Supplementary information for: Osmolyte Induced Changes in Peptide Conformational Ensemble Correlate with Slower Amyloid Aggregation: a Coarse-Grained Simulation Study

Shahar Sukenik[a,b], Liel Sapir[a], and Daniel Harries[a]

[a]Institute of Chemistry and the Fritz Haber Research Center, The Hebrew University, Jerusalem, Israel; [b]Current address: Department of Chemistry, University of Illinois, Urbana, IL 61801, USA.

S1. Simulations

Table S1 – All coarse grained simulations conducted in this study and corresponding composition

	peptide[a]	Water[a]	Sorbitol[a]	Time (μs)	Runs
Water	1	2252	0	2	5
1.5 m sorbitol	1	2311	250	2	5
1.1 m sorbitol	1	2461	200	2	5
0.8 m sorbitol	1	2611	150	2	5
0.65 m sorbitol	1	2686	125	2	5
0.4 m sorbitol	1	2812	83	2	5
Water large	216	486432	0	1	4
Sorbitol large	216	499176	54000	1	3
0.1 m sorbitol	0	36000	256	1	1
0.2 m sorbitol	0	28800	416	1	1
0.3 m sorbitol	0	28800	624	1	1
0.4 m sorbitol	0	28800	832	1	1
0.5 m sorbitol	0	17776	640	1	1
0.6 m sorbitol	0	28800	1248	1	1
0.7 m sorbitol	0	28800	1448	1	1
0.8 m sorbitol	0	28800	1656	1	1
0.9 m sorbitol	0	28800	1864	1	1
1.0 m sorbitol	0	14400	1040	1	1
1.1 m sorbitol	0	28800	2280	1	1
1.2 m sorbitol	0	28800	2488	1	1
1.3 m sorbitol	0	28800	2696	1	1
1.4 m sorbitol	0	28800	2904	1	1
1.5 m sorbitol	0	13328	1440	1	1

[a] number of particles in simulations

Simulation details

All simulations were performed with the GROMACS package.[1–4] The Replica Exchange Molecular Dynamics (REMD) simulation[5] of the peptide in pure water was conducted using the AMBER ff99sb-ildn-NMR force field[6–9] and the TIP4Pew water model.[10] All bond lengths to hydrogen atoms were kept constant with the LINCS algorithm.[11] The simulation consisted of 44 replicas spanning the temperature range $277-465K$, where temperatures were assigned so as to achieve an acceptance rate of 0.2.[12] The dynamics were carried with periodic boundary conditions using a time-step of 2 fs, and replica exchange attempts every 1 ps. Temperatures were held constant using the Nosé-Hoover thermostat with a 1 ps coupling constant.[13,14] Electrostatic calculations were performed using the particle-mesh Ewald (PME) method with 1 Å grid spacing.[15,16] Van der Waals interactions were truncated smoothly with a switching distance of 10 Å and a cutoff distance of 12 Å, and by also accounting for long-range dispersion corrections. The REMD simulation was performed in the NVT ensemble using the volume calibrated by an isobaric-isothermic ensemble simulations at ambient temperature, T = 298 K, and pressure of P = 1 atm. Pressure in this NPT production was kept constant using the Parrinello-Rahman barostat with a 1 ps coupling constant and compressibility of $4.5\cdot10^{-5}$ bar^{-1}.[17,18] Of the 500 ns collected in the REMD production (per replica), only the last 300 ns were used for analysis.

For the molecular dynamics simulations of aqueous sorbitol, we used the TIP4Pew water model and the GLYCAM06 force-field.[19] The partial charges of sorbitol (Figure S1) were reassigned following the standard procedure[20–22] by using the Gaussian 03 package[23], R.E.D tools,[24,25] and Resp program.[26] The derived partial charges are listed in Table S1. The sorbitol simulations were run in the isobaric-isothermic ensemble, as mentioned above. All other simulation details are as detailed above. For the charge calculations we used snapshots of the latter half of a 200 ns simulation.

Coarse-grained simulations were run using MARTINI forcefield v2.1[27,28], with a timestep of 0.04 ps for binary solutions and of 0.01 ps for peptide simulation. The relatively short timestep for peptide simulations is used to allow stable dihedral angles calculations[29]. Simulation settings were set according to recommended MARTINI parameters. Non-bonded parameters were used with a cutoff of 1.2 nm and a shifted function from 0.9 to 1.2 nm. Temperature was controlled using the V-rescale algorithm, with a time constant of 1 ps. Pressure was controlled using a Parinello-Rachman barostat with a time constant of 12 ps, compressibility of $3\cdot10^{-4}$ bar^{-1}, and a reference pressure of 1 bar for all simulation boxes.

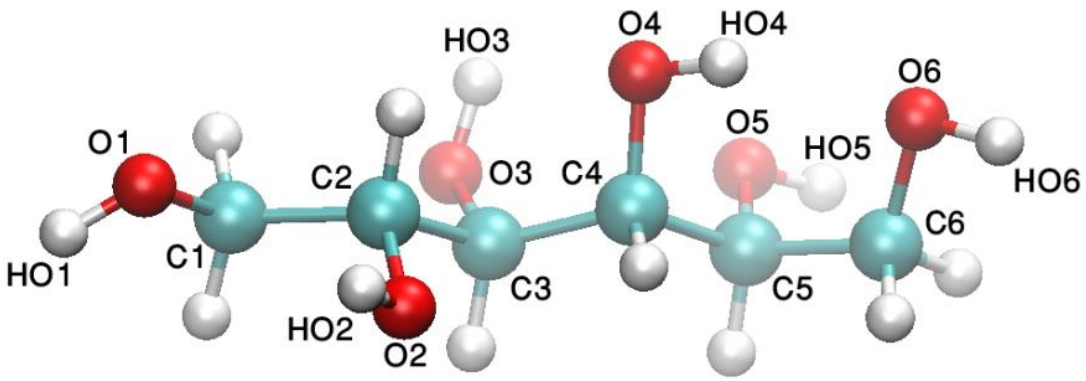

Fig. S1. A sorbitol molecule in the CPK representation.

Table S2. The derived partial charges of sorbitol in the GLYCAM force field. Atom names are as in Figure S1.

