THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

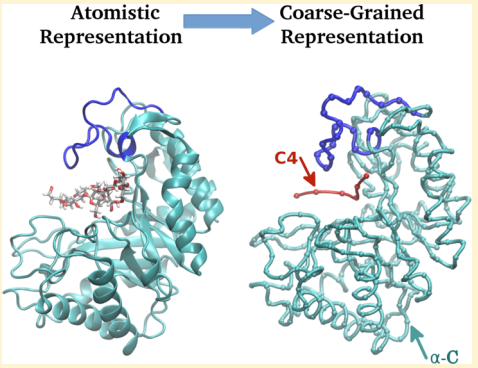
Polysaccharide–Protein Complexes in a Coarse-Grained Model

Adolfo B. Poma,* Mateusz Chwastyk, and Marek Cieplak

Institute of Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warsaw, Poland

Supporting Information

ABSTRACT: We construct two variants of coarse-grained models of three hexaoses: one based on the centers of mass of the monomers and the other associated with the C4 atoms. The latter is found to be better defined and more suitable for studying interactions with proteins described within α-C based models. We determine the corresponding effective stiffness constants through all-atom simulations and two statistical methods. One method is the Boltzmann inversion (BI) and the other, named energy-based (EB), involves direct monitoring of energies as a function of the variables that define the stiffness potentials. The two methods are generally consistent in their account of the stiffness. We find that the elastic constants differ between the hexaoses and are noticeably different from those determined for the crystalline cellulose Iβ. The nonbonded couplings through hydrogen bonds between different sugar molecules are modeled by the Lennard-Jones potentials and are found to be stronger than the hydrogen bonds in proteins. We observe that the EB method agrees with other theoretical and experimental determinations of the nonbonded parameters much better than BI. We then consider the hexaose-Man5B catalytic complexes and determine the contact energies between their the C4–α-C atoms. These interactions are found to be stronger than the proteinic hydrogen bonds: about four times as strong for cellohexaose and two times for mannohexaose. The fluctuational dynamics of the coarse-grained complexes are found to be compatible with previous all-atom studies by Bernardi et al.



I. INTRODUCTION

Polysaccharides, such as cellulose, are a major component of plant cell walls and, therefore, of biomass.[1] Their degradation into simple sugars is accomplished by bacteria that produce the necessary enzymes known as cellulases. The cellulases are either secreted into the solvent or act together when combined into cellulosomes.[2] Cellulose, $(C_6H_{10}O_5)_n$ with n of at least several hundred, is a long, unbranched polymer built of the molecules of D-glucose, denoted as D-GLC, that are connected by the β $(1 \rightarrow 4)$ glycosidic bonds.[3] The polymers typically form microcrystalline and polymorphic microfibrils[4−6] in which particular polymers are placed alongside one another. Their structure can be generated numerically by using a special toolkit.[7] Form I is the one that exists in nature and is thus called native. It is made of 36 chains[4,5] that are stabilized by a network of hydrogen bonds (HBs),[8] and it comes in two distinct allomorphs, Iα and Iβ. Iα is dominant in bacterial and algal celluloses, whereas Iβ is in higher plants. The Iα allomorph has a single cellulose chain in a triclinic cell, whereas Iβ has two chains in a monoclinic cell. Most cellulosic materials contain crystalline and amorphous domains[9] in proportions depending on source and preparation. Most of the reactants penetrate only the amorphous domains.

It is difficult to study such large aqueous biosystems at long time scales through all-atom simulations, especially if the systems involve interactions with proteins.[10] One way out is to consider coarse-grained (CG) models of polysaccharides and proteins and to introduce the implicit solvent.

There are several successful examples of implementing this approach, but only for single-component system (either protein or polysaccharide). These methods can be grouped according to the underlying physical principle which is employed to carry out the coarse graining. In this way, we find methods which target structural (or energy) features of the all-atom model and a very different class of methods which aim to reproduce thermodynamics data (e.g., oil/water partitioning coefficient). One example of the first methods for proteins is the UNRES model[11] and the second is the MARTINI method.[12] Their extensions to deal with polysaccharides can be found in refs 13−15. However, what needs to be developed is a CG framework that can describe protein–polysaccharide complexes in a unified manner.

Here we address this task by considering a still coarser description in which a monosaccharide, such as glucose, is described by one atom in analogy to an amino acid being replaced by a single bead located at the α-C atom as in many structure-based models of proteins.[16−24] Such a description would provide a framework for a unified approach to polysaccharide–protein systems in which effective atoms correspond to groups of comparable sizes. Our CG scheme

Received: June 26, 2015
Revised: August 18, 2015
Published: August 20, 2015

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

is set at a larger length scale than in the MARTINI method, in which an effective atom is typically assigned to 3−4 atoms.

An example of models built along these lines has been developed by Srinivas et al.[25] as well as Fan and Maranas.[26] In both of these models, monosaccharides are represented by effective atoms that are placed either at the centers of mass (CM) of the entire monomeric unit[25] or at the CM of the ring.[26] The interactions between the consecutive monomers are described by three potentials, V_b, V_θ, and V_ϕ, representing the harmonic pseudobond, the bond angle potential, and the torsional (or dihedral) potential, respectively. They involve summations over all bonds and angles, but the individual term contributions are

$$V_b = \frac{1}{2}k_r(r - r_0)^2$$

$$V_\theta = \frac{1}{2}k_\theta(\theta - \theta_0)^2$$

and

$$V_\phi = \frac{1}{2}k_\phi(\phi - \phi_0)^2$$

correspondingly. (Other possible forms of V_ϕ will be discussed later.) The quantities with the subscript 0 denote the equilibrium values of r, θ, and ϕ whereas k_r, k_θ, and k_ϕ are the corresponding force constants. Srinivas et al.[25] determine all of these parameters by the Boltzmann inversion (BI) method.[27] They have evolved the system of 36 celluloses of 80 units (in the crystalline case) in all-atom simulations with the CHARMM force field[28] and the TIP3P water molecules[29] and determined the probability distributions of unit-to-unit distances and the relevant angles. These probabilities are then fitted to the Boltzmann factors that involve the effective empirical potentials. Srinivas et al.[25] find that the results depend on whether the cellulosic material is amorphous (denoted as AM) or crystalline.

In the latter case, they also depend on the chain type: origin (OR) or center (CE) celluloses. The names relate to taking different chains from the monoclinic unit cell: OR is from the origin (corner) of the cell and CE is from its center. In the approach of Fan and Maranas,[26] one assumes that the CG parameters are the same for OR, CE, and AM, but Srinivas et al. find variations between the forms. Both teams used the BI method.

Here, to compare with the cellulose, we consider three different hexaoses and construct CG models for them. Their structures, together with a schematic CG representation, are shown in Figure 1. We first ask how the resulting effective parameters differ from those obtained for the cellulose and to what extent they are affected by the variants of stereochemistry that differentiate between the hexaoses. Specifically, we consider cellohexaose, mannohexaose, and amylohexaose. They all contain six monosaccharide units with identical chemical composition. Each unit is a pyranose but the D-GLC units in cellohexaose are connected by the β (1 → 4) glycosidic bonds whereas in amylohexaose by the α (1 → 4) bonds. Thus, amylohexaose has an overall helical shape while cellohexaose stays linear. On the other hand, in mannohexaose the units are connected like in cellohexaose, but the mannoses (D-MAN) are stereoisomers (or epimers) of D-GLC: on the C2 position the OH group in D-MAN points in the opposite direction than in D-GLC. We shall show that the

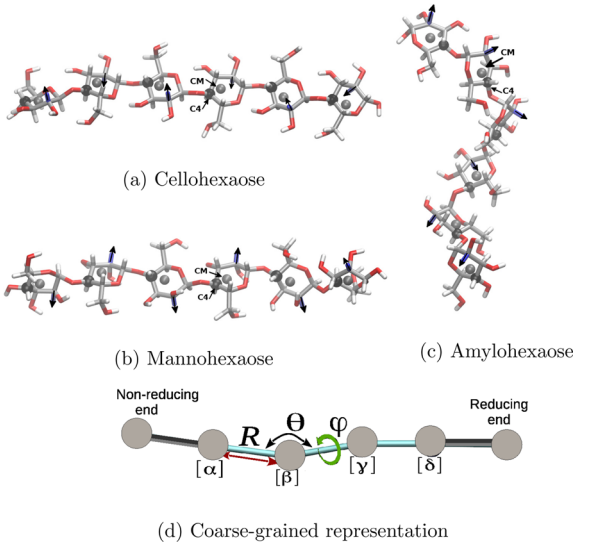


Figure 1. Molecular dynamics snapshots of polysaccharides studied: (a) cellohexaose, (b) mannohexaose, and (c) amylohexaose. The red, gray, and white colors correspond to the atoms of O, C, and H, respectively. The black arrows show the orientation of the OH bonds at the C2 positions. The orientations in mannohexaose are opposite to those in cellohexaose. The orientations in amylohexaose follow the contour of the helix. Panel (d) illustrates the CG description in which we replace one monomer by one effective atom. The atom can be placed either at the CM or at the locations of the C4 carbon atoms. The CG variables are shown for the four consecutive effective particles, labeled, respectively, by α, β, γ, and δ in the central part of the chain.

stereochemistry does affect the parameters in the CG potentials.

In order to determine the parameters of the CG model, we use two methods: the BI and by an explicit determination of the mean energy as a function of its defining quantities: r, θ, and ϕ. We call this second method "energy based" (EB). We also consider two choices for the locations: at the CM of a monomer and at atom C4 on the sugar ring. C4 is selected due to its proximity to the CM. The C4-based representation is more convenient when analyzing structure fluctuations and more in tune with associating the degrees of freedom with the α-C atoms in proteins.

We find that the BI and EB methods are consistent with one another and thus provide their mutual checks. However, the EB method comes with larger error bars and sometimes is hard to use. We find that switching from the CM to the C4 atoms brings a greater precision to the parameters. Other choices of the representative atom, like C1, reduce this precision.

