JCTC Journal of Chemical Theory and Computation

Open Access
This article is licensed under CC-BY 4.0

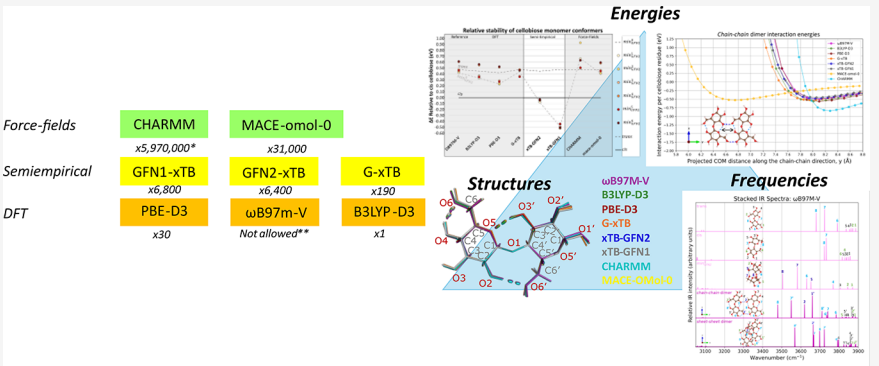
pubs.acs.org/JCTC Article

Intra- and Intermolecular Interactions in Cellobiose: Benchmarking Accelerated Computational Methods for Cellulose

Felicia Andersson,* Shokirbek Shermukhamedov, Valentina Basoli, Fabiana Arduini, Kersti Hermansson,* and Peter Broqvist*

Cite This: J. Chem. Theory Comput. 2026, 22, 12223−12238 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

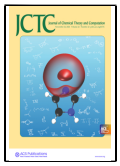


ABSTRACT: Access to accelerated computational methods at the electronic and atomistic scales is important in materials modeling, as it will allow more realistic system descriptions. Against this background, a systematic study of a range of electronic/atomistic methods applied to cellobiose monomers and dimers is presented, with the aim of identifying methods that achieve a significant acceleration of cellulose computations compared to hybrid DFT methods while retaining adequate accuracy. We benchmark such methods, namely the semiempirical quantum-mechanical tight-binding GFN*n*-xTB (n = 1,2) and G-xTB methods, the MACE-OMol-0 machine-learning potential, and the CHARMM36 force field, against three DFT references (ωB97M-V, B3LYP-D3, PBE-D3) for gas-phase cellobiose monomers and dimers, focusing on structures, energies, and vibrational properties. Cellobiose monomers and dimers, as presented in this work, do not represent the structural complexity of cellulose, but they provide tractable structural models for assessing the qualities of the selected computational models regarding cellobiose−cellobiose intermolecular interactions that are also present in cellulose. We find that (i) the *chain*−*chain* and *sheet*−*sheet* dimer interactions are of comparable magnitude, suggesting that both intermolecular dispersion and hydrogen bonding make important contributions to dimer stability, so accurate treatment of noncovalent interactions the calculations is essential; (ii) GFN2-xTB and G-xTB reproduce most conformations well, with a speed-up of two to three orders-of magnitude compared to the DFT calculations, but fail for puckered, strained conformations, artificially improving apparent dimer-energy agreement with DFT; (iii) MACE reaches near-DFT accuracy with a further order of magnitude speed-up, although in a few cases fails to reproduce the hybrid DFT conformations. This is especially true for the chain−chain dimer structure, which is also reflected in a significant deviation of the *chain*−*chain* energy-distance potential curve, which highlights the importance of system-dependent model validation (as was done here) and suggests a need for model fine-tuning for cellulose systems; (iv) CHARMM36 has some conformational and energetic limitations,but being 30,000× faster than G-xTB (itself orders of magnitude faster than DFT), remains practical for qualitative studies.

1. INTRODUCTION

Cellulose is a pivotal material in numerous applications in materials science and related fields owing to its accessible, sustainable, biodegradable, and renewable nature[1] along with its exceptional properties, such as high tensile strength and elasticity.[2] Needless to say, cellulose itself and its processed products, such as paper and clothing, are materials of significant global utility and economic value, certain to play increasingly important roles in many areas. Two interesting novel applications are paper-based electrochemical devices[3,4] and as a growth medium for tissue engineering,[5−7] which enable sustainable drug analysis and residue monitoring for precision medicine applications. While many of the remarkable

Received: July 9, 2026
Revised: September 15, 2026
Accepted: October 1, 2026
Published: November 6, 2026



ACS Publications
© 2026 The Authors. Published by American Chemical Society

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

properties of cellulose originate from the nature of its micro/nanofiber building blocks and their responses to external conditions, including their surrounding medium,[8] it should be pointed out that these properties also often have their origin in microscopic processes at the molecular level. Computational approaches at the electronic and atomistic scales and multiscale approaches play key roles in gathering information about nanoscale properties of cellulose.