Atoms	Partial charge
C1, C6	0.273
O1, O6	-0.696
HO1,HO6	0.423
C2,C3,C4,C5	0.288
O2,O3,O4,O5	-0.714
HO2,HO3,HO4,HO5	0.426
All aliphatic hydrogens	0.0

S2. REMD simulations analysis

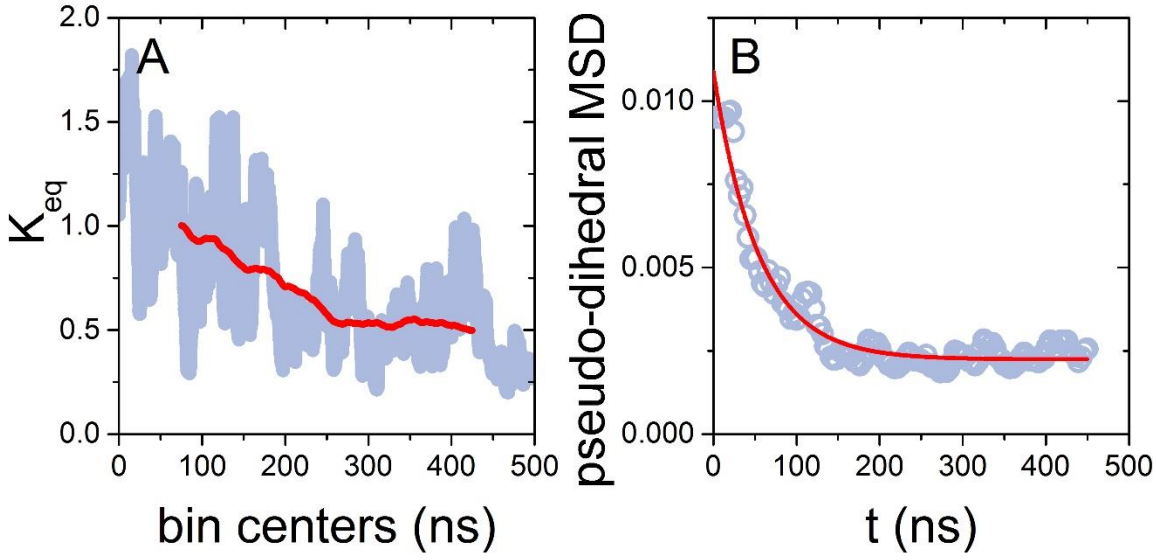

Fig. S2. Convergence of all-atom REMD simulations. (A) Convergence of folding equilibrium constant K_{eq} from the replica at T = 298K. K_{eq} was calculated using the RMSD with respect to a representative folded conformation at each frame, and the cutoff for the folded ensemble was set to RMSD ≤ 2.0 Å. The plot shows the running average of K_{eq}, where each point represents the center of an averaging window of 150 ns, spanning 75 ns to the left and 75 ns to the right of each point (shown in red). This running average is overlaid on the averaged data of non-overlapping, 5 ns windows (grey). The average taken from the 175-325 ns window (centered at 250 ns) already shows the same K_{eq} as 425 ns (within statistical error), demonstrating the equilibrium constant is well converged. (B) Sum of mean-square-differences of all pseudo-dihedral angle distributions between a 50 ns running time window and the final selected dihedral distributions shown in Fig. S3. The red line is an exponential fit to the data in grey. Final pseudo-dihedral angles are taken from the average of the last 300 ns of this REMD replica where both peptide equilibrium (panel A) and dihedral angles (panel B) are well converged.

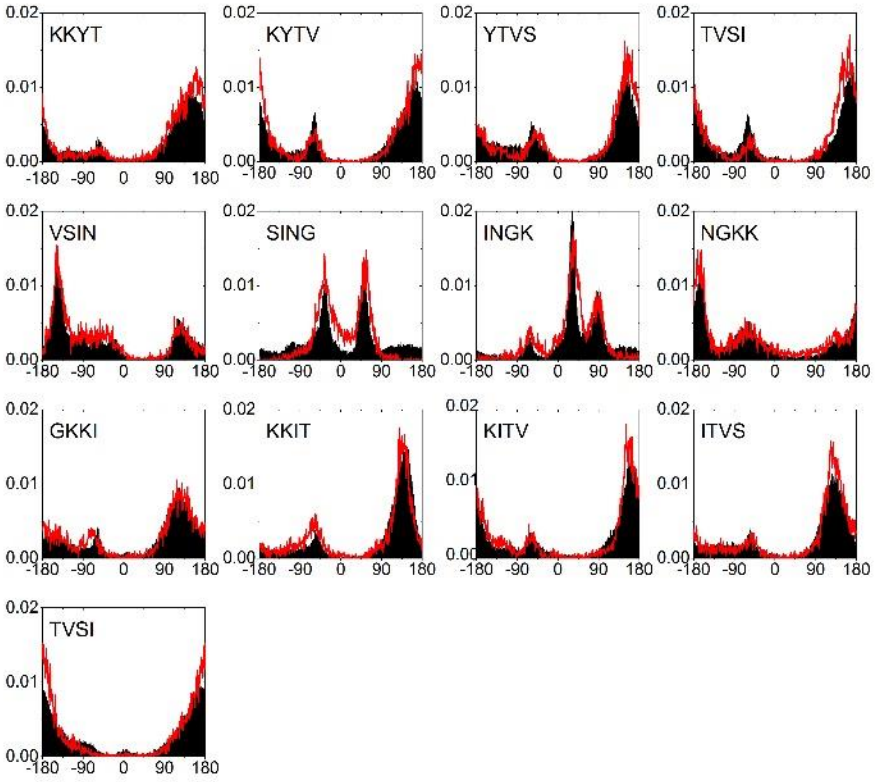

Fig. S3. Resulting pseudo-dihedral angle probability distribution in CG peptide calibrated simulations (red lines) overlaid on distributions from all-atom REMD (black bars) shows good agreement between both models.

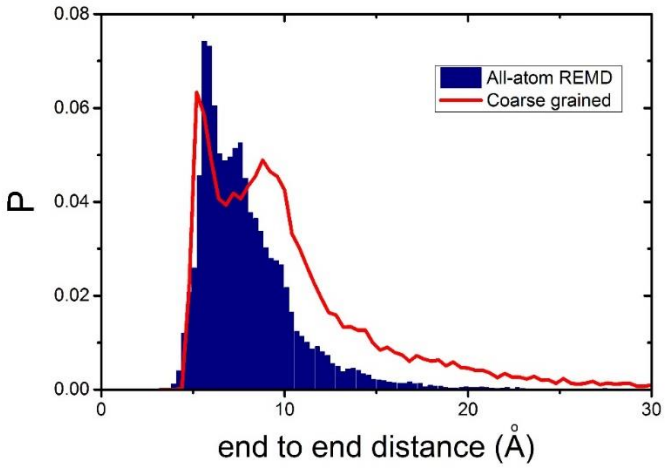

Fig. S4. Comparison of end-to-end distance between CG and all-atom model of met16.