At this stage, we switch the discussion from single hexaoses to systems of interacting polysaccharides. The interactions are through nonbonded couplings, such as HB's. We represent them by the Lennard-Jones potentials with the length parameter σ and the depth of ϵ. One could determine these parameters by considering hexaose dimers, but in order to reduce the noise we extract them from simulations of the crystalline cellulose Iβ. The additional advantage of this approach is that one can compare the CG parameters for the nonbonded interactions within the cellulose sheet and between the sheets. We also compare the derived couplings and elastic parameters to similar quantities in proteins. The coupling strengths are found to be consistent with those used

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

in the structure-based models of proteins.[21−23] Parameters for bonded interactions between two, three, and four α-C atoms were derived using the BI and EB methods. Our set of parameters, especially the k_θ, capture the larger rigidity of α-helices over β-strands reported by Best et al.[30]

Finally, we consider the hexaose-Man5B complex. Protein Man5B is a glycoside hydrolase so it is interesting to find a CG representation for such complexes. We derive effective nonbonded parameters for the sugar−protein interactions and then consider equilibrium root-mean-square fluctuations (RMSF) in the hexaose-Man5B complexes. We show that our CG model leads to results which are consistent with the all-atom results.[10] It distinguishes between various hexaoses and allows for studies at much longer time scales.

II. METHODS

A. All Atom Simulation. The simulations were conducted with version 2.9 of the NAMD molecular dynamics simulation package.[31] The polysaccharides were parametrized by GLYCAM-06 force field.[32,33] The implicit solvent was used for cellohexaose and mannohexaose, whereas amylohexaose required being solvated with 10 000 TIP3P water molecules[29] to avoid instabilities. The proteins were parametrized by the Amber force field.[34,35] The solvation box for the β-hairpin of protein G (PDB: 1GB1) enclosed 14 630 TIP3P water molecules, 13 560 for the tryptophan cage (Trp-cage; PDB: 1L2Y), and for the hexaose−Man5B protein complex (the structure file of Man5B is PDB: 3W0K) it is 23 340. The docking of the hexaose ligands into the binding site in Man5B was performed using Autodock Vina software.[36] The periodic boundary conditions were used to reduce the problem of the finite size effects. Numerical integration of Newton's equations of motion involved the time step of 1 fs and the atomic coordinates were saved every 1 ps for analysis. The system equilibration was carried out in the following way: first 1000 steps of energy minimization were applied and then a short 0.5 ns run in the NPT ensemble was implemented to achieve the atmospheric pressure of 1 bar. For proteins, some restraints on the protein backbone were imposed to stay near the native structure during the NPT step. The production runs for the polysaccharides and proteins were carried out in the NVT ensemble for 40 ns at T = 300 K. The temperature was controlled by the standard Langevin algorithm and the pressure by the Langevin piston pressure control algorithm. The MDenergy plugin from the VMD package[37] was used to compute the contributions of bonded and nonbonded energies. In our simulations of the cellulose Iβ allomorph, we considered 36 chains of 80 monomers each, as in ref 25 and the simulations lasted for 20 ns.

B. Boltzmann Inversion Method. The BI method allows for determination of parameters in a CG model by focusing on some degrees of freedom, q's, such as the distance between the effective atoms or the bond angles formed by three sequentially consecutive effective atoms. The assumption is that in the canonical ensemble corresponding to temperature, T, independent degrees of freedom obey the Boltzmann distribution $P(q) = Z^{-1}\, e^{-U(q)/k_BT}$. Here, $Z = \int e^{-U(q)/k_BT}\, dq$ is the partition function and k_B the Boltzmann constant. $P(q)$ can be determined through the atomistic simulation of the reference system. Once this is done, one can derive the corresponding effective potential $U(q)$, also known as the potential of the mean force, through the inversion $U(q) = -k_BT \ln P(q)$. (Note that Z enters $U(q)$ only as an additive constant.) Srinivas et al.[25] have used an iterative version of the BI method[38,39] in which one derives $P^{(0)}_{CG}(q)$ through simulations in the CG system by assuming a starting effective potential $U^{(0)}(q)$ (typically Lennard-Jones) and then adjusting it iteratively through $U^{(n+1)} = U^{(n)} + k_BT \ln(P^{(n)}_{CG}(q)/P(q))$ until the CG distribution matches the atomistic $P(q)$. Our calculations employ the simple BI method.

C. Energy-Based Approach for Calculation of Effective Bonded Interactions. An alternative method proposed here is to fit the mean atomistic energies to the functional dependence on q as postulated in the CG model. This alternative approach serves as a verification of the results obtained by the BI method and does not assume that the variables q are truly independent. The first example is the effective bond potential, $V_b^{\alpha\beta}$ which is defined between two effective atoms α and β that are placed either on specific atoms, like C4, or at the CM of a molecular unit in a biopolymer. The atoms are separated by a time-dependent distance $r^{\alpha,\beta} = |\mathbf{R}_\alpha - \mathbf{R}_\beta|$ which, generically, will be denoted as r. We assume that

$$V_b^{\alpha\beta}(\mathrm{R}_\alpha, \mathrm{R}_\beta | k_r, r_0) = \frac{1}{2} k_r (r^{\alpha\beta} - r_0^{\alpha\beta})^2 \tag{1}$$

where k_r is the spring constant and $r_0^{\alpha\beta}$ is the equilibrium length of the bond. These two parameters can be determined by evolving the atomistic system and monitoring its total energies, E, that correspond to narrowly defined bins in the values of r. These energies are expected to be distributed in the Gaussian fashion. We plot the mean value $\langle E \rangle$ of the E's obtained within specific bins against r as illustrated in the top left panel of Figure 2 for cellohexaose. We find that the dependence is indeed parabolic and determine the corresponding parameters. The data obtained for the third (central) unit in cellohexaose are expected to be the most reliable, but in order to estimate the error bars we perform the calculations also for the two bonds just off-center.

The CG effective bond angle potential involves three consecutive atoms denoted here as α, β, and γ. It is represented as

$$V_\theta^{\alpha\beta\gamma}(\mathrm{R}_\alpha, \mathrm{R}_\beta, \mathrm{R}_\gamma | k_\theta, \theta_0) = \mathrm{V}_b^{\alpha\beta} + \mathrm{V}_b^{\beta\gamma} + \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{2}$$

where

$$\cos(\theta) = \frac{r^{\alpha\beta} \cdot r^{\beta\gamma}}{|r^{\alpha\beta}||r^{\beta\gamma}|}$$

is the angle between the three molecules (see Figure 1d). The first two terms on the right-hand side of eq 2 are the effective bond potentials for molecules (α and β) and (β and γ). The last term in this equation is the effective bond angle potential which is typically represented by the harmonic potential. The determination of the force bending constant (k_θ) and the equilibrium angle (θ_0) is similar to the determination of k_r and r_0 except that now the three body energies are monitored and the terms $V_b^{\alpha\beta}$ and $V_b^{\beta\gamma}$ are subtracted to get E. The procedure is illustrated in the middle panels of Figure 2 for cellohexaose. The error bars are determined by considering the three choices of the consecutive effective atoms: 1−2−3, 2−3−4, and 3−4−5.

In a similar way, the effective torsion potential can be described by

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

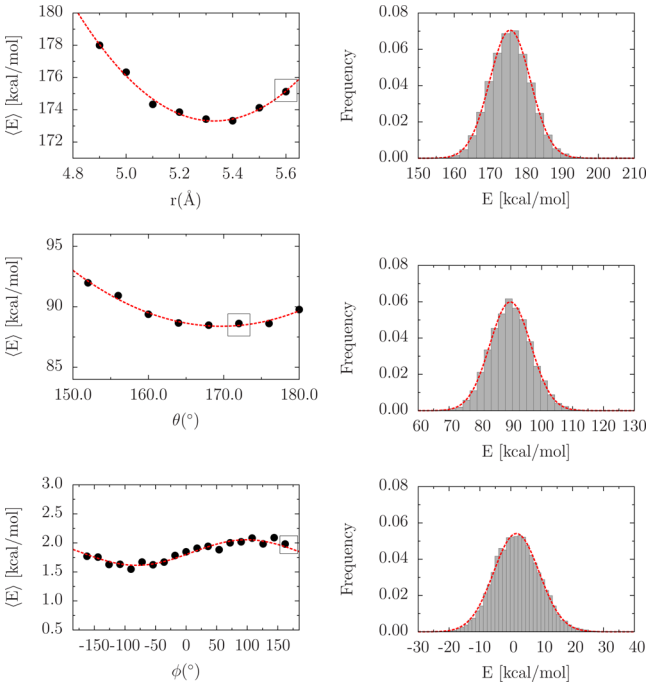


Figure 2. Left panels show the effective potentials computed by the EB method for cellohexaose at T = 300 K using the all-atom implicit solvent simulation. The lines correspond to the parameters listed in Table 1. The right panels show the atomistic energy distributions corresponding to the data points surrounded by squares shown in the panels on the left. The top panel corresponds to the two-body bond potential. The middle panels correspond to the effective three-body interaction describing the bond angle potential. The bottom panels correspond to the four-body interaction describing the dihedral terms.

$$V_{\phi}^{\alpha\beta\gamma\delta}(\mathrm{R}_{\alpha}, \mathrm{R}_{\beta}, \mathrm{R}_{\gamma}, \mathrm{R}_{\delta}|\overline{a}, ...) = V_{b}^{\alpha\beta} + V_{b}^{\beta\gamma} + V_{b}^{\gamma\delta} + V_{\theta}^{\alpha\beta\gamma} + V_{\theta}^{\beta\gamma\delta} + f(\phi) \quad (3)$$

where ϕ represents the torsion angle between the α, β, γ, and δ atoms (see Figure 1d). In order to get the needed E's, we first subtract all of the two-and three-body potentials and then determine the distributions of E's within bins corresponding to ϕ. We consider two functional forms of $f(\phi)$

$$f(\phi) = K_1[1 - \cos(\phi - \phi_0)] \quad (4)$$

and

$$f(\phi) = \frac{1}{2}k_{\phi}(\phi - \phi_0)^2 \quad (5)$$

depending on what dependence is found in a particular system. The cosine form is necessary for cellohexaose (see the bottom panel of Figure 2), mannohexaose (see Figure S1 in Supporting Information − SI), and in AM cellulose, as reported by Srinivas et al.,[25] while the parabolic form works for amylohexaose and the crystalline cellulose. Near the minimum, the cosine form becomes parabolic with $k_{\phi} = K_1$, so we use k_{ϕ} to make comparisons. Since the statistics of the four-body terms in a six-unit chain are small, we determine k_{ϕ} and ϕ_0 through simulations of eight-unit chains.

