Quantitative assessment of amylose dimerization process across force fields

Jaya krishna Koneru and Jagannath Mondal*

Tata Institute of Fundamental Research, Centre for Interdisciplinary Science, 36/P Gopanapalli Village, Serillingampally, Hyderabad-500 107, Telangana, India

E-mail: jmondal@tifrh.res.in

Despite being a key biomacromolecule, carbohydrate has received much less attention from computational community, compared to protein and lipids. This is majorly because of slow development of classical force fields for carbohydrate due to associated complexity in sampling its intrinsic flexibility and lack of an extensive assessment of existing carbohydrate force fields. Towards this end, the current work provides a robust comparison of four carbohydrate force fields (CHARMM36, GROMOS53A6$_{CARBO\_R}$, OPLS-AA, GLYCAM36) by evaluating their ability to simulate the dimerization process of a pair of amylose chain, the key ingredient of starch. The microsecond long molecular dynamics simulations on each of the four force fields capture spontaneous formation of double-helical self-assembled morphology from a pair of well-separated conformations. However, geometrical clustering of the trajectories reveal that these force fields mutually differ in sampling diverse array of conformations ranging from non-helical to partially open helical structures in addition to double-helical structures. Notably, the simulations reveal that relative to CHARMM36 and GLYCAM06 force field, GROMOS53A6$_{CARBO\_R}$ over stabilizes the double-helical self-assembled morphology. CHARMM36 force field predicts significant transition between double-helical and partially open helical structure. The force fields differ in the relative propensity of parallel and anti-parallel double-helical formations. Overall by comparing all force fields in equal footing, this work provides a guided benchmark for carbohydrate simulation.

Keywords: Amylose, carbohydrate force field, helix, self-assembly.

Introduction

Amylose is the polymeric form of glucose, the building block of carbohydrate. As a key constituent of starch, amylose acts as the major contributor to the aqueous insolubility of starch and hence occasionally utilized as a thickener and gel in foods and textiles. The broad range of its utility stems from the underlying conformational flexibility of amylose, giving rise to helicity and self-association. In a key experiment, Rundle and Edwards[1] showed the first evidence of a helical structure for amylose. Later X-ray crystallographic studies provided the crystallographic structure of single and double-helical conformation. Using solution NMR study, Cheetham and Tao[2] showed spontaneous transition between helical and coil conformation of single stranded amylose. There are also experimental reports[3,4] of reversible rotation of amylose alpha glycosidic torsions by 180 degree, generally known as Syn→Anti transition, which interrupts helical shapes causing kinks. The kinks are hypothesized to be associated with coil-like structure. More over, the pyranose ring puckering statistics also play an important role in deciding conformational equilibrium in the amylose[5]. The purpose of the current article is to evaluate how the overall helix-coil conformational heterogeneity, combined with the torsional transitions, effectively impacts the amylose self-aggregation process.

Classical Molecular Dynamics (MD) simulation[6] is a method of choice for sampling the underlying conformational diversity and self-assembly of biopolymers at an atomistic precision, due to its ability to capture fast structural transitions at an atomistic length scale, in absence of high-resolution experimental data. However, the quality of results derived from the MD simulations is often subjected to quality of underlying force field[7] describing bonded and nonbonded interactions. The empirical nature of underlying force field has given rise to multiple variants, often rising from each developer's independent and unique approach.

Despite being one of the major biomolecules of life, carbohydrates have received much lesser attention from the computational community, compared its other biomolecular counterparts, namely protein, nucleic acid and lipids. This is majorly because of the complexity associated with intrinsic flexibility of carbohydrates, which has prevented their exhaustive experimental characterization. Due to the scarcity of X-ray and NMR data, the development process of carbohydrate force fields has moved relatively more slowly than its other biomolecular counterparts. While last decade has seen independent developments of a number of carbohydrate force fields, an exhaustive assessment of the quality of the existing force fields is currently lacking. In the present article, we evaluate the relative ability of four popular carbohydrate force fields, namely CHARMM36[8,9], GLYCAM06[10], GROMOS53A6$_{CARBO\_R}$[10] and OPLS-AA[11] to describe the self-assembly process of a pair of single-stranded amylose dodecamer in aqueous solution. Towards this end, we have generated multiple copies of microsecond-long trajectories to extensively sample the diverse dimerization process. While the formation of double-helical amylose structure is found to be a common feature in all force fields, an atomistic investigation of our simulations show significant diversity in the overall conformational heterogeneities in the dimerization process. As would be revealed in the article later, these force fields mutually differ in sampling diverse array of conformations ranging from non-helical to partially open helical structures in addition to double-helical structures. The kinetics of the processes is also found to be strongly contingent upon the nature of the force field.