S3. Osmotic pressure of sorbitol model

In order to evaluate how well the sorbitol model reproduces experimentally known solution properties, we calculated the osmotic pressure of the sorbitol model in binary solutions with MARTINI waters using two different methods. The first method uses the sorbitol-sorbitol radial distribution function, $g_{ss}(r)$ to calculate the second virial coefficient B, through the relation[30]

$$B = -2\pi N_{av} \int_{r=0}^{\infty} \left(g_{ss}(r) - 1\right) r^2 dr ,$$

with *r* the distance from the center sorbitol's center of mass, and N_{av} Avogadro's constant. Once determined, *B* can be used to calculate the osmotic pressure Π at low sorbitol concentrations using sorbitol molality m_S, and the mass to volume density of the solution, d :

$$\Pi \cong m_s + Bdm_s^2 .$$

A second method to obtain the osmotic pressure is through the use of the Kirkwood-Buff theory of solutions, as we describe in Ref. 31. Briefly, the derivative of the osmotic pressure with respect to water density is related to the Kirkwood-Buff integrals G_{sw}^{∞} and G_{ww}^{∞} (the former integral, for example, is shown for 1 mol/kg Fig. 2D, and discussed in the main text) by:

$$\left(\frac{\partial \Pi}{\partial \rho_w}\right) = \frac{-m_w}{\rho_w\left(1 + \rho_w G_{ww}^{\infty} - \rho_w G_{sw}^{\infty}\right)} .$$

Here, ρ_w is the water-bead number density, and m_w is the pure water-bead molality. Integrating over a second order polynomial fit to the resulting values for the derivative with respect to ρ_w, and setting the integration constant so that $\Pi = 0$ in pure water, yields the simulated osmotic pressure. While this method gives the exact osmotic pressure, it is numerically noisier than the first estimate as two KBIs are involved rather than one.

Fig S5 shows the results from both methods, and compares them with experimental (red line) and ideal (van 't Hoff, black line) osmotic pressures. Both methods used to calculate the pressure from simulations carry errors with them. These are mainly associated with radial distributions integrals, which are notoriously difficult to converge. Despite this, a qualitatively good fit, showing a slight yet positive deviation from ideal osmotic behavior is apparent, in agreement with experiments. The deviation from experimental results increases, however, as sorbitol concentrations increase. At the concentration used in our peptide aggregation study (1 mol/kg) the osmotic pressure shows a close agreement with experiment. Important to the coarse-grained aspect of our simulations, all calculations performed here do not utilize any special or arbitrary fitting factors to account for the 4-to-1 mapping native to MARTINI. We conclude that despite

the known drawbacks of the coarse grained representation, our model manages to reproduce experimental solution thermodynamic parameters to good accuracy.

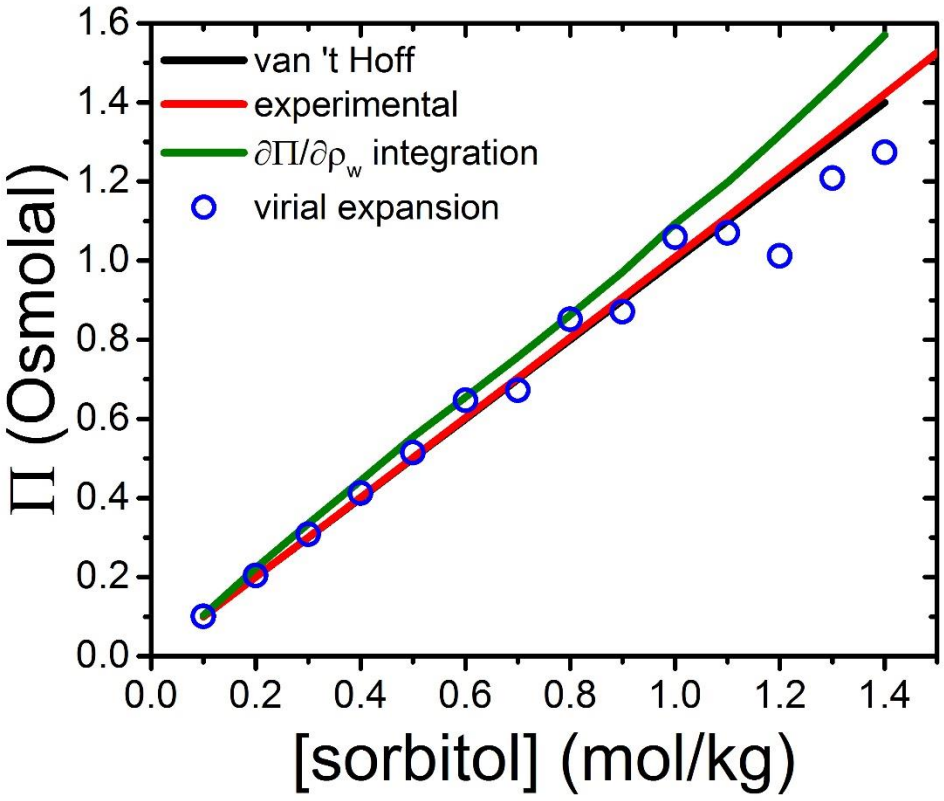

Fig. S5. Comparison between simulated and experimental osmotic pressures. Two methods of calculation for the osmotic pressure, discussed above, are shown as blue symbols and green line, and compared with experimental data shown as the red line. Small positive deviations from ideal osmotic pressure (in black) are observed for both simulated and experimental data at the relevant concentrations used in this study.