The nonbonded interactions (the HB's and ionic bridges) are represented by the Lennard-Jones potentials with the depth of the potential well ϵ and the length parameter σ. For small deviations away from the equilibrium this potential can be represented by an effective harmonic term with k_{nb} such that $\epsilon = k_{nb}(\sigma^{eff})^2 36^{-1}(2^{-2/3})$ and $\sigma = 2^{-1/6}\, r_0$. The parameters are obtained in analogy to the procedure for the bond potential: one gets k_{nb} and r_0 by first fitting to the harmonic potential near the minimum of the mean force, but then one infers about the ϵ from k_{nb}.

D. Principal Component Analysis. The molecular dynamics data are often too noisy for a direct identification of correlated motions that are important for biology. The principal component analysis (PCA) is a technique[40−42] that allows to do so in a simple manner. The PCA protocol is applied to an atomistic trajectory of N particles, $\mathbf{r}(t) = (x_1(t), y_1(t), ..., z_N(t))^T$, where T denotes transposition. First, one constructs the covariance matrix $\mathbf{C}$ of $\mathbf{r}(t)$ defined by

$$\mathbf{C} = \langle(\mathbf{r}(t) - \langle\mathbf{r}(t)\rangle_{t_{tot}})(\mathbf{r}(t) - \langle\mathbf{r}(t)\rangle_{t_{tot}})^T\rangle_{t_{tot}} \quad (6)$$

where $\langle\rangle_{t_{tot}}$ denotes the time average over $0 \le t \le t_{tot}$. Then the symmetric $3N \times 3N$ matrix $\mathbf{C}$ is diagonalized with an orthonormal transformation matrix $\mathbf{R}$ such that

$$\mathbf{R}^T\mathrm{CR} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_{3N}) \quad (7)$$

The columns of $\mathbf{R}$ are the eigenvectors or principal modes. The eigenvalues λ_m are equal to the variance in the direction of the corresponding eigenvector. They can be organized so that $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{3N} \ge 0$. The original trajectory, $\mathbf{r}(t)$, can be projected onto the eigenvector to give the principal components $p_i(t)$, i = 1, ..., $3N$ as follows:

$$\mathbf{p} = \mathbf{R}^T(\mathbf{r}(t) - \langle\mathbf{r}\rangle) \quad (8)$$

Note that $p_1(t)$ represents the first principal component with the largest mean square fluctuation, i.e., with the most dominant motion. In practice, collective movements in proteins are identified by projecting the Cartesian trajectory coordinates along several principal components.

III. RESULTS AND DISCUSSION

A. Potential Parameters for the Hexaoses. Table 1 (top) shows the potential parameters obtained for the three hexaoses at T = 300 K by the two methods and for the two choices of the locations of the effective atom. They are compared to the stiffness parameters obtained for the cellohexaoses which are connected by the intrachain HB's as shown in Figure 3. These HB's are formed typically between the hydroxyl groups ($-O_x-H_x$) such as $O_5\cdots H-O_3$ and $O_6\cdots H-O_2$. The best agreement between both methods is obtained for the C4 representation for which there is very little difference in k_r between cellohexaose and mannohexaose. This outcome makes physical sense because an axial stretching of the chain should not be affected by the direction of the OH group which is perpendicular to the ring. The CM representation does not have this feature. We suggest to use 51 kcal/mol as the common value of k_r for the two hexaoses in the C4 representation. For amylohexaose, one may take half of this value.

In the bond-angle part, however, k_θ for cellohexaose is twice as big as for the mannohexaose and there is not much difference between mannohexaose and amylohexaose. There is also very little difference in k_ϕ between cellohexaose and mannohexaose though the common value depends on the representation: one may take 0.46 kcal/mol when using the C4 one. However, for amylohexaose k_ϕ is about seven times larger. To summarize, for the C4 representation, we suggest

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

Table 1. Parameters Obtained for the CG Description of the Bonded Interactions in Three Hexaoses, the Cellohexaose with the Intrachain HB (Indicated by the HB Superscript) and Cellulose[a]

	CM				C4			
	BI		EB		BI		EB	
r	k_r [kcal/mol/Å^2]	r_0 [Å]	k_r [kcal/mol/Å^2]	r_0 [Å]	k_r [kcal/mol/Å^2]	r_0 [Å]	k_r [kcal/mol/Å^2]	r_0 [Å]
cellohexaose	41.0 ± 2.4	5.305	51.1 ± 6.2	5.331	52.0 ± 2.6	5.340	46.1 ± 6.3	5.310
mannohexaose	29.0 ± 2.2	5.299	36.3 ± 5.4	5.214	51.2 ± 1.2	5.330	45.3 ± 4.5	5.290
amylohexaose	23.4 ± 1.8	4.960	22.4 ± 3.8	4.952	29.2 ± 1.2	4.910	20.8 ± 8.7	4.930
cellohexaose^{HB}	100.8 ± 3.2	5.315	100.9 ± 7.7	5.300	85.4 ± 3.1	5.390	86.0 ± 8.2	5.390
AM cellulose	115.68	5.228*	-		-		-	
OR cellulose	219.31	5.283*	-		120.3 ± 4.3	5.248	104.2 ± 14.3	5.266
CE cellulose	368.10	5.250*	-		120.1 ± 4.1	5.252	102.1 ± 15.0	5.279
cellulose	179.92	5.237†	-					
	BI		EB		BI		EB	
θ	k_θ [kcal/mol/rad^2]	θ_0 [°]	k_θ [kcal/mol/rad^2]	θ_0 [°]	k_θ [kcal/mol/rad^2]	θ_0 [°]	k_θ [kcal/mol/rad^2]	θ_0 [°]
cellohexaose	40.1 ± 3.4	170.12	72.2 ± 32.2	170.30	50.3 ± 4.1	172.01	30.2 ± 15.0	172.70
mannohexaose	27.9 ± 4.4	172.20	38.1 ± 16.3	172.02	25.6 ± 2.1	173.10	16.3 ± 2.3	169.80
amylohexaose	17.1 ± 3.1	141.70	16.6 ± 10.4	140.34	23.8 ± 2.1	143.20	14.0 ± 3.8	137.20
cellohexaose^{HB}	40.5 ± 3.1	170.21	42.00 ± 0.2	170.6	54.3 ± 3.8	165.0	32.8 ± 1.4	164.40
AM cellulose	127.53	163.5*	-		-		-	
OR cellulose	401.52	168.7*	-		377.5 ± 4.1	167.2	359.1 ± 90.0	169.1
CE cellulose	516.25	173.2*	-		361.1 ± 3.4	166.7	281.1 ± 80.4	168.3
cellulose	212.00	175.6†	-					
	BI		EB		BI		EB		
ϕ	k_ϕ [kcal/mol/rad^2]	ϕ_0 [°]	k_ϕ [kcal/mol/rad^2]	ϕ_0 [°]	k_ϕ [kcal/mol/rad^2]	ϕ_0 [°]	k_ϕ [kcal/mol/rad^2]	ϕ_0 [°]	K_1
cellohexaose	0.20 ± 0.04	220.0	0.13 ± 0.10	280.7	0.48 ± 0.02	190	0.30 ± 0.11	267.0	C
mannohexaose	0.17 ± 0.03	190.0	0.20 ± 0.12	252.0	0.42 ± 0.04	197	0.55 ± 0.30	208.8	C
amylohexaose	3.28 ± 0.06	−52.0	6.60 ± 0.23	−54.10	3.12 ± 0.08	−50.0	3.85 ± 0.21	−50.8	
cellohexaose^{HB}	0.36 ± 0.04	195.0	0.31 ± 0.12	195.6	0.90 ± 0.03	200	0.80 ± 0.14	193.0	
AM cellulose	2.68	224.0*	-		-		-		C
OR cellulose	11.0	191.6*	-		12.31 ± 0.10	181.0	4.02 ± 0.60	185.0	
CE cellulose	3.82	187.2*	-		12.47 ± 0.20	180.5	4.30 ± 0.40	182.1	
cellulose	0.60	180.0†	-						

[a]The values in the cells marked by the * symbol are cited after ref 25. The values in the cells marked by the † symbol are cited after ref 26—they have been obtained by using the ring center model which does not distinguish between the AM, OR, and CE forms of the cellulose. The data for the cellulose in the C4 representation have been obtained by us. The symbol "C" (for "cosine") in the last column indicates that the dihedral term is described by the cosine function and the value of K_1 is then equal to k_ϕ. .

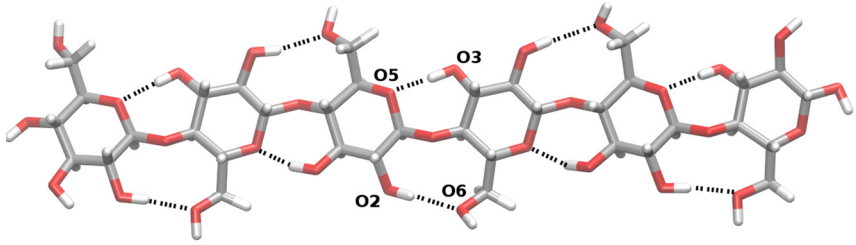


Figure 3. MD snapshot of cellohexaose chain with restraints inducing two kinds of intrachain HB's: O3−H···O5 and O2−H···O6. All such HB's are represented by dashed lines. The O and C atoms are in red and gray colors, respectively.

to use the following sets of values of parameters k_r, k_θ, and k_ϕ in kcal/mol: (a) for cellohexaose 51, 50, and 0.46; (b) for mannohexaose 51, 25, and 0.46; (c) for amylohexaose 25.5, 25, and 3.1, respectively. The system denoted as cellohexaose^{HB} in Table 1 corresponds to cellohexaose to which the intrachain HB's are added as restraints (see Table S1 in SI). The addition makes the cellohexaose stiffer by a factor of 2 when judging by the values of k_r and k_ϕ, but it leaves k_θ unchanged.

