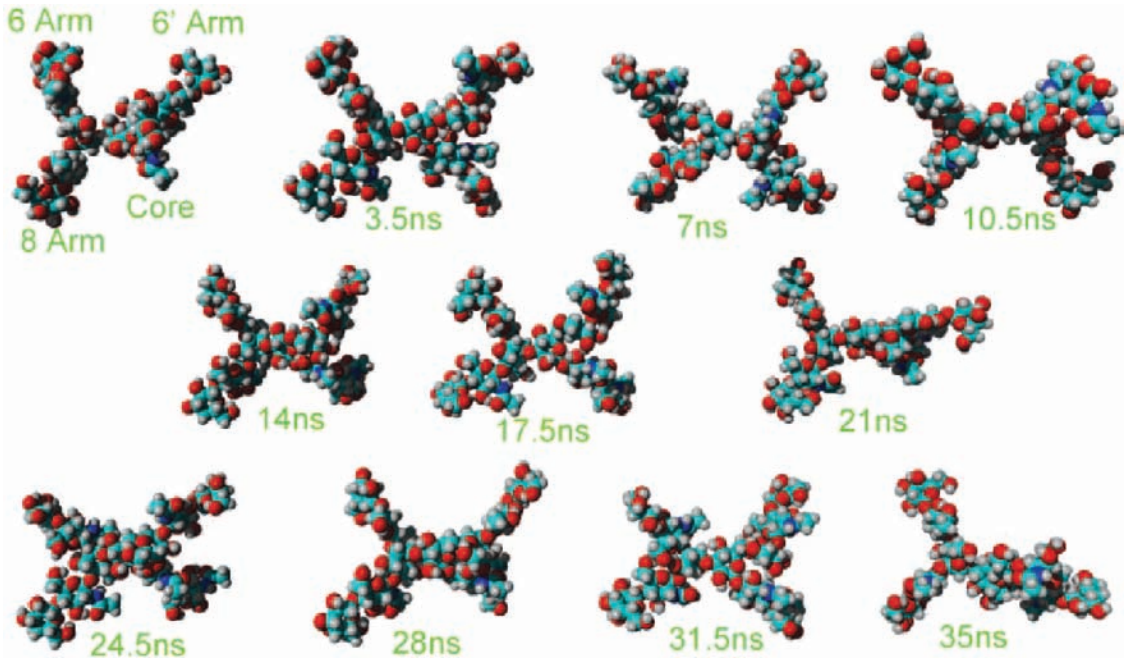

Figure 4. Representative structures of TA derived from the REMD simulations. The orientation of the 6′, 6 and 8 Arms is indicated in the first snapshot and is the same for all of the other conformations.

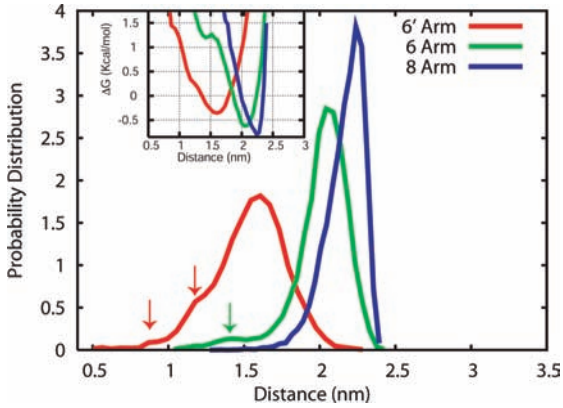

Figure 5. Normalized probability distributions for the distances between the C1 atoms in the terminal Gal's in the 6′, 6, and 8 Arms and the C1 atom in the terminal GlcNAc in the core region. Arrows indicate peaks to which the text refers. The inset shows the corresponding free-energy functions derived from these probability distributions.

predict the subset of these conformers that can bind the three subunits of ASGP-R through the terminal Gal's without steric clashes.