S4. Preferential interaction of sorbitol with the peptide

An additional method to obtain the effect of cosolute on protein folding free-energy is through the direct measurement of the number of sorbitol excluding waters, where

$$\Delta\Gamma = \left[N_W \left(1 - \frac{N_S / N_W}{n_S / n_W} \right) \right]_{folded} - \left[N_W \left(1 - \frac{N_S / N_W}{n_S / n_W} \right) \right]_{unfolded}$$

Here, N relates to the number of molecules in the vicinal area of the peptide, while n is the number of molecules in the bulk area, far from the peptide, and the subscripts S and W represent cosolute and water models respectively. A cutoff between the vicinal and bulk regions is determined by the region of g(r) convergence (Fig. 2D). The number for $\Delta\Gamma$ is obtained by averaging the limiting value of the curve shown in Fig. S6 at the region where no more changes (apart from the undulations known to occur in coarse-grained liquids[32,33], typically 15-20 Å from the peptide surface). The difference $\Delta\Gamma$ can then be related to the *m*-value (as described in the main text) through the relation

$$m = \frac{RT}{13.9} \Delta\Gamma,$$

where 13.9 is the number (in moles) of water beads in 1kg of pure MARTINI water, R is the universal gas constant, and T is the temperature.

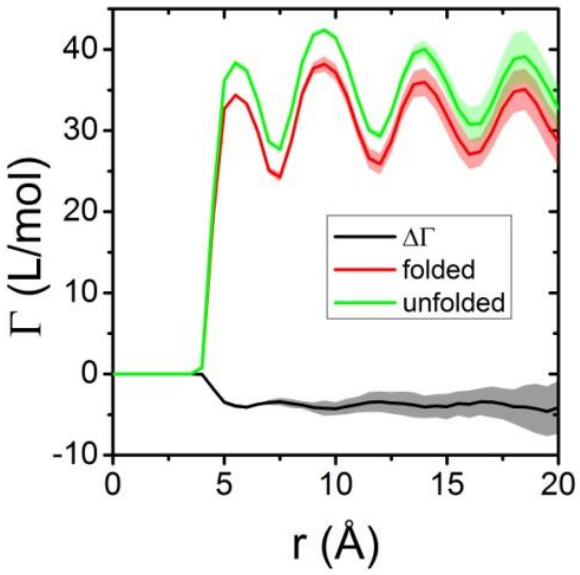

Fig S6. Change in preferential interaction of water with peptide shows difference between the folded and unfolded ensemble in the presence of 1 mol/kg sorbitol. The value of $\Delta\Gamma$ is proportional to the *m*-value discussed in the text. The overall free energy change upon folding in these conditions is $m = -0.9 \pm 0.2$ kJ/mol.

S5. SASA changes upon folding of CG peptide

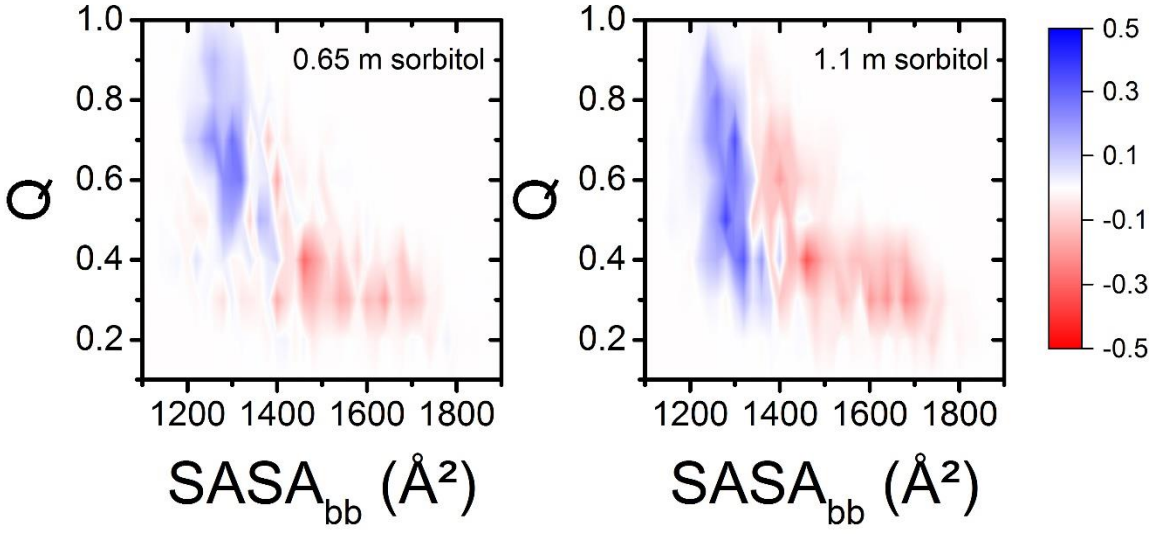

Fig S7. Changes in Q vs. backbone SASA distributions due to the presence of sorbitol. Values are calculated by subtracting the probability distributions in the presence of sorbitol from those in the absence of sorbitol. Red regions show a decrease in probability, while blue regions show an increase, and color intensity indicates the magnitude of distribution change (in percent). The figures clearly show that the presence of sorbitol differentiates high vs. low SASA conformations, but does not have a similarly clear cutoff for Q.

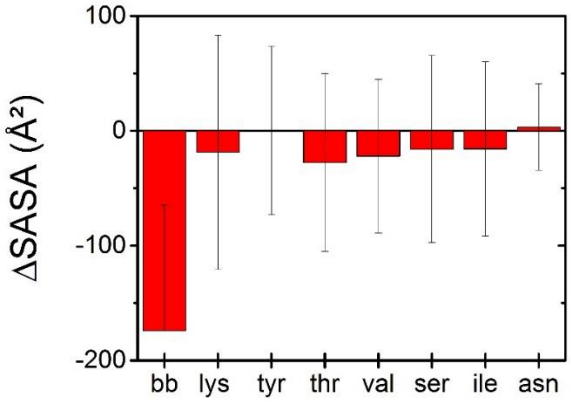

Fig S8. Changes in SASA upon folding pf CG peptide in water for backbone and sidechains of different amino acids. For each ensemble defined by the SASA cutoff specified in the text, an average SASA is obtained using the VMD measure algorithm. The probe size is set to 2.35, the size of the MARTINI water bead. The difference upon folding is shown. Error bars are the propagated standard deviations of the ensemble averages.