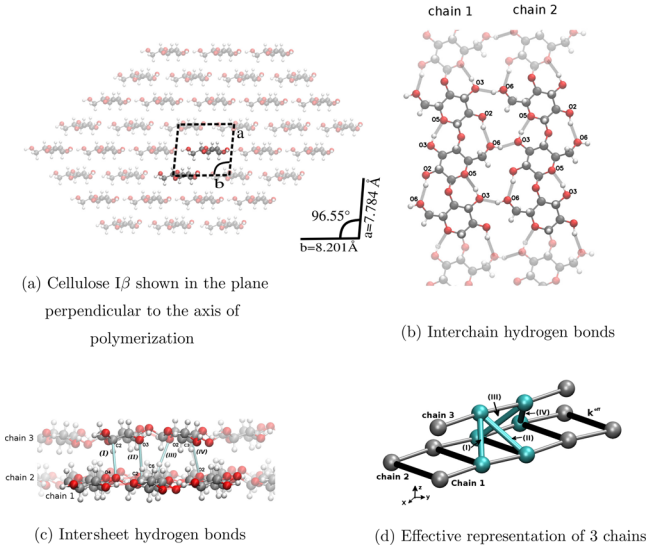
Table 1 shows the results of Srinivas et al.[25] for the cellulose Iβ fibril which gets an extra stabilization provided by the HB's between the chains. The chains form sheets connected by HB's of the O−H···O kind, but the interactions between the sheets are substantially weaker because they are coupled by the C−H···O HB. These couplings will be discussed later. It is interesting to note that k_r for cellohexaose^{HB} is nearly the same as k_r for the amorphous cellulose, but k_ϕ and k_θ are much weaker. This is a signature of the fact that the local ground-state conformations are distinct.

In ref 25, the bonded parameters were derived in the CM representation and their results for the CE, OR, and AM

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

chains differ considerably from each other as shown in Table 1. In particular, the OR and CE chains are much stiffer than AM as evidenced by all elastic constants. If judged by k_ϕ, OR is stiffer than CE, but the values of the other two elastic constants suggest otherwise. However, when we use the C4 representation, the crystalline CE and OR become quite similar elastically which is expected more physically. The EB method gives nearly the same effective k_r and k_θ as the BI method, but k_ϕ is three time smaller—one may take 8 kcal/mol/rad^2 as the compromise value of k_ϕ. The C4 representation is similar in spirit to that used by Fan and Maranas[26] since the degrees of freedom are associated with the ring centers, which is less volatile than the CM of the unit and hence the differences between the types of chains become minor.

We now focus on the nonbonded interactions between two D-GLC monomers. They may arise within a single chain in positions O3−H···O5 and O2−H···O6 as shown in Figure 3 and, as we discussed in the context of cellohexaose[HB], these two HB's are effectively included in the value of the bonded parameter k_r, as they mainly restrain the axial elongation between 2 D-GLC monomers. Other important HB's that are present in cellulose Iβ are the interchain ones (see Figure 4b



(a) Cellulose Iβ shown in the plane perpendicular to the axis of polymerization

(b) Interchain hydrogen bonds

(c) Intersheet hydrogen bonds

(d) Effective representation of 3 chains

Figure 4. Panel (a) shows a view of the 36-chain microfibril model of the cellulose Iβ allomorph in the plane perpendicular to the axis of polymerization. Axes a and b define the monoclinic unit cell formed by two chains: at the origin (OR) and center (CE). Panel (b) shows two D-GLC chains belonging to a sheet. The O3−H···O5 and O2−H···O6 intrachain HB's occur between monomers within one chain and O6−H···O3 interchain HB's are responsible for keeping the chains together. Panel (c) shows typical intersheet HB's in cellulose Iβ; four types of intersheet HB's are drawn by the faint cyan lines and enumerated by the Roman numbers. Panel (d) shows the CG representation of a three-chain subsystem. The thick black lines represent the effective interchain interactions and the cyan lines the intersheet ones.

which shows two chains within a sheet of cellulose) and the intersheet bonds (see Figure 4c) which participate in the structural stabilization of the fiber. For cellulose Iβ, there are two nonbonded energy scales:[43] $\epsilon_{\text{inter-chain}}$ that represents the potential well for the effective interchain interaction within the planar sheets (mostly due to the O6−H···O3 HB), shown in Figure 4b, and $\epsilon_{\text{inter-sheet}}$ for the intersheet interaction, shown in Figure 4c. The combined view of the existing couplings is shown in Figure 4d. The latter energy scale includes HB of the type C−H···O, such as C_2−H···O_4 (Type-I), C_2−H···O_3 (Type-II), C_6−H···O_2 (Type-III), and C_3−H_3···O_2 (Type-IV).

Table 2 lists the nonbonded effective Lennard-Jones parameters ϵ^{eff} and σ^{eff} for cellulose Iβ in water at T = 300 K. They have been obtained for both CM and C4 representations and by the two methods (EB and BI). Fan and Maranas[26] have mapped the nonbonded interaction to the Morse potential with the depth set to 5 kcal/mol, which is representative of a moderate O−H···O HB energy strength in solids.[44] They have considered the radial distribution function (RDF) for the ring centers. The first four peaks in the RDF correspond to the average distances of (1) 5.90, (2) 6.68, (3) 7.69, and (4) 8.32 Å which are meant to correspond to various HB's. The first two peaks correspond to the HB's of types I and III of Figure 4d, respectively. The last peak is for the HB's between the two parallel chains of Figure 4b. However, the third peak corresponds to two chains separated by the lattice constant a, as shown in Figure 4a, which are not connected by any HB. Nevertheless, they use the BI method involving the Gaussian fitting of the widths of the four peaks to derive the curvatures of the HB potentials even for the third peak.

This BI-based procedure certainly overestimates the strength of the intersheet HB's of type C−H···O, which are known to be weaker.[45] On the other hand, Srinivas et al.[25] have obtained the nonbonded potentials in a tabulated (i.e., not analytic) form by using iterative BI methods[38,39] applied to the OR-OR, CE-CE, and OR-CE radial distribution functions. Our nonbonded energy values correspond to HB's inside the cellulose material and not exactly for the hexaoses. Still, they provide an insight into the relationship between the energy scales for the two types of HB's.

To summarize, we suggest taking 7.4 kcal/mol for ϵ^{eff} corresponding to the interaction between parallel D-GLC chains and 2.4 kcal/mol for the HB interactions between the sheets, independent of the CM or C4 representation (CM or C4). For these nonbonded couplings, the BI method works much worse than EB, mainly because it does not take into account the correlation between atoms such as in the pair correlation function. Moreover, the BI method gives only the optimal solution in the limit of a highly diluted system, and clearly some limitation of this technique must arise when dealing with crystalline systems. Thus, this method leads to a factor-of-6 overestimation of the energy parameters, when it is compared with typical HB's in solids.[44,45]

B. Potential Parameters for Proteins. We now use the same BI and EB methodology to derive the stiffness and HB parameters for proteins to gain insight into the energy scales relative to the polysccharides. The degrees of freedom are taken to be associated with the α-C atoms. Unlike the polysaccharides, proteins are atomically inhomogeneous so we focus on parameters for simple secondary structures. We consider two peptides shown in Figure 5: the 16-residue β-hairpin from protein G studied in refs 46−48 and the 20-residue Trp-cage, in which the (1−9) segment is an α-helix and the (11−13) segment is a short 3−10 helix.[49] We also study protein Man5B. It comprises 330 residues and its structure has been solved by means of the X-ray crystallography at 1.60 resolution.[50] Its set of secondary structures consist of 11 α-helices and 12 β-strands. For our

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

Table 2. Effective Nonbonded Parameters between D-GLC Monomers in Cellulose Iβ Computed in the CM and C4 Representation

	CM				C4					
	BI		EB		BI		EB			
nonbonded	k_r [kcal/mol/Å^2]	ϵ^{eff} [Å]	k_r [kcal/mol/Å^2]	ϵ^{eff} [kcal/mol]	k_r [kcal/mol/Å^2]	ϵ^{eff} [Å]	k_r [kcal/mol/Å^2]	ϵ^{eff} [kcal/mol]	σ^{eff} [Å]	HB-type
interchain OR	48.420	46.90	7.680	7.421	48.130	46.62	7.660	7.410	7.440	O6−H···O3
interchain CE	50.860	46.13	7.740	7.500	50.221	48.43	7.714	7.472	7.424	O6−H···O3
intersheet I	28.050	14.26	4.004	2.042	27.940	14.16	3.904	1.960	5.381	C2−H···O4
intersheet II	30.120	15.17	5.050	2.600	30.002	15.17	5.103	2.530	5.376	C2−H···O3
intersheet III	26.720	17.02	3.752	2.400	28.101	17.91	3.803	2.420	6.034	C6−H···O2
intersheet IV	28.270	17.70	4.220	2.640	29.403	18.31	4.200	2.630	5.980	C3−H···O2

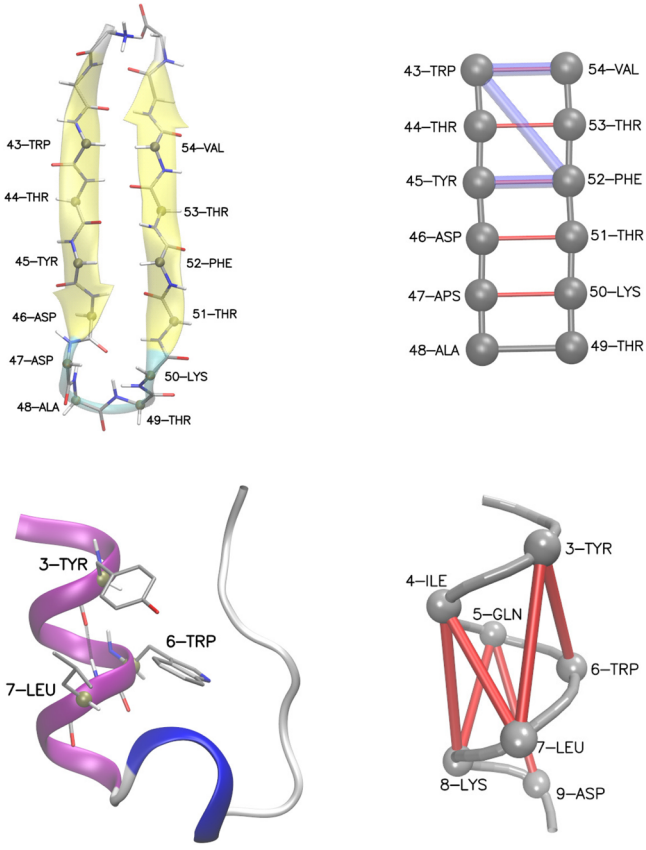


Figure 5. Top left panel shows the β-hairpin fragment of the 16 amino acid residues from the protein G. Top right panel shows the corresponding CG representation based on the positions of the α-C atoms. This hairpin possesses two kinds of the native contacts between the α-C atoms: five HB's (or hydrophilic−hydrophilic) depicted in red and three hydrophobic (or hydrophobic−hydrophobic) depicted in blue. The bottom left panel shows the Trp-cage. 3-TYR forms two HB's with 6-TRP and 7-LEU as highlighted. The bottom right panel shows the CG representation of the helical part of the protein together with the contacts (in red).

studies, we select the (129−164) segment which incorporates one α-helix, one β-strand, and one coil.