Our pairwise structural alignment of the H1 subunit structure of ASGP-R to the QPDWG mutant of MBP A with bound Gal is shown in Figure 6. This alignment resulted in an rmsd of 1.27 Å for all atoms (the value for C_α is ∼0.94 Å). Given the almost perfect overlap of the CRDs in the two proteins, we were able to construct 17 500 models of TA bound to the three H1 subunits by aligning each terminal Gal on TA to the known crystal structure of Gal bound to the QPDWG mutant. Only 22 out of the 17 500 were viable structures for the complex (i.e., ones in which no interprotein subunit steric clashes were detected).

These 22 viable complexes could be separated into two families distinguished by the "relative chirality" of the protein subunit arrangement (see Figure 7). The chirality arises from the inherent flexibility of the three arms of TA. At this point, we replaced the H1 subunit connected to the 8 Arm with our homology model of the H2 subunit. Since the CRDs of H2 and H1 are highly similar to that of QPDWG, variations on the well-established mode of binding between Gal and QPDWG were not expected. Because of the high level of similarity between the two protein structures, no clashes resulted from this substitution.

In the absence of further knowledge of the constraints imposed by the three membrane-bound protein stalk regions and the locations of glycosyl moieties on the H1 and H2 subunit surfaces, both of the families of complexes described above represent reasonable predictions of the trimeric complex. Not only do the two H1 units bind to the 6 and 6′ Arms and the H2 unit to the 8 Arm, but also, the structures accommodate a large glycoprotein attached to the core region of the sugar. This is a requirement for any predictive model, since in the biological system TA is attached to a variety of different proteins through its core.

When bound to TA, all of the candidate complexes are clearly asymmetric. The distances between terminal Gal residues in the 6′ and 6, 6 and 8, and 6′ and 8 Arms are 20.9 ± 1.3, 12.5 ± 1.4, and 23.3 ± 3.1 Å, respectively. Clearly, the largest variability is in the distance between the 6′ and 8 Arms, which in some candidate complexes are separated by 16−25 Å.

The asymmetry in interarm distances seen in each of the 22 candidate complexes is consistent with the prediction of asymmetry made by Lodish[9] on the basis of simple geometric arguments using Rice's distance distributions for TA derived from FRET experiments. Interestingly, a symmetric structure of trimeric MBP A in the absence of oligosaccharide has been crystallized.[53,54] If the results of our study apply more generally to other membrane-bound complexes such as trimeric MBP A, unless the ligand is highly symmetrical, the complex must be able to adapt in order to accept the ligand. A recent article by

(53) Sheriff, S.; Chang, C. Y.; Ezekowitz, R. A. *Nat. Struct. Biol.* **1994**, *1*, 789–794.
(54) Weis, W. I.; Drickamer, K. *Structure* **1994**, *2*, 1227–1240.

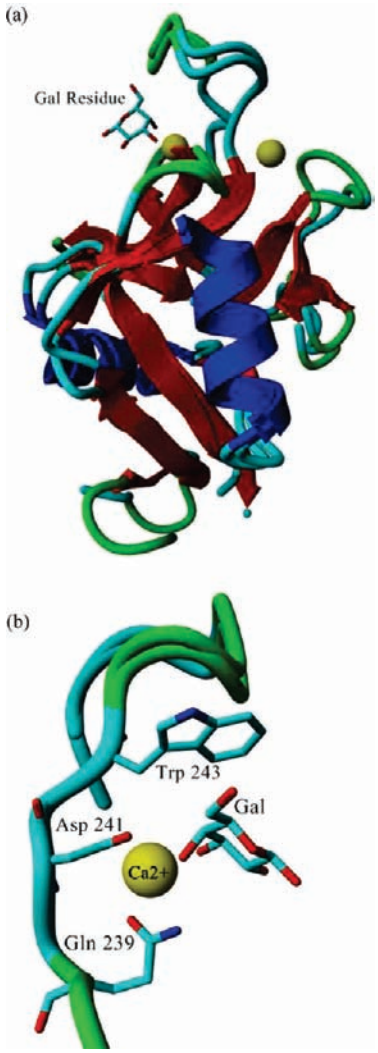

Figure 6. Structural alignment of the H1 subunit of ASGP-R with the QPDWG mutant of MBP A. (a) All-atom alignment of the H1 subunit with the QPDWG mutant. The all-atom rmsd between the two proteins was 1.27 Å. Bound Gal is shown in stick format. (b) Zoom-in view of the active site. The labeled residues are those of H1.