S6. Kinetic model for amyloid growth

The master equation used here to numerically fit the experimental result is based on a master equation originally developed by Knowles et al.[34] to model amyloid aggregation kinetics. We previously modified this scheme to include monomer detachment from formed fibrils. This addition was necessary to explain our experimental observables for met16 aggregation in the presence of various osmolytes and polymers, including sorbitol.[35] Here we add yet another term, the attachment of formed fibrils to each other. This is necessary because our simulated trajectories show that a major part of elongation results from attachment of formed fibrils to each other. The final form of the kinetic scheme is formulated as a discrete master equation for clusters of size *i*

$$\frac{dP_i}{dt} = k_{on}\left[P_1\left(P_{i-1} - P_i\right)\right] + k_{off}\left[P_{i+1} - P_i\right] + k_b\left[\sum_{j=i+2}^{N} P_j + P_{2i} - \left(i-3\right)P_i\right] +$$

$$+k_a\left[\sum_{k+j=i} P_j P_k - P_i\left(\sum_{j=2}^{N} P_j\right) - P_i^2\right]$$

Here P_i are probabilities of monomers to exist in a cluster of size i , k_{on} and k_{off} are rates of monomer attachment and detachment to formed fibrils, and k_b and k_a are rates of fibril breakage and attachment to other fibrils, respectively. An additional term of the form $k_n P_1^2$ describes nucleation from two monomers, where the nucleus is of size *i=2*, and exists for monomers and dimers only. The rate of fibril-to-fibril attachment, k_a was necessary to fit our simulations results but was not necessary to fit the aggregation data of our experimental work. This is because the ThT fluorescence assay used as a probe in our experiments cannot discern events of fibril attachment.

References:

(1) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.

(2) 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* (16), 1701–1718.

(3) Juréus, A.; Langel, Ü. Galanin and Galanin Antagonists. *Acta Chim. Slov.* **1996**, *43* (1), 51–60.

(4) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications*. September 1995, pp 43–56.

(5) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1-2), 141–151.

(6) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65* (3), 712–725.

(7) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958.

(8) Li, D.-W.; Brüschweiler, R. NMR-Based Protein Potentials. *Angew. Chemie* **2010**, *122* (38), 6930–6932.

(9) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8* (4), 1409–1414.

(10) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665–9678.

(11) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.

(12) Garcia, A. E.; Herce, H.; Paschek, D. Simulations of Temperature and Pressure Unfolding of Peptides and Proteins with Replica Exchange Molecular Dynamics. In *Annual Reports in Computational Chemistry*; Chemistry, D. C. S. B. T.-A. R. in C., Ed.; Elsevier, 2006; Vol. Volume 2, pp 83–95.

(13) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52* (2), 255–268.

(14) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.

(15) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089.

(16) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995**, *103* (November), 8577–8593.

(17) Parrinello, M. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182.

(18) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50* (5), 1055–1076.

(19) 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* (4), 622–655.

(20) Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. Solvated Ensemble Averaging in the Calculation of Partial Atomic Charges. *J. Comput. Chem.* **2001**, *22* (11), 1125–1137.

(21) Francl, M. M.; Carey, C.; Gange, D. M. Charges Fit to Electrostatic Potentials . 11 . Can Atomic Charges Be Unambiguously Fit to Electrostatic Potentials ? *J. Comp. Chem.* **1995**, *17* (3), 367–383.

(22) Yongye, A. B.; Foley, B. L.; Woods, R. J. On Achieving Experimental Accuracy from Molecular Dynamics Simulations of Flexible Molecules: Aqueous Glycerol. *J. Phys. Chem. A* **2008**, *112* (12), 2634–2639.

(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, J. T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Hratchian, J. E. K. H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 09, Revision B.01. *Guassian Inc*. Gaussian, Inc.: Wallingford CT 2010.

(24) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39* (suppl), W511–W517.

(25) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821–7839.

(26) Bayly, C. I.; Bayly, C. I.; Cieplak, P.; Cieplak, P.; Cornell, W. D.; Cornell, W. D.; Kollman, P. A.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges-the RESP Model. *J. Phys. Chem* **1993**, *97*, 10269–10280.

(27) Marrink, S.-J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824.

(28) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819–834.

(29) Seo, M.; Rauscher, S.; Pomès, R.; Tieleman, D. P. Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- and Elastin-like Peptides. *J. Chem. Theory Comput.* **2012**, *8* (5), 1774–1785.

(30) McQuarrie, D. A.. *Statistical Mechanics*; University Science Books, 2000.

(31) Sapir, L.; Harries, D. Linking Trehalose Self-Association with Binary Aqueous Solution Equation of State. *J. Phys. Chem. B* **2011**, *115* (4), 624–634.

(32) Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. A. Kirkwood–Buff Coarse-Grained Force Fields for Aqueous Solutions. *J. Chem. Theory Comput.* **2012**, *8* (5), 1802–1807.

(33) Mukherji, D.; van der Vegt, N. F. A.; Kremer, K.; Delle Site, L. Kirkwood–Buff Analysis of Liquid Mixtures in an Open Boundary Simulation. *J. Chem. Theory Comput.* **2012**, *8* (2), 375–379.

(34) Knowles, T. P. J.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326* (5959), 1533–1537.

(35) Sukenik, S.; Harries, D. Insights into the Disparate Action of Osmolytes and Macromolecular Crowders on Amyloid Formation. *Prion* **2012**, *6* (1), 26–31.