1. Effective Bonded Interactions in Peptides. Figure 6 shows the sequential dependence of the bonded parameters for the segment in the Man5B protein. Table 3 shows the average values of these parameters for several examples of the secondary structures and loops—the average is over the sites in the structures. Both the BI and EB methods give nearly the same results for the k_r parameter −200 kcal/mol/Å^2, independent of the location in the sequence. However, the angular elastic constants depend on the type of the secondary structure. On average, the α-helix is found to be stiffer than the β-strand. This is consistent with the finding of Best et al.[30] However, we observe that the difference is captured by the factor of 2 instead of 4, independent of the method used.

A common CG description of the dihedral part of the peptide bond has been inspired by all atom force fields where dihedral angles from the protein backbone are modeled by a Fourier expansion such as

$$\sum_{\text{dih}} \frac{V_n}{2}[1 + \cos(n\phi - \delta_n)]$$

Here, V_n denotes the height of the energy barrier, δ_n is the equilibrium dihedral angle, and n indicates the type of symmetry around the dihedral angle—typically, it corresponds to a threefold periodicity. It seems natural to use the same functional form to describe the effective dihedral CG potential. At a first approximation, the distances and angles behave nearly quadratically around their equilibrium values and the harmonic form for this potential can be used. However, the effective dihedral angles show a larger flexibility across the dihedral space than in all-atom simulation. This has been demonstrated, for instance, in ref 52 through an exhaustive exploration by means of all-atom replica exchange molecular dynamics of the atomistic structure of the β-hairpin fragment in a silk fiber. The overall simulation time reached 0.5 μs, and a good conformation sampling of dihedral space was obtained. The effective dihedral potential was parametrized by $V(\phi) = \sum_{\text{dih}} A + B\cos(\phi) + C\cos^2(\phi) + D\cos^3(\phi)$, with a proper set of parameters A, B, C, and D derived by the BI method. The set of the parameters is specific for amino acids in the β-strands and in the turn.

Another parametrization of dihedral potential has been proposed by Clementi et al.:[18] $V(\phi) = \sum_{\text{dih}} K_\phi^1[1 - \cos(\phi - \phi_0^n)] + K_\phi^3[1 - \cos(3(\phi - \phi_0^n))]$. By construction, this potential facilitates the search for native dihedral angles (ϕ_0^n), but it also allows for intermittent excursions to dihedral angles associated with high energy situations. In our studies we have not observed any of the above-mentioned functional forms for dihedral potential in the α-C representation. Instead, the harmonic potential around the equilibrium (or native) dihedral angle appears to describe the dynamics in the folded state in an adequate way. As we see in Table 3, the average bonded parameters for the peptide segments agree between the methods, within the error bars, and only in the case of α-helices in Trp-cage and Man5B does the BI method give force constants which are twice as big as the EB method. The

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

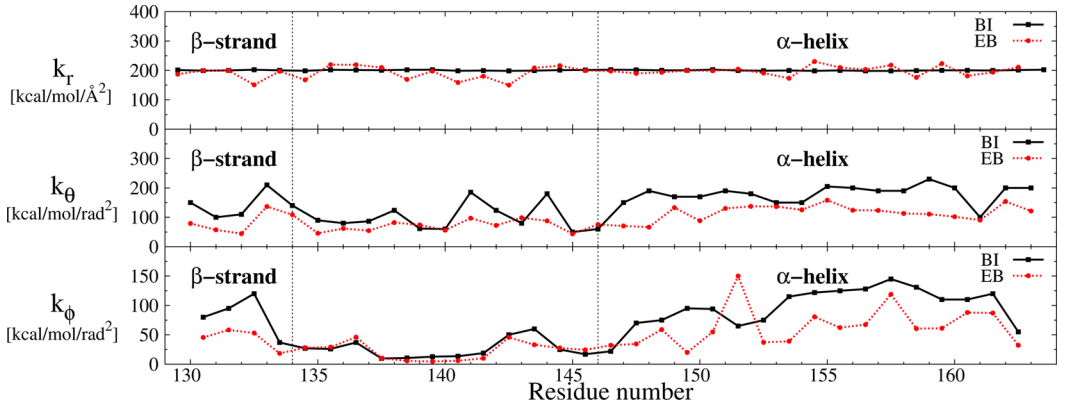


Figure 6. Sequential dependence of the effective bonded parameters k_r, k_θ, and k_ϕ for segment (129−164) in the Man5B. The elastic constants are associated with the smallest site index of the two consecutive residues involved. In this protein fragment is found a β-strand which is connected to an α-helix by a coil. Black solid and dashed red lines correspond to the BI and EB methods, respectively.

Table 3. Average Bonded Parameters for the Indicated Secondary Structures, Trp-Cage and Man5B Loops[a]

	α-C			
	BI		EB	
	k_r [kcal/mol/Å²]	r_0 [Å]	k_r [kcal/molÅ²]	r_0 [Å]
α-helix (Trp-cage)	207.0 ± 2.2	3.870	205.1 ± 8.2	3.873
α-helix (Man5B)	200.1 ± 1.3	3.871	199.1 ± 15.4	3.870
β-strand (Man5B)	200.4 ± 1.3	3.862	184.0 ± 18.3	3.864
β-hairpin	201.4 ± 8.3	3.860	190.1 ± 13.2	3.856
Loop (Trp-cage)	203.5 ± 2.1	− −	193.10 ± 17.4	− −
Loop (MAN5B)	200.5 ± 1.4	− −	187.10 ± 23.7	− −
	BI		EB	
	k_θ [kcal/mol/rad²]	θ_0 [°]	k_θ [kcal/mol/rad²]	θ_0 [°]
α-helix (Trp-cage)	231.5 ± 9.4	89.1	110.0 ± 7.2	88.5
α-helix (Man5B)	182.2 ± 29.0	91.1	119.6 ± 23.4	92.5
β-strand (Man5B)	142.0 ± 43.2	116.8	79.4 ± 35.3	118.1
β-hairpin	58.0 ± 6.2	131.3	25.8 ± 5.4	130.1
Loop (Trp-cage)	123.50 ± 20.2	− −	81.22 ± 32.3	− −
Loop (MAN5B)	107.10 ± 34.3	− −	73.20 ± 12.1	− −
	BI		EB	
	k_ϕ [kcal/mol/rad²]	ϕ_0 [°]	k_ϕ [kcal/mol/rad²]	ϕ_0 [°]
α-helix (Trp-cage)	90.6 ± 5.8	51.4	29.2 ± 7.2	50.5
α-helix (Man5B)	110.4 ± 22.0	49.3	64.7 ± 25.4	50.13
β-strand (Man5B)	83 ± 16.5	−148.1	52.3 ± 5.3	−146.7
β-hairpin	26.3 ± 19.2	−152.8	23.9 ± 9.8	−148.4
Loop (Trp-cage)	28.3 ± 2.1	− −	20.4 ± 7.3	− −
Loop (Man5B)	13.1 ± 4.2	− −	7.3 ± 3.4	− −

[a]Parameters are obtained by the BI and EB methods.

parameters obtained for the loops by the two methods are consistent with one another within the error bars. They can be used to properly describe the stiffness throughout the protein as the loops are regions that do not get extra stabilization through contact interactions. Figure 7 shows the

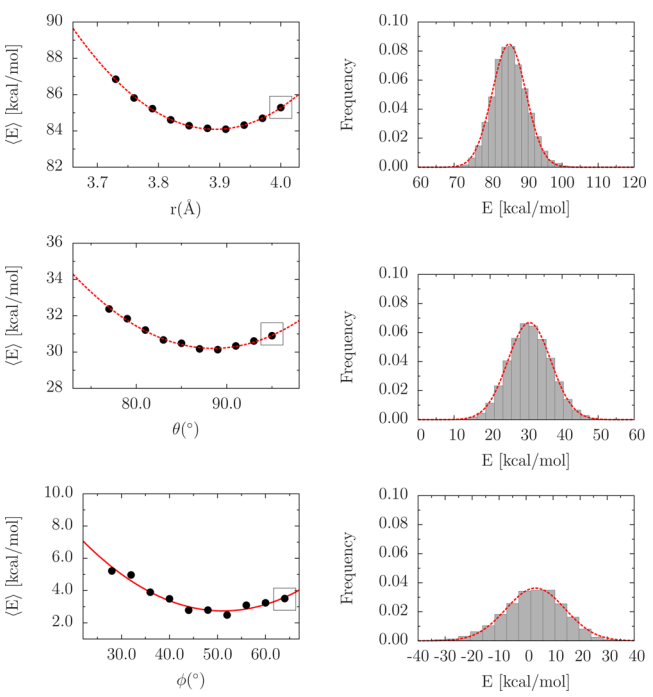


Figure 7. Left panels show the effective potentials computed by the EB method for the segment (3−9) in Trp-cage at T = 300 K using the all-atom explicit solvent simulation. The lines correspond to the parameters listed in Table 1. The right panels show the atomistic energy distributions corresponding to the data points surrounded by squares shown in the panels on the left. The top panel correspond to the two-body bond potential. The middle panels correspond to the effective three-body interaction describing the bond angle potential. The bottom panels correspond to the four-body interaction describing the dihedral terms.

effective quadratic energy profile for bonds, bond angles, and dihedral angles computed by the EB method in Trp-cage. Note that the harmonic description for dihedrals captures deviations from the native dihedral angle. Similar plots were obtained for the β-hairpin and Man5B proteins (data not shown).