Experimental

Simulation model and methods:

Amylose dimer models were manually built from their single-stranded analogs for their respective force fields. The single-stranded structures were obtained from GLYCAM carbohydrate builder (http://glycam.org/tools/molecular-dynamics/oligosaccharide-builder/build-glycan?id=1).A pair of amylose dodecamer, at 20 Å separation, was then solvated in a cubic box with side length 68 Å with TIP3P water model[12] for all force fields except GROMOS53A6$_{CARBO\_R}$, for which SPC water model[13] was used. Fig. 1 depicts the initial configura-
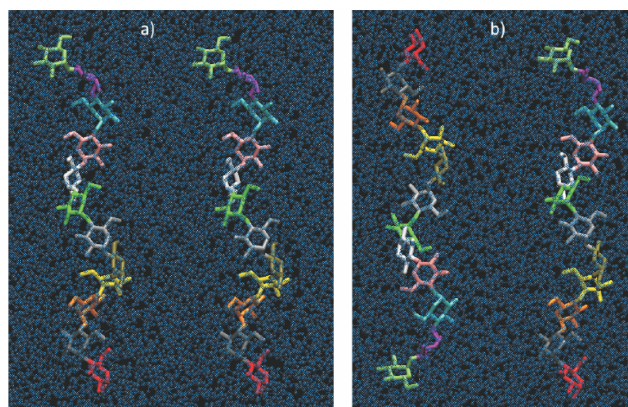

Fig. 1. Initial configurations of a pair of well-separated amylose dodecamer in (a) parallel orientation and (b) anti-parallel orientation. The pairs are solvated in water.

tions for the simulations. All simulations are performed using GROMACS 5.1.2[14] simulation package in NPT ensemble.

The components were coupled separately with the thermostat. The temperature was maintained at 298 K by employing Nose Hoover temperature coupling scheme[15,16] using a coupling constant of 1 ps. An isotropic pressure coupling using Parrinello-Rahman protocol[17] was implemented to maintain the pressure at 1 bar. Verlet cutoff[18] scheme was implemented for Lenard-Jones interaction extending upto 1.2 nm, while particle mesh Ewald and reaction field (only GROMOS53A6$_{CARBO\_R}$) schemes were implemented for treating electrostatic interactions. All bond lengths involving hydrogen atoms of the dimers were constrained using the LINCS[19] algorithm and water hydrogen bonds were fixed using the SETTLE approach[20]. Simulations were performed using the leapfrog integrator with a time step of 2 fs and initiated by randomly assigning the velocities of all particles from a Maxwell-Boltzman distribution. For each force field, the aqueous solution of well-separated amylose dimer was subjected to multiple independent simulations, each 500 ns-1 μs long. The un-biased trajectories were used to compute the radius of gyration of the amylose dimer, the centre of mass distance between the chains and number of inter-chain hydrogen bonds. The aggregated trajectories were subjected to geometric clustering using k-means clustering[21] algorithm. The dimer radius of gyration, center of mass distance and head-head/head-tail distances between two chains were used as the clustering metrics.

Results and discussion

In this work, first we investigate whether a pair of dodecamer of amylose chains, initially mutually well separated (Fig. 1), spontaneously self-assemble during the period of computer simulations. In this regard, as discussed in the method section, we have compared our investigations across four different carbohydrate force fields, namely CHARMM36, GLYCAM06, OPLS-AA and GROMOS53A6$_{CARBO\_R}$.

