PCCP

Molecular Dynamics Simulations for Glass Transition Temperature Predictions of Polyhydroxyalkanoate Biopolymers

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-06-2020-003163.R1
Article Type:	Paper
Date Submitted by the Author:	15-Jul-2020
Complete List of Authors:	Bejagam, Karteek; Los Alamos National Laboratory, Iverson, Carl; Los Alamos National Laboratory, Chemistry Divison Marrone, Babetta; Los Alamos Naional Laboratory, Bioscience Pilania, Ghanshyam; Los Alamos National Laboratory,

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Molecular Dynamics Simulations for Glass Transition Temperature Predictions of Polyhydroxyalkanoate Biopolymers†

Karteek K. Bejagam,*[a] Carl N. Iverson,[b] Babetta L. Marrone,[c] and Ghanshyam Pilania*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Polyhydroxyalkanoates (PHAs) represent an emerging class of biosynthetic and biodegradable polyesters that exhibits a considerable potential to replace petroleum-based plastics towards a sustainable future. Despite the promise, general structure-property mappings within this class of polymers remain largely unexplored. An efficient exploration of this vast chemical space calls for the development and validation of predictive methods for accurate estimation of a diverse range of properties for PHA-based polymers. Towards this aim, here we present and validate results of our molecular dynamics (MD) simulations based approach aimed at predicting glass transition temperatures (T_g) of PHA-based polymers. Since generally-available and widely-used polymer forcefields exhibit a relatively poor performance for T_g predictions, we developed a new forcefield by modifying the polymer consistent force field (PCFF) *via* refining a selected set of torsion potentials of the polymer backbone using accurate density functional theory (DFT) computations. After carefully assessing the dependence of critical simulation parameters, such as, polymer chain length, number of polymer chains, supercell size, thermal quenching rate used in the simulation, applicability and transferability of the modified PCFF (m-PCFF) force-field is demonstrated by directly comparing the computed T_g predictions of various polymers with different chemistries, polymer side chain lengths and functional groups forming the polymer side chains against the respective experimentally measured values. Furthermore, transport properties such as self-diffusion coefficient and viscosity are computationally determined and their well-known correlations with the target property is demonstrated. Lastly, we employed the developed approach to predict T_g values for a number of yet-to-be-synthesized PHA-based polymers with a diverse set of functional groups in the polymer side chain. The results are further rationalized by correlating the predicted T_g values with the inter-chain H-bond formation tendencies of the different side chain functional groups. This work represents an important first step towards computationally guided design of PHA-based functional polymers and opens up new directions for a systematic investigation of composition- and configuration-dependent structure-property relationships in more complex binary and ternary copolymer systems.

1 Introduction

Plastics are ubiquitous in our daily life, so much so that a world without plastics is unimaginable today. The use of plastics is growing at an unprecedented rate and as of 2017 an estimated total of 8300 million metric tons of virgin plastics have been produced.[1] Only a tiny fraction of all the plastics produced (∼9 %) is currently being recycled and much of the remaining ends in landfills or oceans.[1] While plastic products exhibit many favorable properties such as lightweight, water resistance, flexibility, strength, durability, lowcost, and electrical and thermal insulation, most are non-renewable, non-degradable and extremely harmful to the ecosystem.[2,3] Thus there is a pressing need to discover alternative polymeric materials which are biodegradable and can be synthesized without taxing the environment.

Polyhydroxyalkanoates (PHAs) are one such class of biodegradable polyesters that are naturally produced by microorganisms, such as cyanobacteria, as a reserve of energy and carbon storage.[4,5] Owing to available chemical diversity within PHAs, this

[a] *Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Email: karteekbeja@lanl.gov, gpilania@lanl.gov*
[b] *Chemistry Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.*
[c] *Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.*

† Electronic Supplementary Information (ESI) available: Additional computational details and procedure for the refinement of torsion potentials were discussed. The figures illustrating the dependence of T_g on system size, polymer chain-length, and quenching rates were also provided. See DOI: 00.0000/00000000.

class of polymers can, in principle, exhibit a wide range of tunable properties such as glass transition temperature (T_g), melting temperature, and mechanical properties, including Young's modulus, tensile strength, % elongation at break, yield strength etc.[6–9] In fact, the polyesters and related classes of polymers have already shown a potential for applications in a range of diverse fields, including pharmaceutical, biomedical, and packaging industries.[10–12] However, barring a handful of specific chemistries, the vast chemical space available in PHAs remains largely unexplored and untapped.

While a high throughput exploration effort that is based entirely on an experimental trial-and-error-based route in search of PHA chemistries with targeted functionalities is unlikely to be successful, if at all practical, first principles based quantum mechanical computations still remain much too expensive to study structure-property relationships in polymer chemical spaces with candidates reaching far beyond millions. Towards this end, atomistic simulations based on classical force fields (FFs) offer an alternative route, with a manageable tradeoff between computational cost and predictive accuracy, to explore a wide range of PHA chemistries in a systematic manner. While molecular dynamics (MD) simulations based on FFs have been widely and successfully used in the past to study a variety of polymer properties for a diverse range of chemistries[13,14], the parametrized FFs that are employed to describe details of the underlying potential energy surface for a particular chemistry often suffer from poor generalizability and, therefore, are non-transferable, as such. Thus, the applicability of a given FF needs to be carefully validated before one can employ such MD simulations to make predictions on new chemistries and start exploring general chemical trends across a wide range of chemistries.

As a first step in this direction, we evaluate and compare performance of generally-available and widely-used polymer FFs towards the prediction of glass transition temperature (T_g)—a key functional property dictating the thermomechanical behavior in polymers[15]. Subsequently, develop a new forcefield by modifying the polymer consistent force field (PCFF) *via* refining a selected set of potential energy contributions describing the polymer backbone using *ab initio* density function theory (DFT) computations. We then assess the dependence of all critical simulation parameters, such as polymer chain length, number of polymer chains, supercell size, thermal quenching rate, on the predicted T_g. Further, we demonstrate the applicability and transferability of the developed FF by directly comparing the computed T_g predictions of various polymers with different chemistries, polymer side chain lengths and functional groups forming the polymer side chains against the respective experimentally measured values. As an additional validation, well known correlations of certain transport properties such as self-diffusion coefficient and viscosity with the target property are also demonstrated. Finally, we employed the developed approach to predict T_g values for a number of yet-to-be-synthesized PHA-based polymers with a diverse set of functional groups in the polymer side chains and show that the results of the simulations can be rationalized by correlating the predicted T_g with the polarizability and inter-chain H-bond formation tendencies of the different side chain functional groups.