2. Effective Nonbonded Interaction in Peptides. In the α-helix (1−9) segment of Trp-cage there are 11 native contacts

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

Table 4. Effective Nonbonded Lennard-Jones Parameters Computed for the Native Contacts in Trp-Cage and β-Hairpin[a]

Native Contact		α-C BI k_r [kcal/mol/Å^2]	α-C BI ϵ^{eff} [kcal/mol]	α-C EB k_r [kcal/mol/Å^2]	α-C EB ϵ^{eff} [kcal/mol]	σ^{eff} [Å]	nature of contact
Trp-cage (α-helix)							
3-TRY	6-TRP	5.102	1.740	2.234	0.762	4.413	HB
3-TYR	7-LEU	9.63	6.060	4.722	2.972	5.996	HB,HP
4-ILE	7-LEU	5.230	1.802	2.330	0.803	4.437	HB,HP
4-ILE	8-LYS	9.10	5.680	4.430	2.764	5.970	HB,HP
5-GLN	8-LYS	4.020	1.252	1.774	0.553	4.220	HB,HP
5-GLN	9-ASP	7.10	3.930	3.130	1.732	5.623	HB
β-hairpin							
43-TRP	54-VAL	9.82	2.803	5.510	1.573	4.039	HP
44-THR	53-THR	7.72	4.000	3.120	1.616	5.440	HP
45-TYR	52-PHE	10.22	2.914	4.580	1.307	4.038	HB
46-ASP	51-THR	7.40	3.360	2.860	1.300	5.092	HP
47-ASP	50-LYS	8.74	2.360	3.800	1.025	3.925	HB
43-TRP	52-PHE	5.42	3.660	2.560	1.730	6.212	HP

[a]The chemical nature of the contact (the last column) is determined with the CSU method.[51] The hydrophobic−hydrophobic contacts are denoted as HP and the hydrophilic−hydrophilic as HB, corresponding to putative hydrogen bonds.

between residues (i,j) where j is equal to i+3 or i+4. The contacts correspond to HB's between the hydrogen group acceptor O═C(i) and the group donor N−H(j). We focus only on the 6 contacts in the (3−9) segment as these are the most stable during all atom simulations. In the β-hairpin we also study 6 of the most stable native contacts in the segment (43−54). These contacts are listed in Table 4 together with the obtained effective energy parameters. The EB method yields smaller values of ϵ^{eff} than BI—by a factor of 2. However, the EB values are consistent with the theoretically derived energy scales for the N−H···O═C HB in aqueous peptides: on the order of ∼1.58 kcal/mol in a β-sheet and 1.93 kcal/mol in an α-helix.[53] They are also consistent with experimental results.[54] Comparable energy scales in HB's have been used in structure-based CG models of proteins.[18,23,24] Note that the BI method has been found to overestimate the strength of nonbonded interactions also for the cellulose Iβ—the method assumes statistical factorization of the degrees of freedom.

Note that contact energy analysis shows a clear heterogeneity in the parameter ϵ^{eff}, which can be attributed to the specific nature of the contact. If we take an average over the contacts in the β-hairpin, we get 1.6 kcal/mol for the hydrophobic−hydrophobic (HP) case and 1.2 kcal/mol for the hydrophilic−hydrophilic (HB) one. In the helical segment of the Trp-cage, two well-separated effective energy scales coexist. The larger (2.5 kcal/mol on average) is associated with the pairs of type i,i+4 and the smaller with i,i+3 (0.7 kcal/mol on average).

In order to estimate a single characteristic energy scale for each peptide we take the average over all native contacts. In this way, for β-hairpin the nonbonded energy scale is about 1.43 ± 0.30 kcal/mol and for the helical segment −1.60 ± 0.90 kcal/mol. Altogether, one may take 1.5 kcal/mol as the average strength of the native contacts between α-C atoms. It is interesting to note that studies of protein stretching at constant speed have yielded $\epsilon^{eff} \approx 1.6$ kcal/mol when comparing to experimental results on the mechanostability, after making extrapolations to the experimental speeds.[23]

C. Hexaose−Man5B Complex. *1. Parametrization of the Hexaose−Protein Contacts.* We now characterize the C4−α-C contacts between the catalytic pocket of Man5B and a hexaose. The identification of the contacts is made through the atomic overlap criterium.[21,22] We represent all heavy atoms in the protein and in the sugar chain by spheres. The radii of the spheres are equal to the van der Waals radii multiplied by 1.24 to account for attraction. The radii of the spheres in the protein are taken from Tsai et al.,[55] and in the sugar we use the GLYCAM-06 all-atom force field[32,33] (see the van der Waal radii in Table S2 in SI).

A contact is declared to exist if two clusters of spheres, one formed by a residue and another by the sugar monomer, overlap. In this way, we find 29 contacts in the cellohexaose-Man5B and 28 contacts in the mannohexaose-man5B complexes. These contacts and their respective nonbonded binding energies are listed in Tables 5 and 6, respectively. The contacts form networks that are represented schematically in Figure 8.

In the case of cellohexaose, the central D-GLC$_3$ makes 10 contacts and D-GLC$_4$ 7 contacts, whereas the other D-GLC monomers generate no more than 4 contacts. In the case of mannohexaose, the well connected central region expands: D-MAN$_3$ makes 7 contacts, D-MAN$_4$, 7, and D-MAN$_5$, 6, but the connectivity of the last monomer drops from 4 to 2. The differences reflect the opposite orientation of the OH group on the C2 atoms. About 50% of the sugar−protein contacts are common for both maps.

The effective binding energy of the sugar−protein contacts will be denoted as ϵ^{eff}_{SP}. They are determined by using only the EB method. For the cellohexaose−Man5B complex, it varies between 0.44 and 8.640 kcal/mol. The maximal value is for the contact between the third D-GLC monomer and HIS-83. For the mannohexaose−Man5B complex, it varies between 0.77 and 4.62 kcal/mol resulting in an overall weaker coupling of the mannohexaose compared to the cellohexaose. This finding is consistent with the large-scale all atom simulations by Benardi et al.[10] who suggest that the stronger binding of cellohexaose inhibits the enzymatic activity in Man5B complexes.

In CG simulations, it is convenient to use uniform values of ϵ^{eff}_{SP}. For mannohexaose, we propose to take the double value

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

Table 5. List of 29 Protein−Sugar Contacts in the Cellohexaose−MAN5B complex[a]

		C4 and α-C		
D-GLC$_n$	AA-residue	**EB** k_r [kcal/mol/Å^2]	σ^{eff} [Å]	ϵ_{SP}^{eff} [kcal/mol]
D-GLC$_1$	TYR-12	1.50	7.812	1.602
D-GLC$_1$	VAL-13	1.30	6.528	0.970
D-GLC$_1$	TRP-210	1.23	7.613	1.217
D-GLC$_2$	TYR-12	2.48	8.450	3.100
D-GLC$_2$	TRP-210	2.40	5.696	1.363
D-GLC$_2$	TRP-291	4.30	7.672	4.430
D-GLC$_3$	TYR-12	3.03	10.900	6.290
D-GLC$_3$	HIS-84	5.28	9.670	8.640
D-GLC$_3$	ASN-136	5.50	8.470	6.900
D-GLC$_3$	GLU-137	4.01	7.090	3.524
D-GLC$_3$	TYR-198	4.00	7.050	3.380
D-GLC$_3$	HIS-205	4.95	8.530	6.303
D-GLC$_3$	TRP-210	3.80	6.900	3.080
D-GLC$_3$	GLU-258	5.61	5.940	3.460
D-GLC$_3$	TRP-291	9.19	6.960	7.792
D-GLC$_3$	PHE-297	3.70	9.020	5.270
D-GLC$_4$	GLU-137	3.40	6.714	2.682
D-GLC$_4$	GLY-178	3.90	4.971	1.730
D-GLC$_4$	ARG-196	4.10	7.030	3.546
D-GLC$_4$	TYR-198	3.06	5.610	1.690
D-GLC$_4$	MET-201	3.80	7.760	4.002
D-GLC$_4$	HIS-205	3.10	7.930	3.314
D-GLC$_4$	GLU-258	6.20	7.290	5.766
D-GLC$_5$	GLN-199	0.92	6.782	0.740
D-GLC$_5$	MET-201	3.16	5.810	1.890
D-GLC$_6$	ASN-140	0.50	8.020	0.562
D-GLC$_6$	ASN-180	0.53	7.040	0.440
D-GLC$_6$	GLN-199	0.51	9.014	0.730
D-GLC$_6$	GLN-228	0.61	7.290	0.570

[a]The last two columns show the length and energy parameters of the corresponding Lennard-Jones potential.

Table 6. Similar to Table 5 but for the 28 Protein−Sugar Contacts in the Mannohexaose−Man5B Complex

		C4 and α-C		
D-MAN$_n$	AA-residue	**EB** k_r [kcal/mol/Å^2]	σ^{eff} [Å]	ϵ_{SP}^{eff} [kcal/mol]
D-MAN$_1$	TYR-12	1.32	9.060	1.842
D-MAN$_2$	TYR-12	2.40	8.732	3.110
D-MAN$_2$	ASN-92	3.10	6.992	2.652
D-MAN$_2$	TRP-210	2.10	5.496	1.110
D-MAN$_2$	TRP-291	3.60	8.570	4.624
D-MAN$_2$	PHE-297	0.8	8.726	1.066
D-MAN$_3$	ILE-91	3.40	6.556	2.560
D-MAN$_3$	ASN-92	2.50	6.480	1.840
D-MAN$_3$	GLU-137	2.40	7.551	2.394
D-MAN$_3$	TYR-198	1.80	7.270	1.664
D-MAN$_3$	HIS-205	3.10	8.270	3.710
D-MAN$_3$	TRP-210	1.80	6.950	1.521
D-MAN$_3$	TRP-291	2.20	8.974	3.100
D-MAN$_4$	ILE-91	2.70	8.160	3.144
D-MAN$_4$	GLU-137	1.50	7.571	1.504
D-MAN$_4$	GLY-178	5.90	4.100	1.732
D-MAN$_4$	ASN-180	4.40	6.452	3.210
D-MAN$_4$	TYR-198	4.90	6.550	3.680
D-MAN$_4$	MET-201	4.10	6.740	3.260
D-MAN$_4$	HIS-205	2.10	8.674	2.770
D-MAN$_5$	ASN-140	1.32	8.192	1.550
D-MAN$_5$	GLY-177	4.20	6.104	2.740
D-MAN$_5$	GLY-178	2.50	6.413	1.800
D-MAN$_5$	ASN-180	2.70	7.284	2.510
D-MAN$_5$	MET-201	3.30	6.342	2.322
D-MAN$_5$	TRP-211	1.50	10.663	2.984
D-MAN$_6$	GLN-228	1.60	5.670	0.900
D-MAN$_6$	GLN-227	1.54	5.312	0.770

of ϵ^{eff} derived for proteins, i.e., about 3 kcal/mol. For cellohexaose, we propose to double it again—to 6 kcal/mol.