First we consider the effect of using CHARMM36 force field for modeling the self-assembly behavior of amylose pair. Starting with a pair of well-separated amylose in their parallel orientation, we find that the two chains approach each other and spontaneously form a double-helical self-assembled morphology within the time scale of current MD simulations. For quantitative characterization of the dynamical self-assembly process, we compute the radius of gyration of the combined pair of amylose chains and monitor the distance between centres of mass between them. Figs. 2 and 3 depicts the time profile of these two metrics in a representative trajectory for all force fields. We find that the values of both the metrics gradually decrease with time, suggesting the progression of dimerization process with time. Since hydrogen bonds play a main role in formation of the double-helical morphology, we have also computed the number of hydrogen bonds between two amylose chains to track the formation of double helical structure. As depicted in Fig. 4, the number of hydrogen bonds progressively increases with time, suggesting the formation of double helical morphology. However, in the case of CHARMM36 force field, we also notice occasional transition of perfect double helical structure to partially open double helical structure as well. These are characterized by the sporadic drop in number of hydrogen bonds on the way to form double-helical structure. Interestingly for the specific case of CHARMM36 force field, we find that the formation of double-helical morphology is strongly contingent upon the initial orientation of the two strands to be mutually parallel. When simulation starts with both the strands in anti-parallel orientation, we find that the two amylose strands do not form any double-helical assembly for CHARMM36 force field.
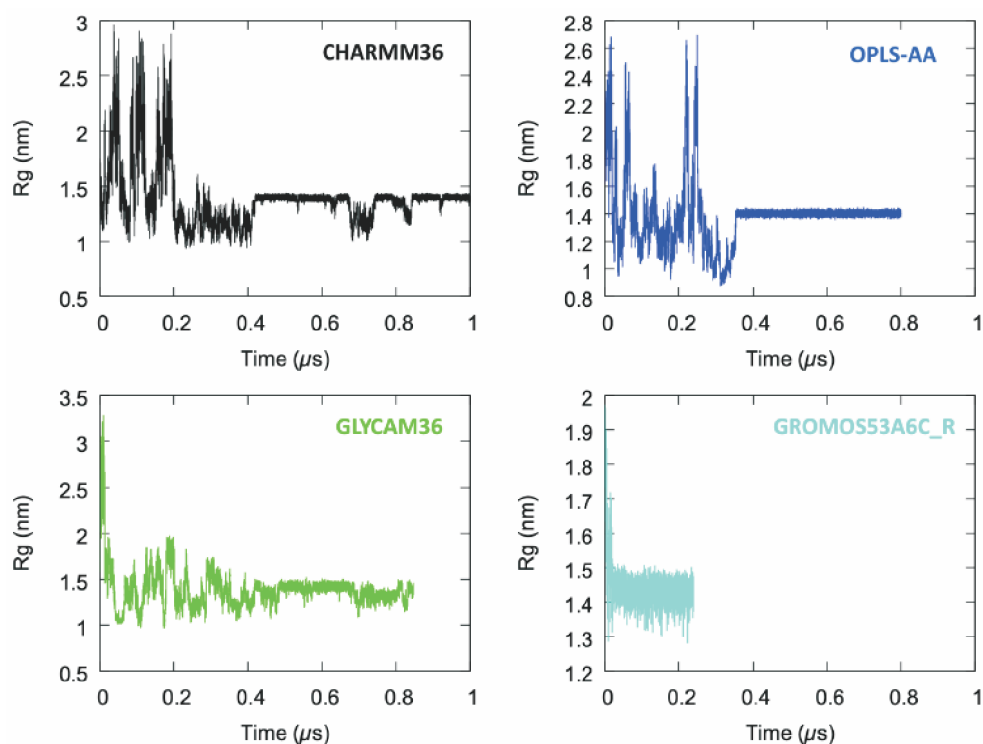