2 Methodology

Three most commonly used FFs for organic polymer MD simulations, namely, the CHARMM General Force Field (CGenFF),[16,17] the Generalized Amber Force Field (GAFF),[18,19] and the Polymer Consistent Force Field (PCFF)[20] were considered to evaluate T_g values for a prototypical PHA polymer poly(4-hydroxybutyrate) or P4HB. The chemical structure of the polymer repeat unit for P4HB, along with all other PHA polymers considered in this study, is shown in **Figure S1** of the Supplementary Information accompanying this manuscript. Using each FF, MD simulations for P4HB in an all-atom representation were performed using Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) 30 April 2019 version.[21] An amorphous polymer box, with three dimensional periodic boundary conditions, consisting of 20 polymer chains with each chain consisting of 100 monomer units (100-mer) was generated at a density of 1.0 g/cm^3 using Enhanced Monte Carlo (EMC) software (version 9.4.4).[22] A representative snapshot of an initial configuration for P4HB generated using EMC and rendered using VMD[23] is shown in **Figure 1 a**. In each case, equations of motion were integrated with a time-step of 1.0 fs using velocity-Verlet algorithm, long-range corrections to pressure and energy were included and long-range Coulombic interactions were treated using the particle-particle particle-mesh (PPPM) method.[24] Temperature and pressure were controlled by employing Nos*é*-Hoover thermostat and barostat, respectively.[25,26] To start with, the initial configurations were equilibrated at 700 K for 4 ns to relax the potential energy (see **Figure S2**) and subsequently quenched from 700 K to 140 K at a rate of 20 K/ns in a NPT ensemble. In agreement with experimental observations of biosynthesized PHA polymers[27,28], for all polymer chemistries with chiral carbon, a monomer unit with an R-chiral center was employed for an amorphous polymer model construction. Further details of our computational methodology are provided in the Supplementary Information.

In addition to crucial simulation parameters, such as, quench rate, polymer chain length, number of polymer chains used in the simulation cell (all of which are discussed in detail in the next section), T_g predictions can also be sensitive to a specific randomly generated amorphous configuration used in an MD run. To estimate the level of uncertainly as a result of using a particular randomly generated polymer configuration, we emploed five different initial configurations of P4HB, all consisting of 20 100-mer chains in MD simulations with a quenching rate of 20 K/ns. Our results (presented in **Figure S3** in the Supplementary Information) suggest the T_g predictions for a given chemistry can be made within a standard deviation of 2.25 K, leading to a reasonable accuracy for predicting relative trends across different PHA chemistries.

To determine the transition temperature, we follow the variation in the polymer density as a function of temperature during the course of the MD simulations. As the temperature is systematically decreased, the density of the polymer supercell increases. However, the rate of change is not uniform across the entire temperature range. At the T_g, the polymer undergoes a transformation from a soft and rubbery substance to a glassy, rigid and brit-

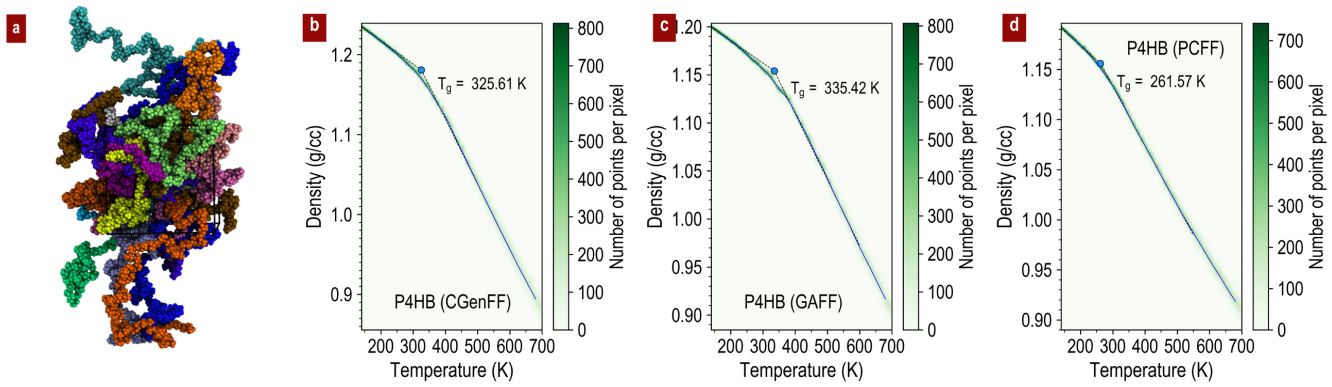


Fig. 1 (a) Snapshot of the initial configuration of an amorphous polymer box generated using Enhanced Monte Carlo (EMC)[22]. The system consists of 20 chains, each with 100 monomer units, and were colored differently for clarity. Density as a function of temperature obtained from the quenched simulations of P4HB employing **(b)** CGenFF, **(c)** GAFF, and **(d)** PCFF parameters.

tle material as the temperature is lowered.[15,29–33] During this transition, there is a dramatic increase in the mobility of polymer chains and this is reflected in temperature dependence of several physical properties, including, density, molar volume, viscosity, mechanical strength, specific heat.[34] This transition shows up as a slope change in the simulated density versus temperature curve. In the present case, density values were stored for every 0.1 ps of total simulation length of 28 ns. Next, we binned and averaged computed densities for every 5 K over the entire temperature range from 140 K to 700 K. For each simulation run, T_g was then determined by fitting straight lines to a very high accuracy (with a coefficient of determination $R^2 > 0.99$ or higher) on either side of the region of slope change.

3 Results and Discussions

3.1 T_g Predictions Using Various Force Fields

Figures 1 b,c,d present our T_g simulation results for P4HB using CGenFF, GAFF, and PCFF, respectively. A comparison of the predicted T_g values of 325.61 K (CGenFF), 335.42 K (GAFF), and 261.57 K (PCFF), with the corresponding experimentally-measured value of 225.3±2.8 K[7,35,36] shows that while CGenFF and GAFF lead to large errors in the target property (44.5 % and 48.8 %, respectively) PCFF-predicted T_g is relatively much closer to the measured value (albeit, still over estimated with an error of 16.1 % with respect to the experimental value). Given the superior performance of PCFF in the T_g prediction of P4HB relative to the other two FFs, going forward we select PCFF and take a closer look at the description of the underlying potential energy surface within this FF.