Appendix 1: Forcefield paramaters for met16 peptide including REMD-calibrated dihedrals in GROMACS format

```
; MARTINI 2.1 Coarse Grained topology file for "met16"
; Sequence:
; KKYTVSINGKKITVSI
; Secondary Structure:
; CCCCCCCCCCCCCCCC

[ moleculetype ]
; Name         Exclusions
Protein         1

[ atoms ]
    1    P5     1   LYS    BB     1  0.0000 ; C
    2    C3     1   LYS   SC1     2  0.0000 ; C
    3    Qd     1   LYS   SC2     3  1.0000 ; C
    4    P5     2   LYS    BB     4  0.0000 ; C
    5    C3     2   LYS   SC1     5  0.0000 ; C
    6    Qd     2   LYS   SC2     6  1.0000 ; C
    7    P5     3   TYR    BB     7  0.0000 ; C
    8   SC4     3   TYR   SC1     8  0.0000 ; C
    9   SC4     3   TYR   SC2     9  0.0000 ; C
   10   SP1     3   TYR   SC3    10  0.0000 ; C
   11    P5     4   THR    BB    11  0.0000 ; C
   12    P1     4   THR   SC1    12  0.0000 ; C
   13    P5     5   VAL    BB    13  0.0000 ; C
   14   AC2     5   VAL   SC1    14  0.0000 ; C
   15    P5     6   SER    BB    15  0.0000 ; C
   16    P1     6   SER   SC1    16  0.0000 ; C
   17    P5     7   ILE    BB    17  0.0000 ; C
   18   AC1     7   ILE   SC1    18  0.0000 ; C
   19    P5     8   ASN    BB    19  0.0000 ; C
   20    P5     8   ASN   SC1    20  0.0000 ; C
   21    P5     9   GLY    BB    21  0.0000 ; C
   22    P5    10   LYS    BB    22  0.0000 ; C
   23    C3    10   LYS   SC1    23  0.0000 ; C
   24    Qd    10   LYS   SC2    24  1.0000 ; C
   25    P5    11   LYS    BB    25  0.0000 ; C
   26    C3    11   LYS   SC1    26  0.0000 ; C
   27    Qd    11   LYS   SC2    27  1.0000 ; C
   28    P5    12   ILE    BB    28  0.0000 ; C
   29   AC1    12   ILE   SC1    29  0.0000 ; C
   30    P5    13   THR    BB    30  0.0000 ; C
   31    P1    13   THR   SC1    31  0.0000 ; C
   32    P5    14   VAL    BB    32  0.0000 ; C
   33   AC2    14   VAL   SC1    33  0.0000 ; C
   34    P5    15   SER    BB    34  0.0000 ; C
   35    P1    15   SER   SC1    35  0.0000 ; C
   36    Qa    16   ILE    BB    36 -1.0000 ; C
   37   AC1    16   ILE   SC1    37  0.0000 ; C

[ bonds ]
; Backbone bonds
    1     4      1   0.35000   400 ; LYS(C)-LYS(C)
    4     7      1   0.35000   400 ; LYS(C)-TYR(C)
    7    11      1   0.35000   400 ; TYR(C)-THR(C)
   11    13      1   0.35000   400 ; THR(C)-VAL(C)
   13    15      1   0.35000   400 ; VAL(C)-SER(C)
```

```
   15    17      1   0.35000   400 ; SER(C)-ILE(C)
   17    19      1   0.35000   400 ; ILE(C)-ASN(C)
   19    21      1   0.35000   400 ; ASN(C)-GLY(C)
   21    22      1   0.35000   400 ; GLY(C)-LYS(C)
   22    25      1   0.35000   400 ; LYS(C)-LYS(C)
   25    28      1   0.35000   400 ; LYS(C)-ILE(C)
   28    30      1   0.35000   400 ; ILE(C)-THR(C)
   30    32      1   0.35000   400 ; THR(C)-VAL(C)
   32    34      1   0.35000   400 ; VAL(C)-SER(C)
   34    36      1   0.35000   400 ; SER(C)-ILE(C)
; Sidechain bonds
    1     2      1   0.33000  5000 ; LYS
    2     3      1   0.28000  5000 ; LYS
    4     5      1   0.33000  5000 ; LYS
    5     6      1   0.28000  5000 ; LYS
    7     8      1   0.32000  5000 ; TYR
   15    16      1   0.25000  7500 ; SER
   19    20      1   0.32000  5000 ; ASN
   22    23      1   0.33000  5000 ; LYS
   23    24      1   0.28000  5000 ; LYS
   25    26      1   0.33000  5000 ; LYS
   26    27      1   0.28000  5000 ; LYS
   34    35      1   0.25000  7500 ; SER

[ constraints ]
    8     9      1   0.27000 ; TYR
    8    10      1   0.27000 ; TYR
    9    10      1   0.27000 ; TYR
   11    12      1   0.26000 ; THR
   13    14      1   0.26500 ; VAL
   17    18      1   0.31000 ; ILE
   28    29      1   0.31000 ; ILE
   30    31      1   0.26000 ; THR
   32    33      1   0.26500 ; VAL
   36    37      1   0.31000 ; ILE

[ angles ]
; Backbone angles
    1     4     7      2    127    25 ; LYS(C)-LYS(C)-TYR(C)
    4     7    11      2    127    25 ; LYS(C)-TYR(C)-THR(C)
    7    11    13      2    127    25 ; TYR(C)-THR(C)-VAL(C)
   11    13    15      2    127    25 ; THR(C)-VAL(C)-SER(C)
   13    15    17      2    127    25 ; VAL(C)-SER(C)-ILE(C)
   15    17    19      2    127    25 ; SER(C)-ILE(C)-ASN(C)
   17    19    21      2    127    25 ; ILE(C)-ASN(C)-GLY(C)
   19    21    22      2    127    25 ; ASN(C)-GLY(C)-LYS(C)
   21    22    25      2    127    25 ; GLY(C)-LYS(C)-LYS(C)
   22    25    28      2    127    25 ; LYS(C)-LYS(C)-ILE(C)
   25    28    30      2    127    25 ; LYS(C)-ILE(C)-THR(C)
   28    30    32      2    127    25 ; ILE(C)-THR(C)-VAL(C)
   30    32    34      2    127    25 ; THR(C)-VAL(C)-SER(C)
   32    34    36      2    127    25 ; VAL(C)-SER(C)-ILE(C)
; Backbone-sidechain angles
    2     1     4      2    100    25 ; LYS(C)-LYS(C) SBB
    1     4     5      2    100    25 ; LYS(C)-LYS(C) SBB
    4     7     8      2    100    25 ; LYS(C)-TYR(C) SBB
    7    11    12      2    100    25 ; TYR(C)-THR(C) SBB
   11    13    14      2    100    25 ; THR(C)-VAL(C) SBB
   13    15    16      2    100    25 ; VAL(C)-SER(C) SBB
   15    17    18      2    100    25 ; SER(C)-ILE(C) SBB
   17    19    20      2    100    25 ; ILE(C)-ASN(C) SBB
   21    22    23      2    100    25 ; GLY(C)-LYS(C) SBB
   22    25    26      2    100    25 ; LYS(C)-LYS(C) SBB
```

```
   25    28    29      2    100    25 ; LYS(C)-ILE(C) SBB
   28    30    31      2    100    25 ; ILE(C)-THR(C) SBB
   30    32    33      2    100    25 ; THR(C)-VAL(C) SBB
   32    34    35      2    100    25 ; VAL(C)-SER(C) SBB
   34    36    37      2    100    25 ; SER(C)-ILE(C) SBB
; Sidechain angles
    1     2     3      2    180    25 ; LYS
    4     5     6      2    180    25 ; LYS
    7     8     9      2    150    50 ; TYR
    7     8    10      2    150    50 ; TYR
   22    23    24      2    180    25 ; LYS
   25    26    27      2    180    25 ; LYS

[ dihedrals ]
; Backbone dihedrals
[dihedrals]
; KKYT
1        4        7        11       1        0        3.70061  1
1        4        7        11       1        0        0.88488  2
1        4        7        11       1        0        0.913    3
1        4        7        11       1        0        0.16143  4
1        4        7        11       1        0        0.21993  5
1        4        7        11       1        90       -1.37761 1
1        4        7        11       1        90       2.05615  2
1        4        7        11       1        90       0.01273  3
1        4        7        11       1        90       -0.07164 4
1        4        7        11       1        90       -0.42108 5
; 1      4        7        11       1        90       6.44013  0
; KYTV
4        7        11       13       1        0        3.76407  1
4        7        11       13       1        0        0.92105  2
4        7        11       13       1        0        2.06649  3
4        7        11       13       1        0        0.63389  4
4        7        11       13       1        0        -0.11833 5
4        7        11       13       1        0        -0.46887 6
4        7        11       13       1        90       0.13201  1
4        7        11       13       1        90       2.5362   2
4        7        11       13       1        90       0.48809  3
4        7        11       13       1        90       -0.33137 4
4        7        11       13       1        90       -0.32474 5
4        7        11       13       1        90       0.03125  6
; 4      7        11       13       1        90       3.21116  0
; YTVS
7        11       13       15       1        0        3.91116  1
7        11       13       15       1        0        0.03712  2
7        11       13       15       1        0        0.09689  3
7        11       13       15       1        0        0.41527  4
7        11       13       15       1        0        -0.05057 5
7        11       13       15       1        90       0.99996  1
7        11       13       15       1        90       3.33799  2
7        11       13       15       1        90       0.685    3
7        11       13       15       1        90       0.61899  4
7        11       13       15       1        90       -0.32148 5
; 7      11       13       15       1        90       2.97958  0
; TVSI
11       13       15       17       1        0        3.4444   1
11       13       15       17       1        0        -0.3654  2
11       13       15       17       1        0        0.96655  3
11       13       15       17       1        0        0.30801  4
11       13       15       17       1        0        -0.10031 5
11       13       15       17       1        0        -0.12022 6
11       13       15       17       1        0        -0.02915 7
11       13       15       17       1        90       1.54725  1
```

```
11       13       15       17       1        90       2.4196   2
11       13       15       17       1        90       0.0918   3
11       13       15       17       1        90       -0.4578  4
11       13       15       17       1        90       -0.85695 5
11       13       15       17       1        90  -0.43002      6