Fig. 2. Time profile of Radius of Gyration of amylose dimers for different force fields.
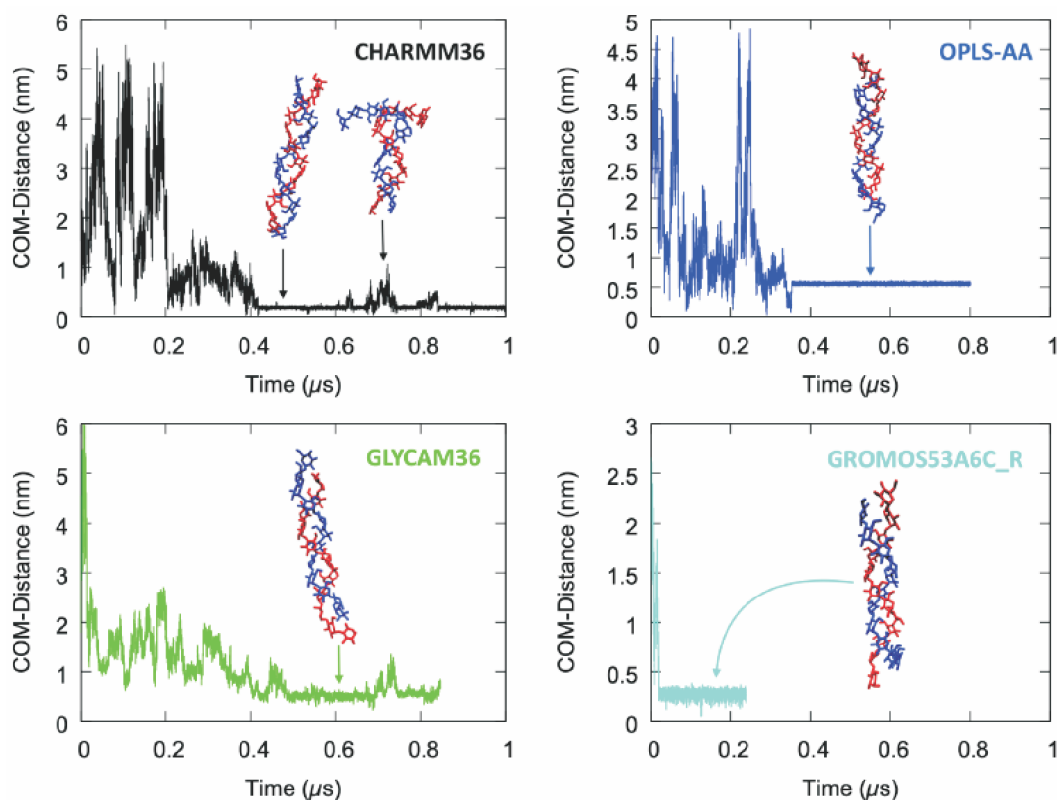
Fig. 3. Distance between the centre of mass of amylose dimers for different force-fields reveals the sensitivity of respective force fields towards dimerization.

While all four force fields employed in the present article predict formation of double-helix formation as described for CHARMM36, we found there are some quantitative differences in the sensitivity of force fields towards helical formation. For example, as shown in Figs. 2-4 in case of GROMOS53A6$_{CARBO\_R}$, the time required for formation of double-helix is very less when compared to other force fields. This suggests that GROMOS53A6$_{CARBO\_R}$ is very biased to the formation of double-helix. The high stability of double-helix and absence of intermediate conformations after the formation of helix in case of OPLS-AA were also noted.

To quantitatively compare the conformational heterogeneities underlying the dimerization process across all the force fields, we geometrically clustered the conformations from all MD trajectories. The COM distance, head-to-head and tail-to-tail distances and number of hydrogen bonds between two chains, as presented in Figs. 2-4, were used as the parameters for performing geometrical clustering. Figs. 5 and 6 show the key snapshots corresponding to cluster midpoint of CHARMM36, OPLS-AA, GLYCAM06 and GROMOS53A6 $_{CARBO\_R}$ respectively, whereas Table 1 refers to the conformational percentages of different conformations of respective force fields arising from geometric clustering.