In general, classical FFs rely on various interatomic interactions of increasing complexity to describe a polymer system, for instance, two body (bond-terms), three-body (angle-terms), four-body (dihedral-terms) and non-bonded interactions.[20,37] Typically, these terms are fitted to capture behavior of either a pre-specified diverse set of polymers or a narrow group of selected chemistries and may not necessarily be transferable to other chemistries. It is, therefore, highly desirable to evaluate (and modify, if necessary) the quality of the fitted details of the potential energy surface in PCFF for PHA-based polymers, before one attempts to model chemical trends across different PHA polymers. Indeed, there are several examples in the literature where slight modifications of the potential energy terms, for a given FF was necessary for an improved description of a target polymer property.[38–40] In particular, four-body torsion potential for the polymer backbone chain can play a critical role in determining the polymer configuration and, thus, its properties.[41–43] In fact, McAliley *et. al.* have demonstrated that the improvement of the torsion potentials in a FF can significantly enhance the accuracy in predicting T_g for polylactic acid.[40] Towards this end, we start by comparing different potential energy profiles for P4HB computed using PCFF with the corresponding ones obtained using more accurate first principles density functional theory (DFT) calculations.

3.2 Refinement of Torsion Potentials

Figures 2 compares potential energy scans, in the case of P4HB, for different polymer backbone dihedrals computed using the PCFF and DFT computations. The DFT calculations were performed with Gaussian-16[44] package employing a B3LYP/6-31+g(d,p) exchange correlation potential.[45–48] A trimer model, shown in **Figure S4** of Supplementary Information, was used in generating the potential energy profiles via geometry optimization of the structure while constraining a specific dihedral angle in discrete steps ranging from -180° to 180°, in steps of 10°. The relaxed geometries were subsequently used as an input to obtain the torsion scan energy profiles using PCFF in LAMMPS.[21] Interestingly, our results show that potential energy profiles for dihedrals including the ketone carbon (C=O) in the polymer backbone exhibit the largest discrepancies when comparing the FF results (gray squares) with those computed using DFT (blue circles), as shown in **Figures 2 a,b**). In particular, the PCFF-computed potential energy profile for the -CH_2-CH_2-C(=O)-O- motif in the backbone, shown in **Figure 2 a**, presents a qualitatively incorrect behavior when compared to that from the DFT calculations. To address this discrepancy, we use the DFT-computed potential energy profiles to refit the torsional parameters of the

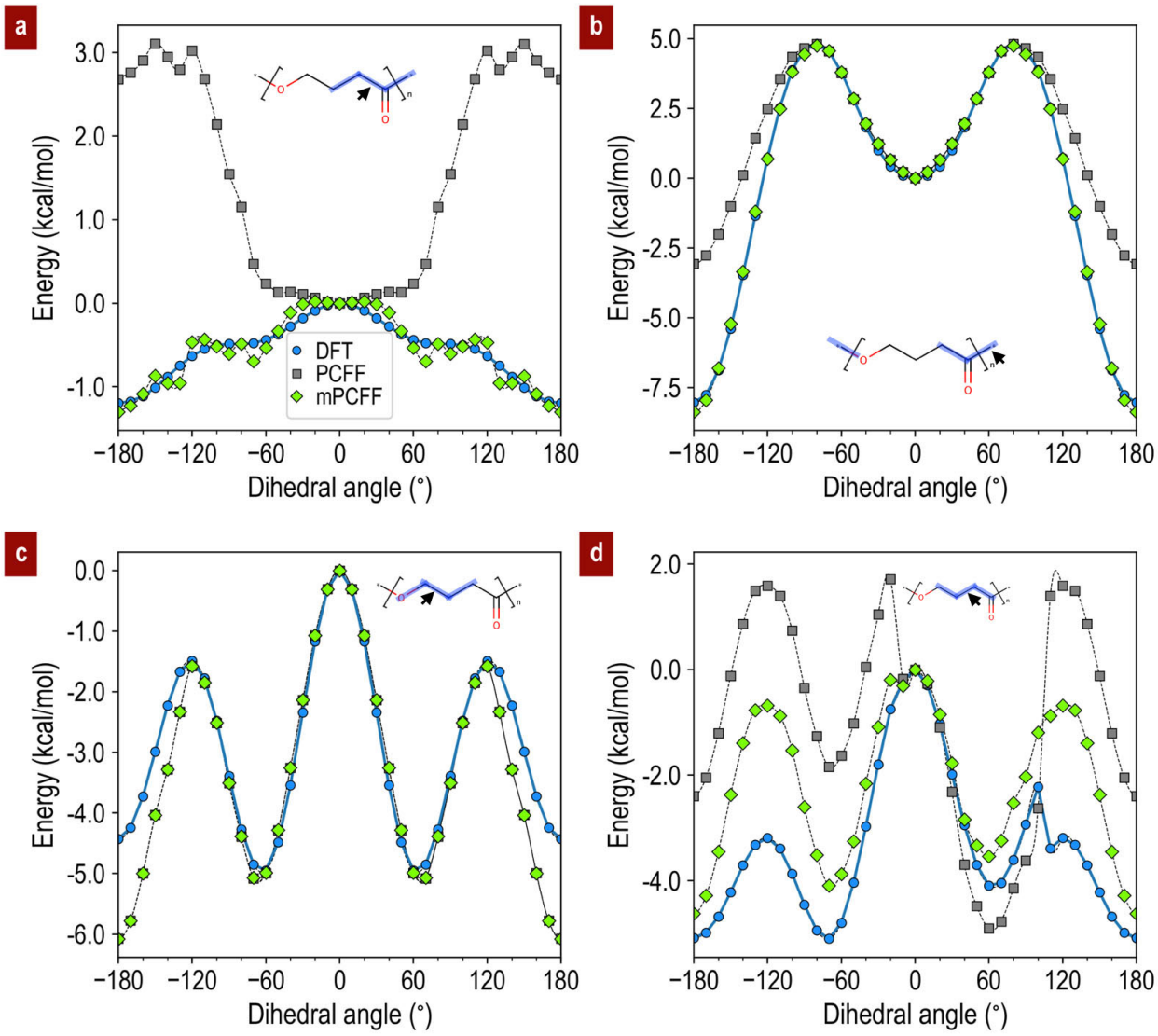


Fig. 2 Potential energy scan profiles for different polymer backbone dihedrals in P4HB computed using DFT (blue circles), PCFF (gray squares), and mPCFF (green diamonds). DFT calculations were performed at B3LYP/6-31+g(d,p) level of theory. The highlighted blue region of the polymer repeat unit shown in each panel **(a-d)** refers to the specific dihedral considered in each case. In each case, the energies of the potential energy scans are referenced to the respective 0° dihedral angle values.

two dihedrals in the vicinity of the backbone ketone group in PCFF (*cf.*, **Figures 2 a,b**). Hereinafter, the modified PCFF is referred to as mPCFF. While details of the adopted fitting procedure are provided in the Supplementary Information, it can be seen from **Figure 2** that the potential energy profiles computed with the mPCFF show a good agreement with corresponding DFT-computed energetics.