Menon and co-workers[55] on DCSIGN supports the latter prediction. Using force-pulling experiments, these investigators showed that as the ligand approaches, it reaches a certain distance at which a significant conformational change in the protein complex occurs. This type of conformational rearrangement has also been observed in immunoglobulins.[56]

To provide further details on the structure and dynamics of the full complex in solution, we carried out two different types of MD simulation studies. The first involved two constant-temperature and -pressure simulations at least 10 ns in duration that compared changes in the H1−H1−H2 complex to those in the H1−H1−H1 homotrimer analogue. The second was a set of SAMD simulations with a temperature ramp covering the range 500−300 K. The purpose of the second study was to measure the structural relaxation of the complex after a computational temperature jump. Two clockwise and two counterclockwise arrangements of the complex were used as initial conditions for the four SAMD simulations.

3.3. Simulated Annealing Studies of the H1−H1−H2−TA Complex. Simulated annealing (SA) is a technique in which the system is first heated to high temperatures in order to disrupt the structure and allow crossing of energy barriers and then cooled to lower temperatures in order to allow the system to relax energetically. SA has been applied, for example, to predict the conformation of protein side chains.[57,58] As described in section 2.4, the distances between the ligand and binding receptors were constrained during the temperature ramp but unconstrained during the last 2.5 ns, during which the temperature of the system was kept constant at 300 K.

The radius of gyration (R_g) as a function of time for each of the models is shown in Figure 8. From this figure it is clear that the three-subunit complex can potentially adopt a variety of conformations while still being compatible with the bound oligosaccharide. This flexibility is evident from the fact that some complexes are compact but others are not. Some of these conformations are much more likely to be physiologically relevant than others, but probing the relative probability of each conformer would require exhaustive simulation studies, both in the presence of the membrane and with the inclusion of the receptor stalk, whose conformation is still unknown. Furthermore, detailed attention would have to be paid to glycosylation at the surface of the receptor proteins, and the results may be specific to the particular protein attached to the TA core. Thus, resolving these issues is beyond the ability of current computational technology. However, several fundamentally important conclusions can be drawn from our studies. First, two different binding chiralities are likely to exist. Second, in all cases the complex is asymmetrical. Third, the ASGP-R system appears to allow for significant fluidity, which is likely required for binding to a highly flexible and multiconformational sugar. This last point has important implications for our understanding of existing crystal structures of other sugar-binding complexes. Specifically, although these complexes may appear to be highly symmetrical, fluidity and the potential for conformational variability may be key requirements for binding to asymmetric and highly flexible *N*-glycans.

3.4. Comparison of Structure and Dynamics of Homo- and Heterotrimeric Complexes. In order to gain a molecular-level understanding of the differences in the energetics of the homo- and heterotrimers, we subjected both systems to MD simulations in explicit solvent at ambient pressure and temperature. In particular, we studied in detail the interactions between TA and the subunit connected to the 8 Arm. Figure 9 shows the total energy of the interaction between TA and the protein subunit connected to the 8 Arm as a function of time. There are significantly larger fluctuations when H2 is connected to the 8 Arm. Interestingly, this apparent on−off behavior observed at ∼1 ns and again at 4 ns in the case of H2 can be attributed solely to angular fluctuations of the side chain of ASP^{290} of H2 due to the interaction of the carboxylate with the Gal of the 8 Arm. This is corroborated by the energy of the interaction between this residue and TA, which is shown as a function of time in Figure 9. Though the H2−TA interaction energy appears to fluctuate more and its average appears to be lower, this does not necessarily mean that the free energy of binding to H2 is