For CHARMM36, if we exclude the conformations when both chains are far apart, parallel perfect double-helical population is ~10% of the total conformational space (Fig. 5a(I)), ~3.2% of conformations are partially open double-helical-intermediate (Fig. 5a(II)) where as ~16.8% are non-helical conformations (Table 1). For the simulations starting with the chains in mutually anti-parallel conformation, we didn't observe the formation of anti-parallel double-helix in case of CHARMM36. Any simulation starting with anti-parallel conformation with CHARMM36 force field either results in the two chains to move away or forms non-helical structure. Rather in such cases, we found that > 80% of population are far apart and only ~16.2% are in non-helical conformations
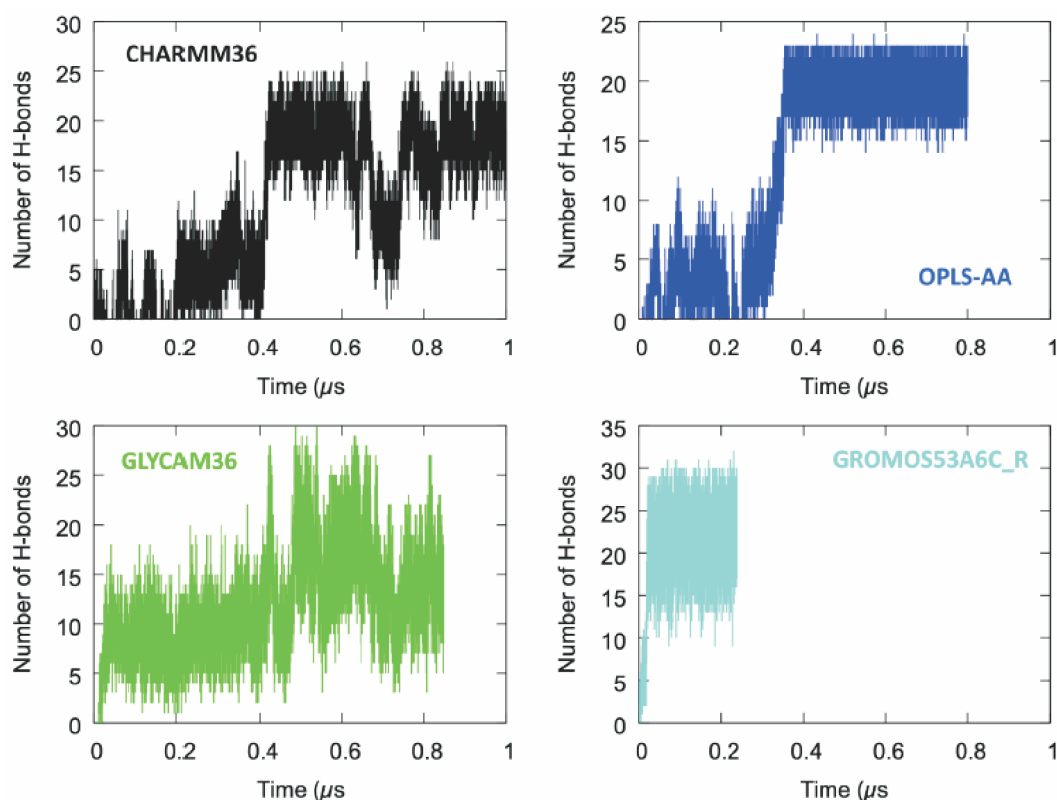
Fig. 4. The time profile of number of hydrogen bonds between two amylose chains for four force fields under investigation.

(Fig. 5b, Table 1). This may be attributed to CHARMM36 favouring intermediate structures for mono-dodecas-accharide.

Point to take note in case of OPLS-AA force field is that, in parallel system there is no formation of partially open double-helical intermediate in our simulations, unlike CHARMM36 force field. The reason for that is high stability of the double helix in OPLS-AA force field. We observed ~27.5% of population are in non-helical conformations and ~18.1% of total conformations formed double helix (Fig. 5c(II), (I) respectively). Interestingly though ~41.4% of conformations are found to be in linear conformation (see Fig. 5c), which is quite stable. On the other hand, when starting with anti-parallel orientation, OPLS-AA expressed very less population ~4.9% of anti-parallel helix and ~58.4% of the conformations are non-helical conformations.

As predicted previously by earlier computational study, GLYCAM06 force field predicts the formation of double-helical morphology, with simulations initiated with well-separated

chains either in parallel or anti-parallel orientations. In our simulations we observed that, in case of parallel helix there is a frequent exchange between the perfect double-helical conformation and it's partially open intermediate. Their combined conformational population is ~40.5% (Fig. 6a). Interestingly the double-helix dissociated after ~400 ns from the time of it's formation, which we didn't observe in other three force fields, which is why ~54.9% of conformations are non-helical conformations (Table 1). For anti-parallel system, we observed the formation of helix very quickly, but for most of the time the system preferred helical intermediate conformation. Overall, after clustering we observed ~52.8% of population to be helical intermediates where as helical conformation is ~3.4% (Fig. 6, Table 1).