3.3 Influence of Simulation Parameters on T_g Predictions

As briefly alluded to in the previous section, the prediction of a polymer's T_g using atomistic simulations can sensitively depend on a number of simulation parameters, such as, number of polymer chains considered in a simulation, polymer chain-length, and thermal quenching rate.[33,40,55] Therefore, before quantifying the improvement achieved in the prediction of T_g for P4HB and other PHA-based polymers with mPCFF (as compared to PCFF), we first carry out a careful study of the dependence and sensitivity of the predicted T_g with respect to these simulation parameters for P4HB, employing mPCFF. More specifically, the following simulation parameter spaces were explored: **(i)** The number of polymer chains in the simulation box were varied as 5, 10, 20, 50, and 100, while keeping the the polymer chain-length and thermal quench rate fixed to 100-mer and 20 K/ns, respectively. **(ii)** To study the effect of polymer chain-length, polymer chain lengths (number of chains in the simulation cell) of 10-mer (200), 30-mer (60), 50-mer (40), 100-mer (20), and 500-mer (5) were simulated at a quench rate of 20 K/ns. Note that the number of chains was varied here in order to have an approximately equal number of total atoms in each of these simulation runs. **(iii)** The dependence of thermal quench rate at a fixed system size and polymer

Table 1 Influence of the simulation parameters on T_g predictions. P4HB polymer chains employing mPCFF parameters were considered for this study.

Number of chains	Chain-length	Quench rate (K/ns)	T_g (K)
	System size		
5	100-mer	20	251.56
10	100-mer	20	252.72
20	100-mer	20	253.16
50	100-mer	20	254.83
100	100-mer	20	256.01
	chain-length		
200	10-mer	20	245.78
60	30-mer	20	258.17
40	50-mer	20	266.93
20	100-mer	20	253.16
5	500-mer	20	255.75
	quench rate		
20	100-mer	10	242.91
20	100-mer	15	252.88
20	100-mer	20	253.16
20	100-mer	50	265.80
20	100-mer	100	270.64

chain length (*i.e.*, 20 chains of 100-mer polymers), where the quench rates of 10 K/ns, 15 K/ns, 20 K/ns, 50 K/ns, and 100 K/ns were considered. The results of predicted T_g values as a function of these simulation parameter sets are summarized in **Table 1** and presented in further detail in the Supplementary Information (**Figures S5-S7**).

The results presented in **Table 1** indicate that for a fixed set of simulation parameters, increasing the number of chains or increasing the quench rate leads to a systematic increase in the predicted T_g value, which is in agreement with previous MD studies in the literature.[40,56,57] Furthermore, simultaneously increasing the chain-length while decreasing the number of chains (to maintain approximately same number of total atoms across the MD simulation runs) leads to a non-monotonic trend in the predicted T_g. This can be explained as a combined effect of the aforementioned trend with respect to the number of chains and the well known Fox-Flory relation capturing T_g dependence with respect to the polymer chain length.[33,55]

Guided by the results of these simulations, for subsequent simulations we choose to consistently employ a polymer supercell containing 20 chains, each constructed with 100 monomer units, and a quenching rate of 20 K/ns,. Note that for accurate predictions, ideally one would like to employ a simulation cell as large as possible with very small thermal quench rates nearing those typically used in experiments (*i.e.*, $<$ 1K/s). However this would be prohibitively computationally expensive and impractical, even with state-of-the-art computational resources. The particular choice of the simulation parameters was largely dictated by this cost-accuracy tradeoff. These simulation parameters were consistently used for all the T_g simulations performed in the present study, including the results reported in **Figure** 1.

Lastly, we note that although the thermal quench rate and system size employed in the present MD simulations significantly differ from those typically used in an experimental setting, due to the systematic trends and conflicting relations of T_g with these simulation parameters—a small system size underestimates T_g, while a higher quench rate leads to an over estimation—a favorable error cancellation is expected. As a result, computed qualitative trends across different PHA chemistries are still expected to be both meaningful and insightful.

3.4 T_g predictions for PHAs with different chemistries

As expected, use of mPCFF leads to a predicted T_g of 253.16 K for P4HB with a 12.4 % error as compared to the corresponding experimentally-measured value (as compared to the value of 261.57 K with the 16.1% error obtained with PCFF), showing a further improvement in the prediction accuracy as a result of an improved description of the torsional parameters in the modified force field. More importantly, however, we find that mPCFF is able to predict correct relative trends in T_g values when comparing different chemistries, while the original PCFF fails. For instance, as reported in **Table** 2, a comparison of predicted T_g val-