11       13       15       17       1        90  -0.11371      7
; 11     13       15       17       1        90       6.65684  0
; VSIN
13       15       17       19       1        0        1.75741  1
13       15       17       19       1        0        -0.15709 2
13       15       17       19       1        0        -0.99829 3
13       15       17       19       1        0        0.52766  4
13       15       17       19       1        0        -0.49089 5
13       15       17       19       1        0        0.14996  6
13       15       17       19       1        0        -0.23281 7
13       15       17       19       1        90       2.16539  1
13       15       17       19       1        90       0.87817  2
13       15       17       19       1        90       1.17224  3
13       15       17       19       1        90       -0.50514 4
13       15       17       19       1        90       0.56317  5
13       15       17       19       1        90       0.22127  6
13       15       17       19       1        90       -0.27671 7
; 13     15       17       19       1        90       7.03003  0
; SING
15       17       19       21       1        0        -0.46157 1
15       17       19       21       1        0        0.70728  2
15       17       19       21       1        0        0.941    3
15       17       19       21       1        0        0.65507  4
15       17       19       21       1        0        0.10295  5
15       17       19       21       1        0        -0.16128 6
15       17       19       21       1        0        -0.11663 7
15       17       19       21       1        90       -0.33144 1
15       17       19       21       1        90       0.15752  2
15       17       19       21       1        90       0.44237  3
15       17       19       21       1        90       0.82642  4
15       17       19       21       1        90       0.40483  5
15       17       19       21       1        90       -0.03855 6
15       17       19       21       1        90       -0.43519 7
; 15     17       19       21       1        90       9.171    0
; INGK
17       19       21       22       1        0        -0.86717 1
17       19       21       22       1        0        0.81001  2
17       19       21       22       1        0        0.6579   3
17       19       21       22       1        0        -0.01495 4
17       19       21       22       1        0        0.10205  5
17       19       21       22       1        0        0.15164  6
17       19       21       22       1        0        -0.01947 7
17       19       21       22       1        0        -0.08432 8
17       19       21       22       1        0        0.02006 9
17       19       21       22       1        90       -2.41202 1
17       19       21       22       1        90       -0.68191 2
17       19       21       22       1        90       -0.39993 3
17       19       21       22       1        90       -0.80828 4
17       19       21       22       1        90       -0.76515 5
17       19       21       22       1        90       0.26492  6
17       19       21       22       1        90       0.63569  7
17       19       21       22       1        90       0.31539  8
17       19       21       22       1        90       -0.02    9
; 17     19       21       22       1        90       15.32339 0
; NGKK
19       21       22       25       1        0        4.16935  1
19       21       22       25       1        0        -1.32069 2
19       21       22       25       1        0        0.99667  3
```

```
19       21       22       25       1        0        -0.25206 4
19       21       22       25       1        0        0.04114  5
19       21       22       25       1        0        -0.10523 6
19       21       22       25       1        0        0.01626  7
19       21       22       25       1        90       0.52452  1
19       21       22       25       1        90       1.18747  2
19       21       22       25       1        90       0.36998  3
19       21       22       25       1        90       -0.56404 4
19       21       22       25       1        90       0.41719  5
19       21       22       25       1        90       -0.34757 6
19       21       22       25       1        90       0.24711  7
; 19     21       22       25       1        90       7.15402  0
; GKKI
21       22       25       28       1        0        3.35857  1
21       22       25       28       1        0        1.28718  2
21       22       25       28       1        0        0.36333  3
21       22       25       28       1        0        0.06896  4
21       22       25       28       1        0        -0.59184 5
21       22       25       28       1        0        -0.31134 6
21       22       25       28       1        0        0.00339  7
21       22       25       28       1        90       -0.57581 1
21       22       25       28       1        90       1.56454  2
21       22       25       28       1        90       0.39271  3
21       22       25       28       1        90       -0.73144 4
21       22       25       28       1        90       -0.17465 5
21       22       25       28       1        90       0.0753   6
21       22       25       28       1        90       0.09612  7
; 21     22       25       28       1        90       7.03307  0
; KKIT
22       25       28       30       1        0        4.3381   1
22       25       28       30       1        0        1.08812  2
22       25       28       30       1        0        -0.11046 3
22       25       28       30       1        0        0.27796  4
22       25       28       30       1        0        -0.50607 5
22       25       28       30       1        0        -0.0532  6
22       25       28       30       1        90       -0.97285 1
22       25       28       30       1        90       3.06055  2
22       25       28       30       1        90       -0.53428 3
22       25       28       30       1        90       -0.04908 4
22       25       28       30       1        90       -0.38271 5
22       25       28       30       1        90       -0.40955 6
; 22     25       28       30       1        90       7.23719  0
; KITV
25       28       30       32       1        0        3.26266  1
25       28       30       32       1        0        -0.35859 2
25       28       30       32       1        0        0.68107  3
25       28       30       32       1        0        0.10521  4
25       28       30       32       1        0        -0.01496 5
25       28       30       32       1        0        -0.15923 6
25       28       30       32       1        0        -0.17012 7
25       28       30       32       1        90       0.63646  1
25       28       30       32       1        90       1.97928  2
25       28       30       32       1        90       -0.14443 3
25       28       30       32       1        90       -0.06725 4
25       28       30       32       1        90       -0.30435 5
25       28       30       32       1        90       -0.06585 6
25       28       30       32       1        90       0.18504  7
; 25     28       30       32       1        90       6.94951  0
; ITVS
28       30       32       34       1        0        3.12156  1
28       30       32       34       1        0        1.65553  2
28       30       32       34       1        0        0.00546  3
28       30       32       34       1        0        0.45185  4
```

```
25       28       30       32       1        0        -0.04926 5
25       28       30       32       1        0        -0.23398 6
25       28       30       32       1        0        -0.2555  7
28       30       32       34       1        90       -0.17809 1
28       30       32       34       1        90       2.6502   2
28       30       32       34       1        90       0.25742  3
28       30       32       34       1        90       0.00349  4
25       28       30       32       1        90       -0.26109 5
25       28       30       32       1        90       -0.12253 6
25       28       30       32       1        90       0.12481  7
; 28     30       32       34       1        90       4.95073  0
; TVSI
30       32       34       36       1        0        4.37967  1
30       32       34       36       1        0        -0.29652 2
30       32       34       36       1        0        -0.13548 3
30       32       34       36       1        0        -0.73646 4
25       28       30       32       1        0        -0.36224 5
30       32       34       36       1        90       -0.05251 1
30       32       34       36       1        90       0.69862  2
30       32       34       36       1        90       -0.09365 3
30       32       34       36       1        90       -0.23855 4
25       28       30       32       1        90       -0.18586 5
; 30     32       34       36       1        90       9.44401  0


; Sidechain improper dihedrals
    7     9    10     8      2      0    50 ; TYR

#ifdef POSRES
#include "posre.itp"
#endif
```