In case of GROMOS53A6$_{CARBO\_R}$ force field, the double-helix conformation is unusually favorable in both parallel and anti-parallel systems (Table 1). Although we observed partially open helical intermediates in both conformations, the results from clustering analysis showed that most favored
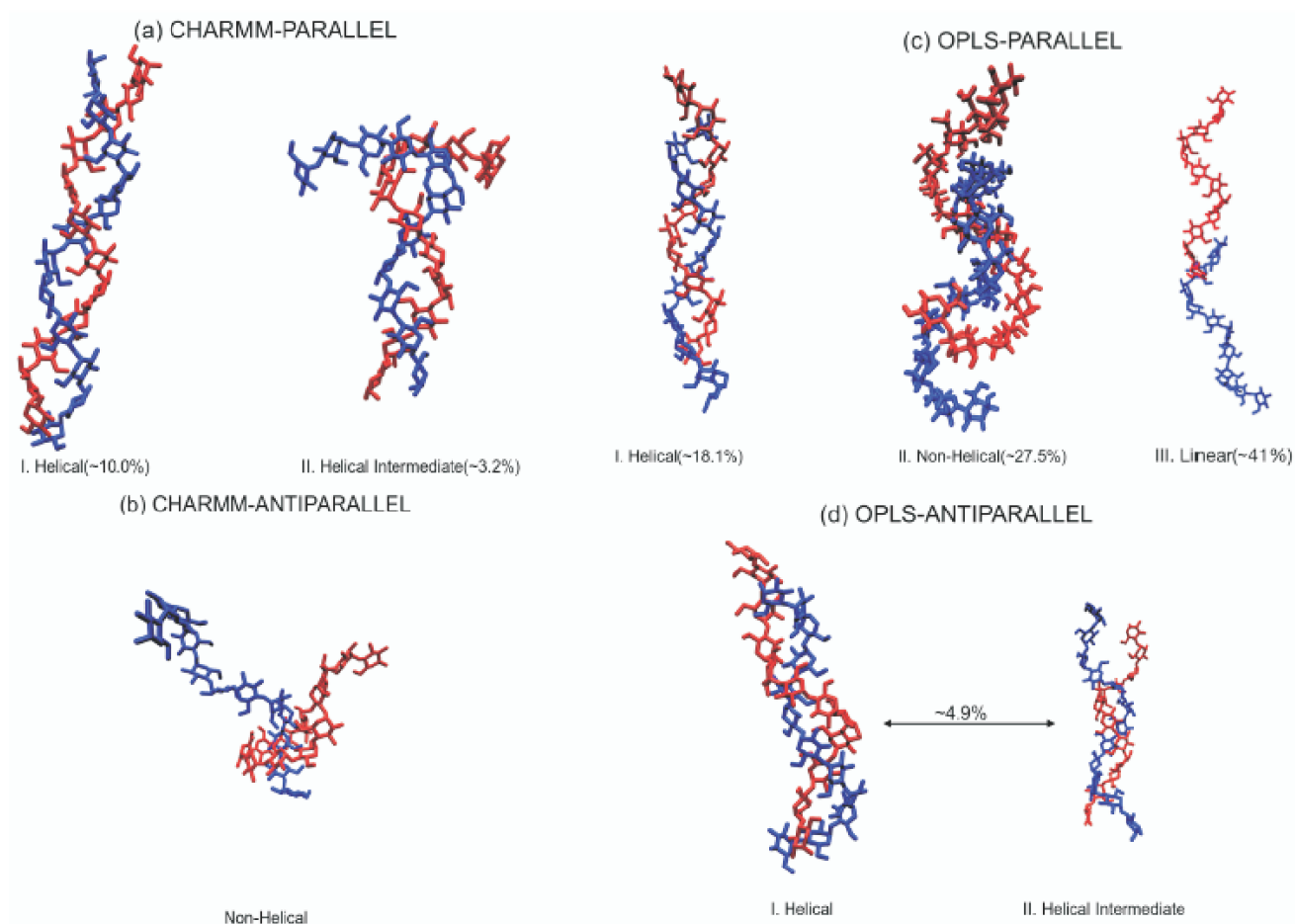
Fig. 5. The representative conformations of dimer as obtained from clustering of trajectories initiated from parallel and anti-parallel conformations for CHARMM36 and OPLS-AA force field.