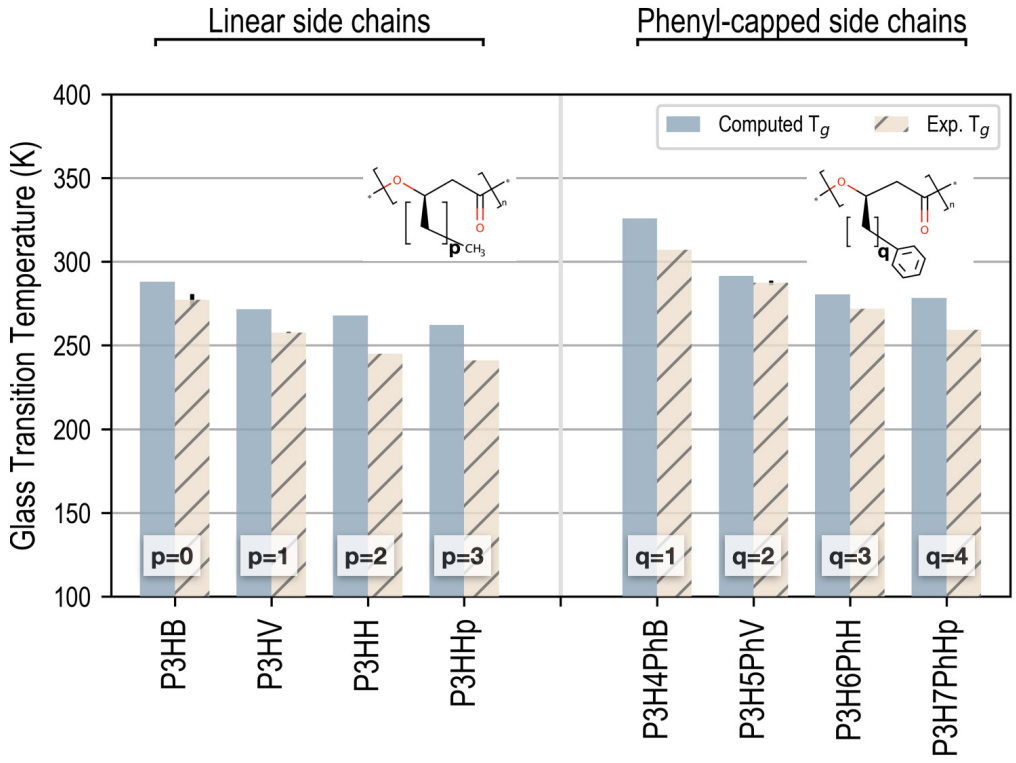


Fig. 3 Comparison of predicted T_gs for PHAs with different chemistries using mPCFF parameters along with their corresponding experimental values. The amorphous polymer box consisted of 20 chains, each with 100 monomer units and temperature was quenched at 20 K/ns to determine the T_g. References for the experimental values are 7,8,35,36,49–54.

Table 2 T_g predictions of all the PHAs studied herein employing mPCFF parameters. In all the simulations, the number of chains was 20, the polymer chain was 100 monomer units, and a quenching rate of 20 K/ns was used. The values in the parentheses were obtained using PCFF parameters. The standard deviation in experimental values are shown in parentheses for the available data. All the T_g's are in K.

Chemical name	abbreviation	Computed T_g	Experimental T_g
poly(4-hydroxybutyrate)	P4HB	253.16 (261.57)	225.3 (2.8) [7,35,36]
poly(3-hydroxybutyrate)	P3HB	287.91 (306.66)	277.1 (3.5) [8,35,49–53]
poly(3-hydroxyvalerate)	P3HV	271.46	257.7 (0.5) [35,52]
poly(3-hydroxyhexanoate)	P3HH	267.8	245.0 [35]
poly(3-hydroxyheptanoate)	P3HHp	262.19	241.0 [35]
poly(3-hydroxy-4-phenylbutyrate)	P3H4PhB	325.94	307.0 [8]
poly(3-hydroxy-5-phenylvalerate)	P3H5PhV	291.36 (284.34)	287.4 (1.4) [8,51,54]
poly(3-hydroxy-6-phenylhexanoate)	P3H6PhH	280.36	271.9 [51]
poly(3-hydroxy-7-phenylheptanoate)	P3H7PhHp	278.33	259.2 [8]

ues for different PHA polymers, namely, poly(4-hydroxybutyrate) (P4HB), poly(3-hydroxybutyrate) (P3HB) and poly(3-hydroxy-5-phenylvalerate) (P3H5PhV) using PCFF and mPCFF shows that while mPCFF is able to correctly predict the experimentally observed relative trends in the T_g values across these polymers (i.e., $T_g^{P4HB} < T_g^{P3HB} < T_g^{P3H5PhV}$) and always systematically overestimates the predicted values with respect to the experimental values, the predictions using PCFF do not follow any systematic trend. For instance, using PCFF the predicted T_g of 284.34 K for P3H5PhV is lower than the experimental value of 287.4 K, while for P3HB the predicted T_g (306.66 K) is significantly higher than that of the experimental observation of 277.1 K.

To further demonstrate the predictive power of mPCFF in predicting relative variation in T_g values of PHA-based polymers, we choose two groups of polymer composition with systematically varying side chain functional group lengths for which accurate

T_g measurements are available. The polymers in both the groups can be represented with -CH(X)-CH_2-C(=O)-O- as the polymer backbone repeat unit with the group 'X' representing a side chain functional group. The two groups are comprised of four linear side chain polymers (with X = -$(CH_2)_p$-CH_3, where $p \in [0, 3]$) and four phenyl-capped linear side chain polymers (with X = -$(CH_2)_q$-C_6H_5 where $q \in [1, 4]$) and exhibit a gradual variation in T_g with respect to the polymer side chain length.[9]

Table 2 and **Figure** 3 compare the computed T_g values for the eight polymers with their corresponding experimental measurements. Note that in **Figure** 3, regardless of the functional group, as the polymer side chain length increases, the measured T_g decreases. Furthermore, for a given number of aliphatic carbons in the side chain, the polymers with a phenyl-capped side chain exhibit a slightly larger T_g. It can be seen from **Figure** 3 that both of these trends are well captured in the computed T_g values. As noted earlier, while the computed T_g is consistently overestimated with respect to the measured value, when focusing on the relative trends in T_g across these chemistries, there is an excellent agreement with the measured values (see **Figure S8** of Supplementary Information).

3.5 Transport Properties

Lastly, to test the well-known and physically-meaningful variation of T_g with other dynamical properties, we consider self-diffusion coefficient (D) and viscosity (η) as two key properties for three different polymers, namely, P4HB, P3HB and P3H5PhV. The specific choice of these three polymers was largely motivated by the fact that these cover a broad range of T_g values, ranging from a low T_g of 225.3 K for P4HB to a relatively higher value of 287.4 K for P3H5PhV. Equilibrium simulations were performed at a temperature of 400 K, where the three polymers are significantly above their T_g and therefore exist in a flexible rubbery state. To determine the self-diffusion coefficient and viscosity, the system was equilibrated in an isothermal isobaric ensemble (NPT) for 10 ns followed by a production run of 50 ns using an NVT ensemble. To reduce the fluctuations, mean-square displacements (MSDs) associated with the center of mass of each polymer chains and viscosities were determined up to a correlation length of 20 ns. The self-diffusion coefficient of each polymer system was then evaluated as one-sixth of the slope of the MSD versus time curve in the diffusive regime (*i.e.*, MSD $\propto$ t) using the Einstein relation, as shown below in **Equation** 1:[58,59]

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\left\langle \left(r_i(t) - r_i(0)\right)^2 \right\rangle. \quad (1)$$

Here $r_i(t)$ represents the position vector of the center of mass for a polymer chain i at a time t and angular bracket is used to denote an ensemble average over all the polymer chains.

In addition to the self-diffusion coefficients, the pressure tensor values for each of the three polymers were stored at 1 fs intervals and subsequently utilized in the Green-Kubo relation, given by **Equation** 2 below, to determine the shear viscosity η from the equilibrium simulations.[60,61]

$$\eta = \frac{V}{k_B T}\int_0^\infty \left\langle \tilde{P_{xy}}(0)\tilde{P_{xy}}(t)\right\rangle dt \quad (2)$$

Here V represents the volume of the simulation cell, k_B is Boltzmann constant, T is the simulation temperature, $\langle ... \rangle$ is ensemble average and $\tilde{P_{xy}}$ is used to denote the off-diagonal element of the stress tensor. Note that the integral in **Equation** 2 over a longer time yields the shear viscosity (η) via computing the stress-stress time autocorrelation. This autocorrelation generally decays rather slowly with time, leading to a plateau region. Thus, here we performed simulations for sufficiently long times to determine the reported shear viscosity results.