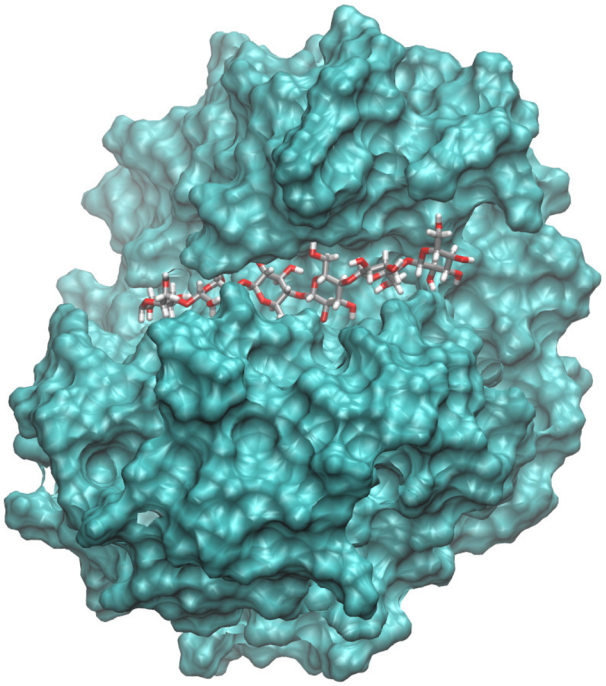
2. Coarse-Grained Studies of Hexaose−Man5B Complex. In this section we study sugar−protein interactions within our CG description. Our main purpose is to determine whether the simplified model captures the essential α-C backbone fluctuations in the presence of a sugar substrate that are related to the enzymatic activity. In sections III and IV, all needed effective parameters for the bonded interactions in sugars and proteins were derived. Here, we choose the C4 representation when dealing with sugar chains and α-C representation for proteins. Table 7 lists the simplified recommended values, discussed in section III, that we use in the CG simulations. In the C4 representation both cellohexaose and mannohexaose are described by almost the same set of bonded parameters; only one angular parameter, the k_θ, is taken to differentiate between them.

For Man5B, we take the following values of the elastic constants in the bonded interactions: k_r = 200 kcal/mol/Å^2, k_θ = 90 kcal/mol/rad^2, and k_ϕ = 10 kcal/mol/rad^2 which are defined for two consecutive α-C, three consecutive α-C, and four consecutive α-C atoms, respectively. We have chosen these parameters based on the following criterion: they should belong to a protein segment which is not stabilized by interactions with secondary motifs (i.e., α-helices or β-strands). Thus, these parameters were obtained by taking only an average over the loop segment (136−143) in Man5B. Equilibrium distance between two α-C atoms was set to 3.8 Å and the angle and dihedral angle equilibrium values were taken from the native conformation of the protein.

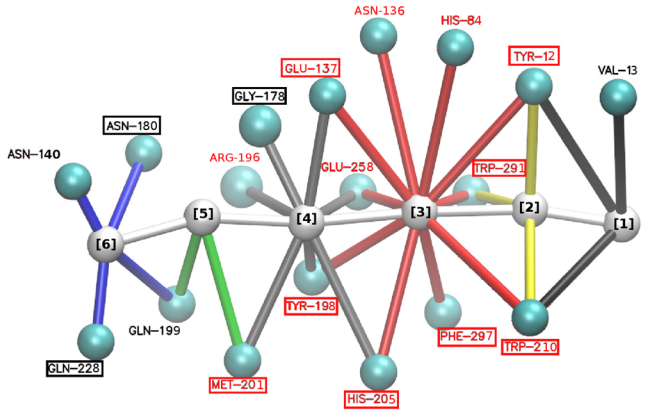
The nonbonded interactions between the α-C atoms in Man5B are included in the Gō-like fashion as described in refs 16, 22, and 23. The depth of the potential well (ϵ^{eff}) is taken to be equal to 1.5 kcal/mol. The sugar−protein interactions have been described in the previous section and we take the simplified average values stated. Thermostating of the complexes is provided by Langevin noise and damping. The T is set to 0.39 ϵ^{eff}/k_B. Each complex was simulated for 10^5 τ steps, where τ is of order 1 ns. This time scale was sufficient to achieve convergence of the RMSF in the positions of individual residues (see Figure S2 in SI). Note that the largest time scale exceeds the all-atom length of simulations by about 3 orders of magnitude.

Figure 9a shows the RMSF of the α-C atoms for the system docked with cellohexaose and mannohexaose and compares it to the situation without docking. The RMSF reveals an apparent loss of flexibility in the loop (segment 200−220), which is known to participate in the cleavage of the substrates[10] by the opening and closing motion of the Man5B catalytic pocket. This comparison cannot be decisive because the RMSF involves several large-scale motions and, in particular, the particular motion that is related to the enzymatic activity. Thus, we have performed the PCA for the undocked and docked CG trajectories. The calculation of the covariance matrix, the eigenvectors and eigenvalues were

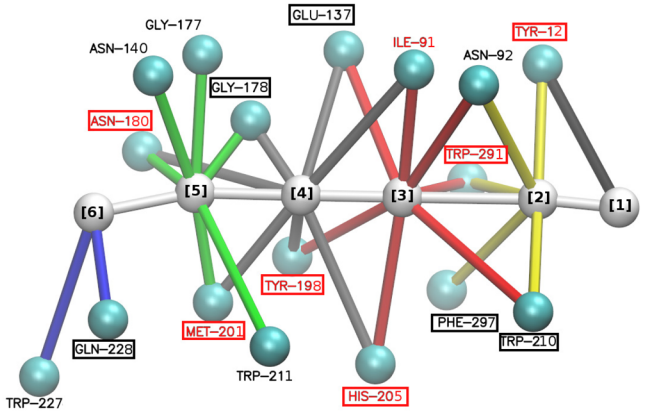
DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041



(a) Docking of cellohexaose in Man5B



(b) Cellohexaose-Man5B



(c) Mannohexaose-Man5B

Figure 8. (a) Snaphot of cellohexaose-Man5B showing that cellohexaose is totally enclosed in the tunnel-shaped active site. Man5B is represented by its molecular surface and cellohexaose in licorice representation. Graphical representation of the sugar−protein contacts in cellohexaose−Man5B (panel b) and mannohexaose−Man5B (panel c). The amino acids residues involved in the same contacts are highlighted within boxes. The red boxes show the amino acids whose interaction energies with a given sugar monomer are larger than 3 kcal/mol. The C4 atom in hexaoses is represented by white spheres connected by sticks. The square brackets show the monomer sequence index. The α-C atoms of the residues are shown in the cyan color.

Table 7. Parameters Used to Perform Our CG Simulation of the Hexaose−Man5B Complex[a]

	k_r [kcal/mol/Å^2]	r_0 [Å]	k_θ [kcal/mol/rad^2]	θ_0 [°]	k_ϕ [kcal/mol/rad^2]	ϕ_0 [°]	ϵ^{eff} [kcal/mol]	$\epsilon_{SP}^{\mathrm{eff}}$ [kcal/mol]
cellohexaose	51	5.33	50	172.40	0.46	190.0	-	6.0
mannohexaose	51	5.31	25	171.50	0.46	197.0	-	3.0
Man5B	200	3.80	90	NC	10	NC	1.5	-

[a]Columns 2−4 provide the values of the three bonded stiffness constants together with the equilibrium values of the coordinates involved. The equilibrium values of the angles and dihedral angles in Man5B are taken from the native conformation (NC). Column 5 gives the value of the contact energy in the protein. Column 6 provides the contact energies for the protein−hexaose contacts.

carried out using g_covar and g_anaeig programs from the GROMACS software package.[57] Using eq 8, we projected the Cartesian trajectory in the direction of first largest eigenvector−RMSF$_{PC}$ for each α-C atom. Figure 9b shows RMSF$_{PC}$ as a function of the sequential position. Note that the pattern for cellohexaose lies just below the one for mannohexase profile, indicating reduced enzymatic activity when Man5B is docked with cellohexaose. This result is in agreement with the all atom simulation done by Bernardi et al. (see Figure 9c).

We have noted that cellohexaose has a different pattern of contacts with Man5B than mannohexaose, but the total number of contacts is similar, 29 and 28, respectively. There are two other differences: the nonbonded coupling with cellohexaose is twice as strong as that with mannohexaose, and also the corresponding k_θ is larger. It is not easy to modify k_θ (it would require making modifications in the atomic force field and redoing the docking procedure), but it is straightforward to double $\epsilon_{SP}^{\mathrm{eff}}$ in mannohexaose. We have checked that the effect of doing so affects the fluctuations only in the catalytic region where it reduces RMSF$_{PC}$ to the

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

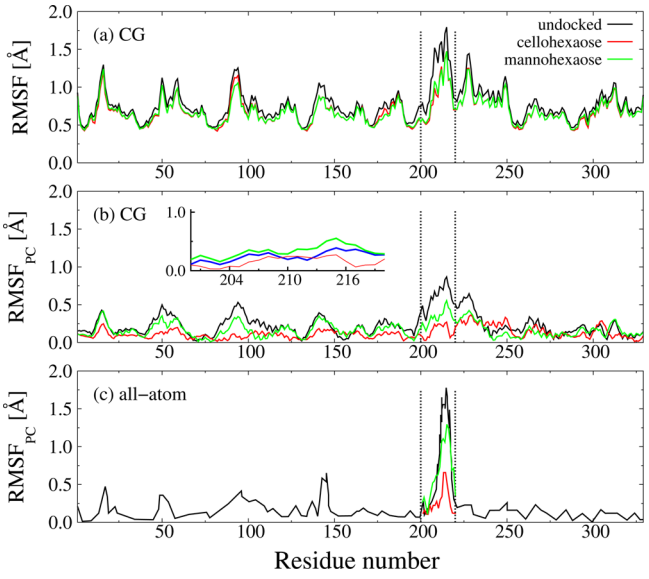


Figure 9. Panel (a) shows RMFS's of the α-C atoms of the hexaose−Man5B complexes. The black line shows RMSF obtained in the CG simulations for the protein without any substrate, i.e., for the undocked system. The red and green lines correspond to the protein docked with the cellohexaose and mannohexaose, respectively. Panel (b) shows the principal component of the RMSF, $RMSF_{PC}$, again divided into the situations. The inset zooms in on the catalytic region where the additional blue line corresponds to mannohexaose with the doubled value of ϵ_{SP}^{eff}. Panel (c) shows the analogue of panel (b), but as computed by Bernardi et al.[10] through 100-ns-long all atom simulations. (The data points were extracted from the published figure by using the g3data software.[56]) In all panels, the catalytic region is indicated by two vertical dashed lines between the 200−220 amino acids in Man5B.

level of cellohexaose, as shown in the inset of Figure 9b). This indicates that it is the pattern and, especially, the strength of the sugar−protein contacts that are decisive for setting the differences in the patterns in RMSF.