Table 1. The populations of various representative dimer conformations in different force field, as obtained from cluster analysis. 'P' and 'AP' denotes parallel and anti-parallel orientations at the initial conformation

Force field and orientation	Double-helix (%)	Open helix (%)	Non-helical (%)	Far-apart (%)	Linear (%)
CHARMM-P	10.0	3.2	16.8	69.8	–
CHARMM-AP	–	–	16.2	83.8	–
OPLS-P	18.1%	–	27.5	13.0	41.4
OPLS-AP	–	4.9	58.4	36.6	–
GLYCAM-P	–	40.5	54.9	4.6	–
GLYCAM-AP	3.4	52.8	35.4	1.21	–
GROMOS-P	71.2	9.1	10.5	9.2	–
GROMOS-AP	91.6	1.1	4.4	3.5	–

conformations for parallel and anti-parallel system is helical conformations with ~71.2% and ~91.1% respectively (Fig. 6c, 6d). Another point to be noted is the time taken for the formation of helix in both parallel and anti-parallel cases is very less (~50–60 ns), which shows the over-sensitivity of GROMOS53A6$_{CARBO\_R}$ towards helical conformation.
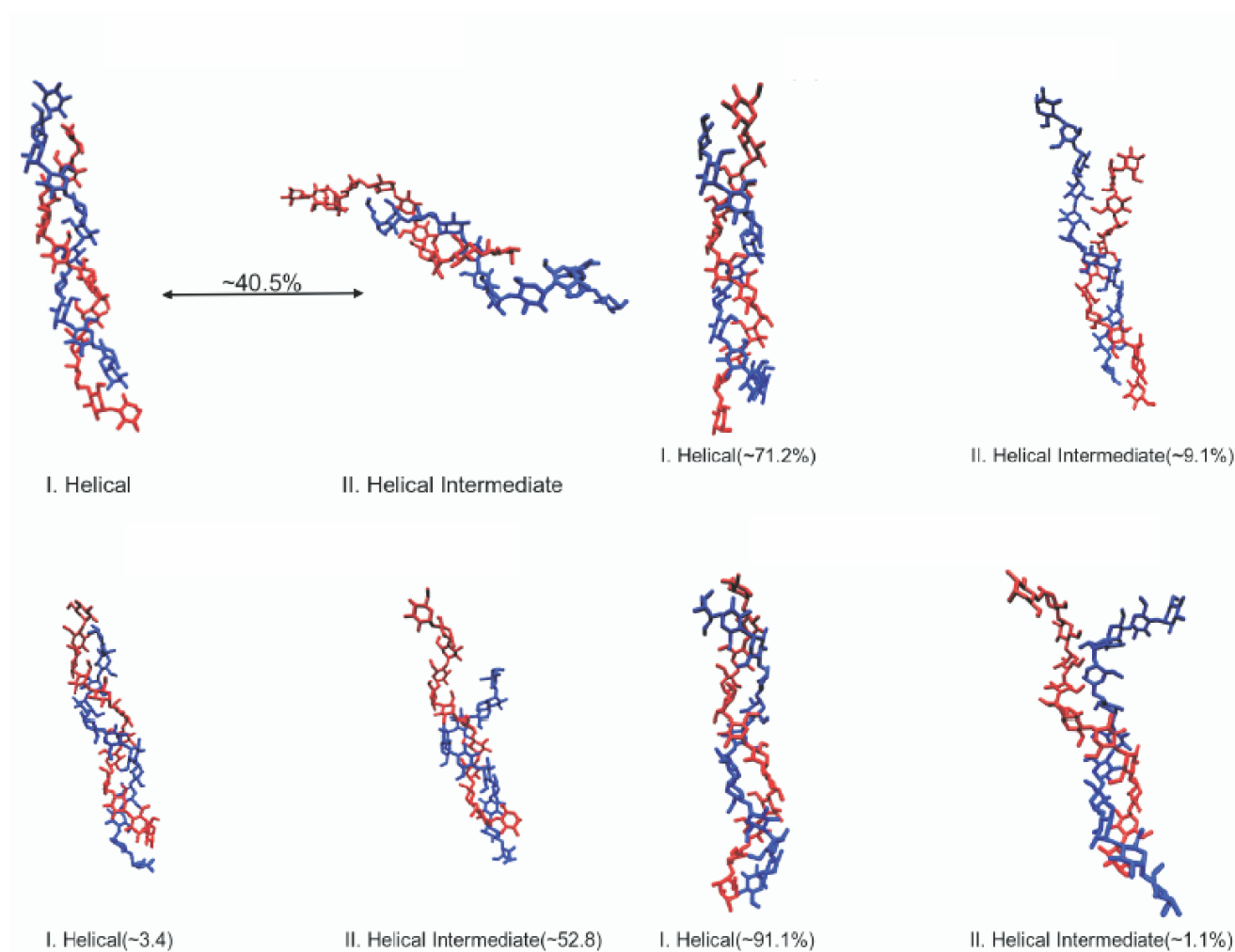
Fig. 6. The representative conformations of dimer as obtained from clustering of trajectories initiated from parallel and anti-parallel conformations. Here the snapshots for GLYCAM06 and GROMOS53A6 $_{CARBO\_R}$ force fields are being shown.