Appendix 2: Bonded parameters for sorbitol CG model

```
[ moleculetype ]
; molname nrexcl
SORB 1

[ atoms ]
;id type resnr residu atom cgnr charge mass
1   P4c  1     SORB   B1   1    0.000  60.0528
2   P4c  1     SORB   B2   2    0.000  60.0528
3   P4c  1     SORB   B3   3    0.000  60.0528

[ constraints ]
; i j   funct   length  forcec.
   1 2 1 0.27 35000
   2 3 1 0.27 35000

[angles]
; i j k funct angle force.c.
   1 2 3 2 142.000 170.0
```

Appendix 3: forcefield parameters for P4c bead type.

```
P4c 72.0 0.000 A 0.0 0.0

  P4c    P4c      1        0.19402E-00       0.20914E-02 ; almost attractive
  P5     P4c      1        0.17246E-00     0.18590E-02 ; semi attractive
  SP5    P4c      1        0.24145E-00       0.26027E-02 ; supra attractive
  P4c    P4       1       0.21558E-00     0.23238E-02 ; attractive
  P4c    BP4      1       0.76824E-00     0.26348E-01 ; supra attractive,s=0.57
  P4c    SP4      1       0.21558E-00     0.23238E-02 ; attractive
  P4c    P3       1       0.21558E-00     0.23238E-02 ; attractive
  P4c    SP3      1       0.21558E-00     0.23238E-02 ; attractive
  P4c    P2       1       0.19402E-00     0.20914E-02 ; almost attractive
  P4c    SP2      1       0.19402E-00     0.20914E-02 ; almost attractive
  P4c    P1       1       0.19402E-00     0.20914E-02 ; almost attractive
  P4c    SP1      1       0.19402E-00     0.20914E-02 ; almost attractive
  P4c    Nda      1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    SNda     1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    Nd       1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    SNd      1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    Na       1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    SNa      1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    N0       1       0.15091E-00     0.16267E-02 ; intermediate
  P4c    SN0      1       0.17246E-00     0.18590E-02 ; semi attractive
  P4c    C5       1       0.13366E-00     0.14408E-02 ; almost intermediate
  P4c    SC5      1       0.13366E-00     0.14408E-02 ; almost intermediate
  P4c    C4       1       0.11642E-00     0.12549E-02 ; semi repulsive
  P4c    SC4      1       0.11642E-00     0.12549E-02 ; semi repulsive
  P4c    C3       1       0.11642E-00     0.12549E-02 ; semi repulsive
  P4c    SC3      1       0.11642E-00     0.12549E-02 ; semi repulsive
  P4c    C2       1       0.99167E-01     0.10690E-02 ; almost repulsive
  P4c    AC2      1       0.99167E-01     0.10690E-02 ; almost repulsive
  P4c    SC2      1       0.99167E-01     0.10690E-02 ; almost repulsive
  P4c    C1       1       0.86233E-01     0.92953E-03 ; repulsive
  P4c    AC1      1       0.86233E-01     0.92953E-03 ; repulsive
  P4c    SC1      1       0.86233E-01     0.92953E-03 ; repulsive
  P4c    Qda      1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    SQda     1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    Qd       1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    SQd      1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    Qa       1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    SQa      1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    Q0       1       0.24145E-00     0.26027E-02 ; supra attractive
  P4c    SQ0      1       0.24145E-00     0.26027E-02 ; supra attractive
```