The results of our simulations for the three polymers are presented in **Figures** 4 (a) and (b). The computed self-diffusion coefficient and viscosity values for P4HB, P3HB, and P3H5PhV are 1.08, 0.71, and 0.13 (in $10^{-12}m^2/s$), and 0.015, 0.109, and 1.34 (in Pa.s units), respectively. These obtained results are well aligned with the intuitive notion that the lower the T_g (relative to the simulation temperature), the more dynamic the polymer becomes, exhibiting a relatively higher diffusion coefficient and a lower viscosity. Further, following the Stokes-Einstein relation[62] the self-diffusion coefficient and viscosity are inversely proportional to each other.

4 T_g predictions on Yet-to-be Synthesized PHAs

While the development and validation of an MD-based atomistic simulation route for the estimation of T_g in the vast PHA polymer chemical space is encouraging, the true potential of such an approach lies in making predictions on hitherto unknown and yet-to-be synthesized polymers. As an illustration, here we choose five different polymers, composed with a common backbone, and distinguished by a set of distinct functional groups as the polymer side chain. For the choice of side chain functional groups, we select carbonyl (-CHO), amine (-NH_2) hydroxyl (-OH), carboxyl (-COOH), and benzoyl (-COPh) groups, as shown in **Figure** 5a. The results of the MD simulations for T_g predictions—employing mPCFF and following the consistent methodology and simulation parameters as established in this work—are presented in **Figure** 5b. The results clearly suggest a strong dependence of T_g on the chemical nature of the side chain functional group and the predicted trend can be qualitatively understood in terms of strengthening of the polymer interchain interactions due to a polar side chain functional group, which leads to a relatively higher T_g. When comparing different functional groups with approximately similar size, as the polarity of the side chain pendent group systematically increases, interchain electrostatic and Hydrogen bonding (H-bonding) interactions gradually grow stronger. As a result of these interactions, neighboring polymer chains are locked strongly in an amorphous structure and are able to push the dynamical transition to relatively higher temperatures, and thus, leading to an increased T_g. We further note here that while PCFF is known to capture non-bonding interactions, including H-bonding interactions[63,64], we also explicitly demonstrate this capability for the mPCFF in case of poly(3 carboxy propionate) or P3CoxyP. **Figure S9** in Supplementary Information clearly ex-

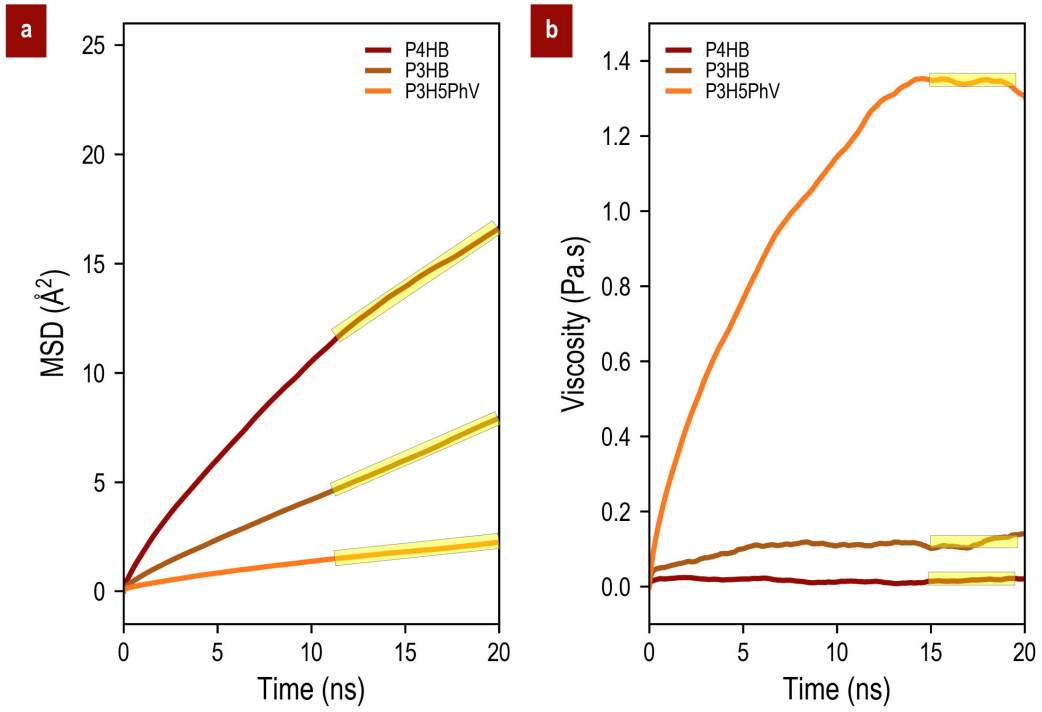


Fig. 4 Comparison of transport properties for three polymer systems in a rubbery state above their T_g. **(a)** Mean-square displacement, **(b)** Viscosity. Simulation cell consists of 20 chains, each with 100 monomer units and equilibrium simulations were carried out at 400 K for 50 ns in an NVT ensemble. The diffusive regions and the plateau regions are highlighted for the three polymer chemistries in (a) and (b), respectively.