IV. CONCLUSIONS

We have constructed a CG description of the hexaose−Man5B complexes and determined all parameters that are necessary to study the dynamics of these complexes. The description involved the C4 and α-C representation for the sugars and the protein, respectively. We made the comparisons of the parameters to those obtained for cellulose Iβ. Additionally, we estimated the nonbonded energy scale in the protein contacts (1.5 kcal/mol).

Our CG simulations have shown a reduced enzymatic activity when Man5B was docked with cellohexaose, in agreement with ref 10. We explain this in terms of the different strength in the contact interactions involved in docking. Enhancement of the enzymatic activity may be obtained by making mutations in the catalytic pocket of Man5B, as these may affect the list of the docking contacts. Our model may help elucidate what these mutations should be.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b06141. Figure S1 (effective stiffness potentials for mannohexaose computed by EB method). Table S1 (parameters used to impose hydrogen bond restraints in cellohexaose). Table S2 (van der Waals radii used for protein and sugar). Figure S2 (Convergence of RMSF for undocked Man5B in coarse grained simulation) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: poma@ifpan.edu.pl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research has been supported by the ERA-NET grant ERA-IB (EIB.12.022) (FiberFuel) and the European Framework Programme VII NMP grant 604530-2 (CellulosomePlus) to M.C. It was also cofinanced by the Polish Ministry of Science and Higher Education from the resources granted for the years 2014-2017 in support of international scientific projects. A. B. Poma and M. Chwastyk were funded through a contract of the ERA-NET grant. The computer resources were financed by the European Regional Development Fund under the Operational Programme Innovative Economy NanoFun POIG.02.02.00-00-025/09.

■ REFERENCES

(1) Bayer, E. A.; Lamed, R.; Himmel, M. E. The Potential of Cellulases and Cellulosomes for Cellulosic Waste Management. *Curr. Opin. Biotechnol.* **2007**, *18*, 237−245.

(2) Bayer, E. A.; Belaich, J. P.; Shoham, Y.; Lamed, R. The Cellulosomes: Multi-Enzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annu. Rev. Microbiol.* **2004**, *58*, 521−554.

(3) Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. Computer Simulations Studies of Microcrystalline Cellulose Iβ. *Carbohydr. Res.* **2006**, *341*, 138−152.

(4) Frey-Wyssling, A. The Fine Structure of Cellulose Microfibrils. *Science* **1954**, *119*, 80−82.

(5) O'Sullivan, A. C. Cellulose: The Structure Slowly Unravels. *Cellulose* **1997**, *4*, 173−207.

(6) Nishiyama, Y.; Langan, P.; Chanzy, H. Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchroton X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2002**, *124*, 9074−9082.

(7) Gomes, T. C.; Skaf, M. S. Cellulose-Builder: a Toolkit for Building Crystalline Structures of Cellulose. *J. Comput. Chem.* **2012**, *33*, 1338−1346.

(8) Shen, T.; Gnanakaran, S. The Stability of Cellulose: A Statistical Perspective from a Coarse-Grained Model of Hydrogen-Bond Networks. *Biophys. J.* **2009**, *96*, 3032−3040.

(9) Ciolacu, D.; Ciolacu, F.; Popa, V. I. Amorphous Cellulose - Structure and Characterization. *Cellulose Chemistry and Technology* **2011**, *45*, 13−21.

(10) Bernardi, R. C.; Cann, I.; Schulten, K. Molecular Dynamics Study of Enhanced Man5B Enzymatic Acitivity. *Biotechnol. Biofuels* **2014**, *7*, 83.

(11) Liwo, A. Coarse Graining: a Tool for Large-Scale Simulations or More? *Phys. Scr.* **2013**, *87*, 058502.

(12) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse Grained Forcefield: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.

(13) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagiela, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; et al. A Unified Coarse-Grained Model of Biological

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* **2014**, *20*, 2306.
(14) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkuizen, L.; Huenenberg, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.
(15) Wohlert, J.; Berglund, L. A. A Coarse-Grained Model for Molecular Dynamics Simulations of Native Cellulose. *J. Chem. Theory Comput.* **2011**, *7*, 753−760.
(16) Hoang, T. X.; Cieplak, M. Sequencing of Folding Events in Go-Like Proteins. *J. Chem. Phys.* **2000**, *113*, 8319.
(17) Nelson Onuchic, J. N.; Nymeyer, H.; Garcia, A. E.; Chahine, J.; Socci, N. D. The Energy Landscape Theory of Protein Folding: Insights Into Folding Mechanisms and Scenarios. *Adv. Protein Chem.* **2000**, *53*, 87−152.
(18) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and en-Route Intermediates for Protein Folding? an Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(19) Karanicolas, J.; Brooks, C. L., III The Origins of Asymmetry in the Folding Transition States of Protein L and Protein G. *Protein Sci.* **2002**, *11*, 2351−2361.
(20) Tozzini, V. Coarse-Grained Models for Proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(21) Sułkowska, J. I.; Cieplak, M. Mechanical Stretching of Proteins—a Theoretical Survey of the Protein Data Bank. *J. Phys.: Condens. Matter* **2007**, *19*, 283201.
(22) Sułkowska, J. I.; Cieplak, M. Selection of Optimal Variants of Go-Like Models of Proteins Through Studies of Stretching. *Biophys. J.* **2008**, *95*, 3174−3191.
(23) Sikora, M.; Sulkowska, J. I.; Cieplak, M. Mechanical Strength of 17 134 Model Proteins and Cysteine Slipknots. *PLoS Comput. Biol.* **2009**, *5*, e1000547.
(24) Takada, S. Coarse-Grained Molecular Simulations of Large Biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137.
(25) Srinivas, G.; Cheng, X.; Smith, J. C. A Solvent-Free Coarse Grain Model for Crystalline and Amorphous Cellulose Fibrils. *J. Chem. Theory Comput.* **2011**, *7*, 2539−2548.
(26) Fan, B.; Maranas, J. K. Coarse-Grained Simulation of Cellulose Iβ With Application to Long Fibrils. *Cellulose* **2015**, *22*, 31−34.
(27) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Addison-Wesley: Massachusetts, 1960; pp 313.
(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(29) Jorgensen, W. L.; Chandrashekar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
(30) Best, R. B.; Chen, Y. G.; Hummer, G. Slow Protein Conformation Dynamics from Multiple Experimental Structure: The Helix/Sheet Transition of Arc Repressor. *Structure* **2005**, *13*, 1755−1763.
(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(32) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.
(33) Tessier, M. B.; DeMarco, M. L.; Yongye, A. B.; Woods, R. J. Extension of the GLYCAM06 Biomolecular Force Field to Lipids, Lipid Bilayers and Glycolipids. *Mol. Simul.* **2008**, *34*, 349−364.
(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(35) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(36) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking With a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2009**, *31*, 455−461.
(37) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(38) Meyer, H.; Biermann, O.; Faller, R.; Reith, D.; Muller-Plathe, F. Coarse Graining of Nonbonded Inter-Particle Potentials Using Automatic Simplex Optimization to Fit Structural Properties. *J. Chem. Phys.* **2000**, *113*, 6264.
(39) Jochum, M.; Andrienko, D.; Kremer, K.; Peter, C. Structure-Based Coarse-Graining in Liquid Slabs. *J. Chem. Phys.* **2012**, *137*, 064102.
(40) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.
(41) Yang, L. W.; Eyal, E.; Bahar, I.; Kitao, A. Principal Component Analysis of Native Ensemble of Biomolecular Structures (PCA_NEST): Insights Into Functional Dynamics. *Bioinformatics* **2009**, *25*, 606−14.
(42) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **1996**, *100*, 2567−2572.
(43) Wertz, J.-L.; Mercier, J. P.; Bédué, O. *Cellulose Science and Technology*; EPFL Press: Lausanne, 2010.
(44) Steiner, T. The Hydrogen Bond in the Solid State. *Angew. Chem., Int. Ed.* **2002**, *41*, 48−76.
(45) Vashchenko, A. V.; Afonin, A. V. Comparative Estimation of the Energies of Intramolecular C-H···O, N-H···O, and O-H···O Hydrogen Bonds According to the QTAIM Analysis and NMR Spectroscopy Data. *J. Struct. Chem.* **2014**, *55*, 636−343.
(46) Munoz, V.; Eaton, W. A. A Simple Model for Calculating the Kinetics of Protein Folding from Three-Dimensional Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11311−11316.
(47) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding Dynamics and Mechanism of β-Hairpin Formation. *Nature* **1997**, *390*, 196.
(48) Chang, I.; Cieplak, M.; Dima, R. I.; Maritan, A.; Banavar, J. R. Protein Threading by Learning. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14350−14355.
(49) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(50) Oyama, T.; Schmitz, G. E.; Dodd, D.; Han, Y.; Burnett, A.; Nagasawa, N.; Mackie, R. I.; Nakamura, H.; Morikawa, K.; Cann, I. Mutational and Structural Analyses of Caldanaerobius Polysaccharolyticus Man5B Reveal Novel Active Site Residues for Family 5 Glycoside Hydrolases. *PLoS One* **2013**, *8*, e80448.
(51) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E.; Edelman, M. Automated Analysis of Interatomic Contacts in Proteins. *Bioinformatics* **1999**, *15*, 327−332.
(52) Schor, M.; Ensing, B.; Bolhuis, P. G. A simple Coarse-Grained Model for Self-Assembling Silk-Like Protein Fibers. *Faraday Discuss.* **2010**, *144*, 127−141.
(53) Sheu, S. Y.; Yang, D. Y.; Selzle, H. L.; Schlag, E. W. Energetics of Hydrogen Bonds in Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12683−12687.
(54) Fersht, A. R.; Shi, J. P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. Hydrogen Bonding and Biological Specificity Analysed by Protein Engineering. *Nature* **1985**, *314*, 235−238.
(55) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The Packing Density in Proteins: Standard Radii and Volumes. *J. Mol. Biol.* **1999**, *290*, 253−266.
(56) Frantz, J. *g3data*, v 1.5.2; 2000.

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041

(57) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701.

DOI: 10.1021/acs.jpcb.5b06141
J. Phys. Chem. B 2015, 119, 12028−12041