Conclusion

Due to its intrinsic flexibility, carbohydrate is computationally much less explored compared to other popular biomolecules, namely protein and lipids, leaving rooms for assessment of existing carbohydrate force field. The current work provides a detailed comparison of carbohydrate force field in their ability to simulate the amylose self-aggregation process. By utilizing the GPU-based acceleration of MD programs, we sample the conformational landscape of the carbohydrate exhaustively and these simulations are able to capture the spontaneous formation of double-helical morphology by the amylose dimers, among other conformations.

Specifically, our simulations show four carbohydrate force fields significantly differ from each other in their behavior in mimicking the amylose self-assembly process. The simulated trajectories reveal that while GROMOS53A6$_{CARBO\_R}$ and OPLS-AA force field promote fast formations of double-helix, CHARMM36 force field shows frequent assembly and disassembly of double-helical conformations. The geometrical clustering of the trajectories provides a quantitative measure of the some intermediates including partially open helix and non-helical structures. We also find that the mutual orientation of the chains has strong force field dependence. GLYCAM06 force field, on the contrary shows a fine balance in all the conformations.

References

1. R. E. Rundle and F. C. Edwards, *J. Am. Chem. Soc.*, 1943, **65**, 2200.

2. N. W. H. Cheetham and L. Tao, *Carbohydr. Polym.*, 1998, **35**, 287.

3. K. Gessler, I. Usón, T. Takaha, N. Krauss, S. M. Smith, S. Okada, G. M. Sheldrick and W. Saenger, *Proc. Natl. Acad. Sci.*, 1999, **96**, 4246.

4. J. Jacob, K. Gebler, D. Hoffmann, H. Sanbe, K. Koizumi, S. M. Smith, T. Takaha and W. Saenger, *Angew. Chem. Int. Ed.*, 1998, **37**, 605.

5. B. M. Sattelle and A. Almond, *Phys. Chem. Chem. Phys.*, 2014, **16**, 8119.

6. S. Hug, *Methods Mol. Biol.*, 2013, **924**, 127.

7. L. Monticelli and D. P. Tieleman, *Methods Mol. Biol.*, 2013, **924**, 197.

8. E. Hatcher, O. Guvench and A. D. MacKerell, *J. Phys. Chem. B*, 2009, **113**, 12466.

9. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell, *J. Chem. Theory Comput.*, 2009, **5**, 2353.

10. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622.

11. D. Kony, W. Damm, S. Stoll and W. F. Van Gunsteren, *J. Comput. Chem.*, 2002, **23**, 1416.

12. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.

13. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Intermolecular Forces*, 1981.

14. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435.

15. S. Nosé, *Mol. Phys.*, 1984, **52**, 255.

16. W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

17. M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182.

18. S. Pail and B. Hess, *Comput. Phys. Comm.s*, 2013, **184**, 2641.

19. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463.

20. S. Miyamoto and K. P. A., *J. Comput. Chem.*, 1992, **13**, 952.

21. S. Lloyd, *IEEE Trans. Inf. Theor.*, 2006, **28**, 129.