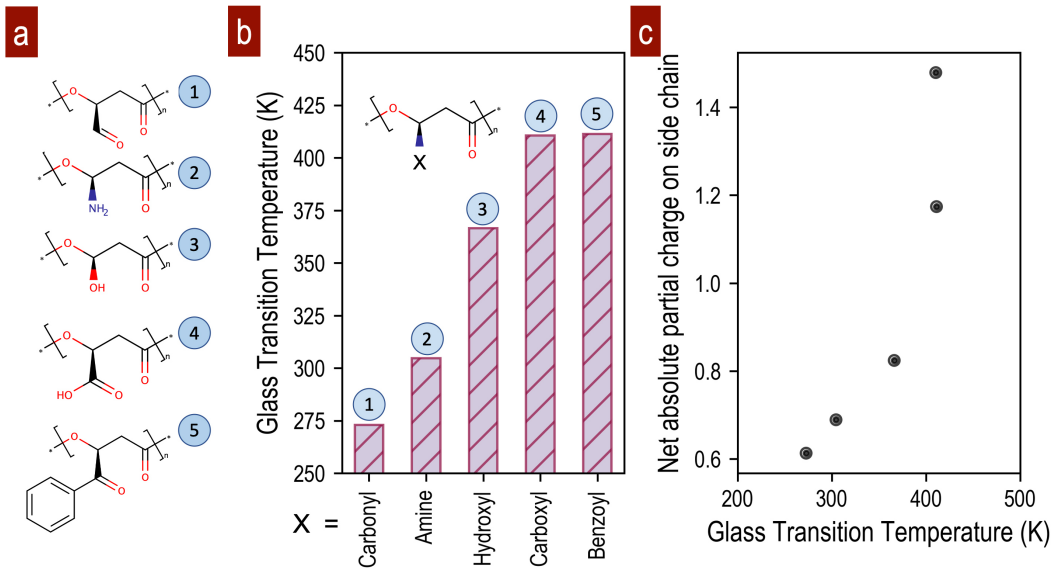


Fig. 5 (a) PHA motifs with five side chain functional groups, (b) T_g predictions from the MD simulations employing mPCFF parameters, (c) Correlation between glass transition temperature (T_g) and net absolute partial charge on the side group.

hibits the presence of a significant level of H-bonding interactions between the carboxyl side chain functional groups, both below (400 K) and above (700 K) the transition temperature. Further, the extent of the interchain H-bonding interactions at the lower temperature is found to be about twice as much as compared to the higher temperature.

To further quantify these chemically intuitive notions, we compute the total absolute atomic Gasteiger partial charges[65,66] for each of the five functional groups as a measure of the group polarity and correlate with the computed T_g values. As shown in **Figure** 5c, there is a strong correlation between the group polarity and T_g, providing a rationale for the predicted trends in the target property.

5 Conclusions

We have developed a systematic approach based on atomistic force-field based MD simulations, that can accurately predict T_g values of PHA-based biopolymers. To accomplish this, we developed a new forcefield by modifying the polymer consistent force field *via* refining a selected set of torsion potentials of the polymer backbone using accurate DFT computations. We then carefully assessed the dependence of critical simulation parameters, such as, polymer configurations, polymer chain length, number of polymer chains, supercell size, thermal quenching rate used in the simulation on the computed T_g values. Next, the applicability and transferability of the designed force-field was demonstrated by directly comparing the computed T_g predictions of various polymers with different chemistries, polymer side chain lengths and functional groups forming the polymer side chains against the respective experimentally measured values. Furthermore, transport properties such as self-diffusion coefficient and viscosity were computationally determined and their well-known correlations with the target property is demonstrated. Lastly, we employed the developed and validated approach to predict T_g values for a number of yet-to-be-synthesized PHA-based polymers with a diverse set of functional groups in the polymer side chain. The results were further rationalized by correlating the predicted T_g values with side chain functional group polarity and inter-chain H-bond formation tendency for a fixed polymer backbone motif. This approach can easily be extended to polymer properties beyond T_g and can potentially open new avenues for a high throughout exploration in PHA-based biodegradable chemistries with targeted functionalities. Upon success, such a computational approach can allow for identification and screening of novel polymer chemistries offering an optimal balance between durability and environmental degradability for various applications, providing a potential sustainable solution for the daunting plastic problem.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Authors gratefully acknowledge the support from Laboratory Directed Research and Development program of Los Alamos National Laboratory under project # 20190001DR. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). Computational support for this work was provided by LANL's high performance computing clusters.

References

1 R. Geyer, J. R. Jambeck and K. L. Law, *Science Advances*, 2017, **3**, e1700782.
2 M. Haward, *Nature Communications*, 2018, **9**, 667.
3 C. Wilcox, E. Van Sebille and B. D. Hardesty, *Proceedings of the National Academy of Sciences*, 2015, **112**, 11899–11904.
4 R. Verlinden, D. Hill, M. Kenward, C. Williams and I. Radecka, *Journal of Applied Microbiology*, 2007, **102**, 1437–1449.
5 S. Chanprateep, *Journal of Bioscience and Bioengineering*, 2010, **110**, 621 – 632.
6 Z. A. Raza, S. Abid and I. M. Banat, *International Biodeterioration & Biodegradation*, 2018, **126**, 45 – 56.
7 D.-C. Meng, Z.-Y. Shi, L.-P. Wu, Q. Zhou, Q. Wu, J.-C. Chen and G.-Q. Chen, *Metabolic Engineering*, 2012, **14**, 317 – 324.
8 S. Mizuno, S. Katsumata, A. Hiroe and T. Tsuge, *Polymer Degradation and Stability*, 2014, **109**, 379 – 384.
9 G. Pilania, C. N. Iverson, T. Lookman and B. L. Marrone, *Journal of Chemical Information and Modeling*, 2019, **59**, 5013–5025.
10 J. C. Middleton and A. J. Tipton, *Biomaterials*, 2000, **21**, 2335 – 2346.
11 H. Seyednejad, A. H. Ghassemi, C. F. van Nostrum, T. Vermonden and W. E. Hennink, *Journal of Controlled Release*, 2011, **152**, 168 – 176.
12 C. W. Pouton and S. Akhtar, *Advanced Drug Delivery Reviews*, 1996, **18**, 133 – 162.
13 K. K. Bejagam, G. Fiorin, M. L. Klein and S. Balasubramanian, *The Journal of Physical Chemistry B*, 2014, **118**, 5218–5228.
14 M. O. Steinhauser, *The Journal of Chemical Physics*, 2005, **122**, 094901.
15 T. Nonoyama, Y. W. Lee, K. Ota, K. Fujioka, W. Hong and J. P. Gong, *Advanced Materials*, **32**, 1905878.
16 K. Vanommeslaeghe and A. D. MacKerell, *Journal of Chemical Information and Modeling*, 2012, **52**, 3144–3154.
17 K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell, *Journal of Chemical Information and Modeling*, 2012, **52**, 3155–3168.
18 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157–1174.
19 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *Journal of Molecular Graphics and Modelling*, 2006, **25**, 247 – 260.
20 H. Sun, *The Journal of Physical Chemistry B*, 1998, **102**, 7338–7364.
21 S. Plimpton, *Journal of Computational Physics*, 1995, **117**, 1 – 19.
22 P. J. in 't Veld and G. C. Rutledge, *Macromolecules*, 2003, **36**, 7358–7365.
23 W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics*, 1996, **14**, 33–38.
24 R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, Taylor & Francis, Inc., USA, 1988.
25 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.
26 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
27 M. Winnacker, *European Journal of Lipid Science and Technology*, 2019, **121**, 1900101.
28 M. S. Reeve, S. P. McCarthy and R. A. Gross, *Macromolecules*, 1993, **26**, 888–894.
29 Y. Zhang, R. D. Adams and L. F. M. da Silva, *The Journal of Adhesion*, 2013, **89**, 785–806.