(55) Menon, S.; Rosenberg, K.; Graham, S. A.; Ward, E. M.; Taylor, M. E.; Drickamer, K.; Leckband, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11524–11529.
(56) Bilgiccer, B.; Thomas, S. W., 3rd; Shaw, B. F.; Kaufman, G. K.; Krishnamurthy, V. M.; Estroff, L. A.; Yang, J.; Whitesides, G. M. *J. Am. Chem. Soc.* **2009**, *131*, 9361–9367.
(57) Correa, P. E. *Proteins* **1990**, *7*, 366–377.
(58) Lee, C.; Subbiah, S. *J. Mol. Biol.* **1991**, *217*, 373–388.

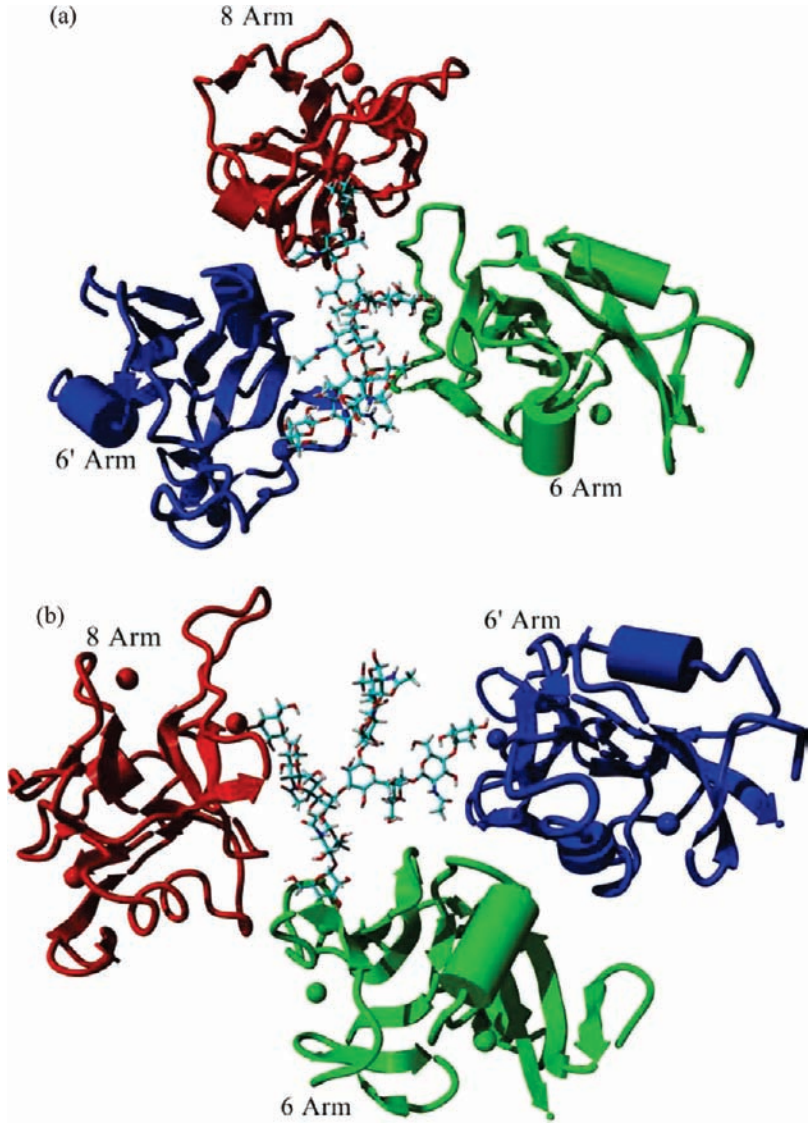

Figure 7. (a) Clockwise and (b) counterclockwise arrangements of the receptor subunits. Attached to the 6 and 6′ Arms are H1 subunits (green and blue, respectively), and attached to the 8 Arm is an H2 subunit (red). These two types of arrangements, clockwise and counterclockwise, were derived from our studies. Since the protein subunits are large, it is unlikely that there is exchange between these conformations once the sugar is bound to the protein trimer.

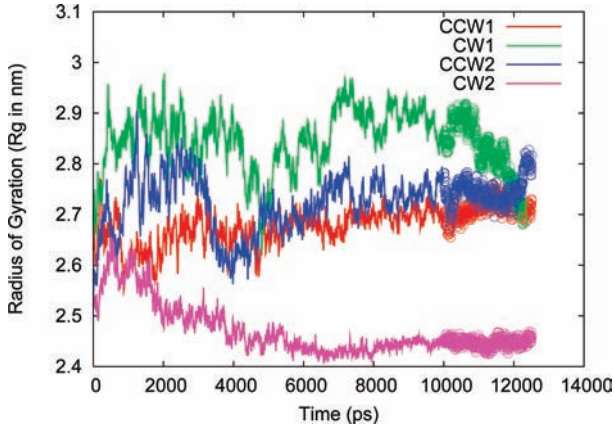

Figure 8. Radii of gyration as functions of time for the four models of TA bound to H1−H1−H2 from our SAMD simulations. The first 10 ns (solid lines) correspond to SAMD, whereas the subsequent 2.5 ns (circles) are unconstrained *NPT* simulations at 300 K. CCW refers to the counterclockwise model and CW to the clockwise model shown in Figure 7.

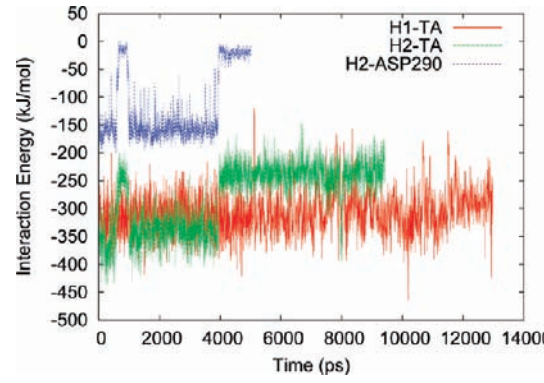

Figure 9. Comparison of TA−H1 subunit and TA−H2 subunit interaction energies at the 8 Arm. The red line represents the energy of the interaction between TA and the H1 at the 8 Arm in the homo-oligomeric complex as a function of time during the simulation. The green line represents the energy of the interaction between TA and the H2 at the 8 Arm in the hetero-oligomeric complex. The blue line represents the interaction between TA and ASP^{290} of H2.

lower. To determine whether that is true, entropy and interactions with the solvent must also be considered.

4. Conclusions

In this study, an exhaustive set of conformations of TA in solution were derived. These models display significant flexibility and are consistent with previous FRET experiments. On the basis of the crystal structure of QPDWG bound to Gal, a model of TA bound to H1 was derived. Although the H2 subunit in ASGP-R has never been crystallized, we built a model for it on the basis of homology, using H1 as a template. From the exhaustive set of conformations derived for TA in solution, all of the models able to accommodate two H1 subunits and one H2 subunit (attached at the 8 Arm) as well as a large protein attached to the core of the oligosaccharide were constructed. All of the predicted complexes are asymmetric, and significant variation in the way TA can be presented to its binding partners appears to be possible on the basis of our SA studies. If one assumes that in the absence of TA the triprotein system in the membrane is symmetrical, as in the case of the recently crystallized MBP A, then these proteins must undergo significant spatial rearrangements in order to bind TA, most likely in a stepwise process. This type of conformational rearrangement has been shown to occur in DCSIGN.[55] Our work is a first step toward understanding a very complicated system. We hope that these findings will guide additional experiments. Finally, the approach described in this article is quite general and can be used to study a large set of very important but poorly understood multivalent transmembrane complexes. We are in the process of applying this procedure to study other receptors.

Acknowledgment. E.J.F. thanks the Carver College of Medicine and the Department of Biochemistry for startup funds used to support this project. We thank Christine M. Blaumueller, Ph.D., for her help in editing the manuscript.

JA1021766