30 J. Han, R. H. Gee and R. H. Boyd, *Macromolecules*, 1994, **27**, 7781–7784.

31 K.-q. Yu, Z.-s. Li and J. Sun, *Macromolecular Theory and Simulations*, 2001, **10**, 624–633.

32 S. Yang and J. Qu, *Polymer*, 2012, **53**, 4806 – 4817.

33 J. Zhang, Y. Liang, J. Yan and J. Lou, *Polymer*, 2007, **48**, 4900 – 4905.

34 J. Meyer, *Polymer Engineering & Science*, 1973, **13**, 462–468.

35 H.-h. Wang, X.-r. Zhou, Q. Liu and G.-Q. Chen, *Applied Microbiology and Biotechnology*, 2011, **89**, 1497–1507.

36 D. P. Martin and S. F. Williams, *Biochemical Engineering Journal*, 2003, **16**, 97 – 105.

37 K. K. Bejagam, G. Fiorin, M. L. Klein and S. Balasubramanian, *The Journal of Physical Chemistry B*, 2014, **118**, 5218–5228.

38 P. Friederich, M. Konrad, T. Strunk and W. Wenzel, *Scientific Reports*, 2018, **8**, 2559.

39 K. H. DuBay, M. L. Hall, T. F. Hughes, C. Wu, D. R. Reichman and R. A. Friesner, *Journal of Chemical Theory and Computation*, 2012, **8**, 4556–4569.

40 J. H. McAliley and D. A. Bruce, *Journal of Chemical Theory and Computation*, 2011, **7**, 3756–3767.

41 M. Bulacu, N. Goga, W. Zhao, G. Rossi, L. Monticelli, X. Periole, D. P. Tieleman and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2013, **9**, 3282–3292.

42 M. Bulacu and E. van der Giessen, *The Journal of Chemical Physics*, 2005, **123**, 114901.

43 M. Bulacu and E. van der Giessen, *Phys. Rev. E*, 2007, **76**, 011807.

44 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, 2016, Gaussian Inc. Wallingford CT.

45 J. W. G. Bloom and S. E. Wheeler, *Journal of Chemical Theory and Computation*, 2014, **10**, 3647–3655.

46 K. H. DuBay, M. L. Hall, T. F. Hughes, C. Wu, D. R. Reichman and R. A. Friesner, *Journal of Chemical Theory and Computation*, 2012, **8**, 4556–4569.

47 A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648–5652.

48 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

49 C.-K. Kang, S. Kusaka and Y. Doi, *Biotechnology Letters*, 1995, **17**, 583–588.

50 A. Czerniecka-Kubicka, W. Frącz, M. Jasiorski, W. Błażejewski, B. Pilch-Pitera, M. Pyda and I. Zarzyka, *Journal of Thermal Analysis and Calorimetry*, 2017, **128**, 1513–1526.

51 M. Ishii-Hyakutake, S. Mizuno and T. Tsuge, *Polymers*, 2018, **10**, 1267.

52 X.-W. Shen, Y. Yang, J. Jian, Q. Wu and G.-Q. Chen, *Bioresource Technology*, 2009, **100**, 4296 – 4299.

53 Z. Wei, L. Liu and M. Qi, *Reactive and Functional Polymers*, 2006, **66**, 1411 – 1419.

54 K. Fritzsche, R. W. Lenz and R. C. Fuller, *Die Makromolekulare Chemie*, 1990, **191**, 1957–1965.

55 K. O'Driscoll and R. A. Sanayei, *Macromolecules*, 1991, **24**, 4479–4480.

56 Q. Yang, X. Chen, Z. He, F. Lan and H. Liu, *RSC Adv.*, 2016, **6**, 12053–12060.

57 D. Hossain, M. Tschopp, D. Ward, J. Bouvard, P. Wang and M. Horstemeyer, *Polymer*, 2010, **51**, 6071 – 6083.

58 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, USA, 1989.

59 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, Inc., USA, 1st edn, 1996.

60 C. J. Mundy, J. I. Siepmann and M. L. Klein, *The Journal of Chemical Physics*, 1995, **102**, 3376–3380.

61 M. Mondello and G. S. Grest, *The Journal of Chemical Physics*, 1997, **106**, 9327–9336.

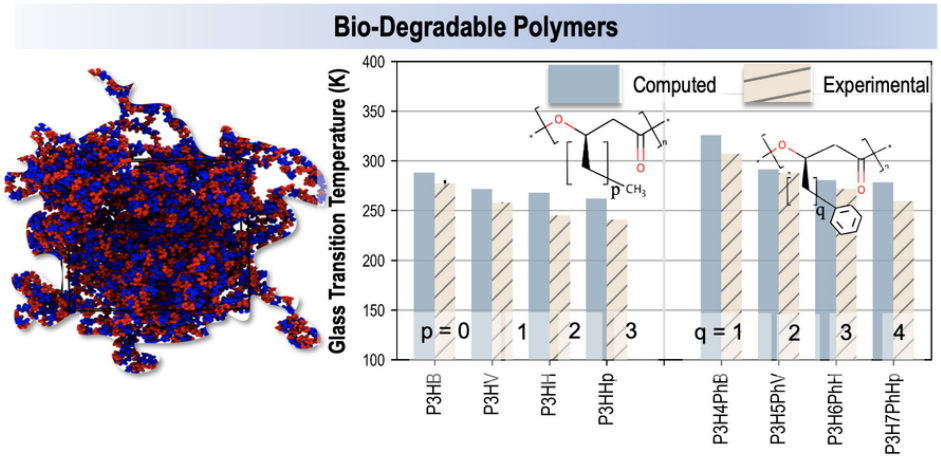
62 J. Brillo, A. I. Pommrich and A. Meyer, *Phys. Rev. Lett.*, 2011, **107**, 165902.

63 W. Chen, G. C. Lickfield and C. Q. Yang, *Polymer*, 2004, **45**, 1063 – 1071.

64 M. Chen, B. Coasne, R. Guyer, D. Derome and J. Carmeliet, *Nature Communications*, 2018, **9**, 3507.

65 J. Gasteiger and M. Marsili, *Tetrahedron*, 1980, **36**, 3219 – 3228.

66 G. Landrum, *RDKit: Open-source cheminformatics http://www.rdkit.org*.



79x39mm (300 x 300 DPI)