However, electronic and atomistic simulations for paper-based materials and cellulose are challenging, both because of the lack of detailed structural information available from experiments and the enormous complexity of these materials, which requires long time scales for statistically converged results. The balancing of computational efficiency and accuracy is challenging; see, for example, ref 9 for a review. In this context, new modeling techniques that are both fast and capable of capturing the essential chemistry of cellulose are needed. One recent strategy, and a quickly expanding research field, is the generation and use of machine-learned interatomic potentials (MLIPs, i.e., machine-learned "force fields"), constructed with the aim of capturing all the essential interactions of an underlying, often quantum-mechanical, method, but at a much lower computational cost (see, for example ref 10). We present such results in the current study, namely from the recently developed family of Message Passing Atomic Cluster Expansion (MACE) foundation models.[11] We are not aware of any applications of MACE to cellulose systems in the current literature.

The MACE models are advanced and elaborate force fields, and like all force field expressions, they only mimic the presence of the electrons. Classical Molecular Dynamics (MD) simulations in the literature are traditionally based on much simpler, often physics-based, mathematical expressions. Despite their sometimes limited accuracy and their difficulties in treating, for example, bond breaking reliably, they are widely used in the cellulose community due to their efficiency and direct link to experimental properties. Classical force fields have been widely used to explore the structure and dynamics of cellulose in condensed matter[12] as well as the structure, dynamics, and dissolution of cellobiose.[13−16] Commonly used force fields include CHARMM36,[17] GLYCAM,[18] GROMOS,[19] and OPLS.[20] These force fields differ in their parametrization strategies, and each has been successfully used to investigate structural and dynamical properties of polysaccharides, although none performs optimally across all relevant cellulose properties.[21] Here, we use CHARMM36 merely as an example of a force field well established for cellulose systems in the literature. CHARMM36 was fitted with special emphasis on the description of the glycosidic bonds[17] and has a generally good performance for small carbohydrate structures and conformational properties, but with known limitations in quantitative conformational energetics.[22,23] Overall, it is a suitable choice among classical force fields for investigating carbohydrate structures and interactions at the atomistic level.

The overall weakness of force fields, be they ML-based or physics-based, is of course that they sacrifice the explicit presence of the electrons and therefore cannot provide explicit electron density information. Therefore, approximate interaction models that are both cost-effective and retain the electrons are of great interest, and constitute an active field of research. Here "cost-effective" means that they should be considerably less CPU/GPU-intensive than ab initio MD simulations, which evaluate forces and energies quantum-mechanically throughout the simulation propagation (see, e.g., refs 24,25 as an example from the cellulose field).

One such family of approximate computational methods that has shown promise is the semiempirical self-consistent-charge (SCC) density functional tight-binding (DFTB) method.[26,27] Like density functional theory (DFT), SCC-DFTB treats the electron density explicitly, but in an approximate manner, with some parts of the Hamiltonian being precomputed. This reduces the computational cost and allows for treatment of systems up to a few thousand atoms,[28] thereby bridging the gap between DFT and atomistic force field models. By explicitly including electrons, SCC-DFTB offers many advantages to a classical FF, e.g., by enabling the study of intra- and intermolecular polarization and charge transfer. Grimme et al. have provided an extension to the SCC-DFTB approach, namely the semiempirical extended tight-binding (xTB) method,[28,29] which, depending on the version (GFN1 or GFN2), incorporates the D3(BJ)[30] and revD4[31,32] dispersion corrections, respectively.

Despite its attractive features for large molecular computations, the use of SCC-DFTB for cellulose systems remains quite scarce in the literature.[33−35] Grimme et al. have validated the xTB methods for representative systems,[28,29,36] but to the best of our knowledge, no systematic benchmarking of the xTB methods has yet been carried out for cellulose. We undertake this here.

In this article, we present a systematic study of the structures, energies, and vibrational properties of cellobiose monomers and dimers. The present study is restricted to these isolated gas-phase systems. While they do not represent the structural complexity of cellulose, they provide useful representations of the key intermolecular interactions present in cellulose. In particular, the dimer configurations extracted from cellulose crystals allow us to examine molecular-level interactions associated with chain−chain and sheet−sheet cohesion without the additional contributions of the periodic bulk structure. These systems therefore provide a controlled basis for evaluating the ability of the approximative computational methods to describe intermolecular interactions also present in cellulose, but the smaller system sizes make it feasible to employ higher-level quantum-mechanical methods, thereby providing a robust reference for the present analysis. Our method/model assessment should be viewed as a first, and important, step toward identifying models that adequately describe cellulose building blocks across different environments. We evaluate the CHARMM36 classical force field, semiempirical quantum-mechanical methods, including GFNn-xTB (n = 1,2) and G-xTB, together with the foundation MLIP model MACE-OMol-0. The properties in focus are the molecular structures, intermolecular interactions, structural stability, and molecular vibrations. For all methods throughout the paper, geometries are optimized at 0 K, and vibrational properties are computed from the corresponding harmonic frequencies at these minimum-energy structures. No finite-temperature or dynamical sampling was performed, except in the CREST conformational search, which is described in more detail in Section 2.1. The results are benchmarked against DFT references (ωB97M-V, B3LYP-D3, and PBE-D3). ωB97M-V is our highest-level DFT method, B3LYP-D3 is commonly used for organic molecular systems, and PBE-D3 is commonly used for bulk periodic structures. All the methods used in this study are shown in Figure 1.

Downloaded from pubs.acs.org/jctce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

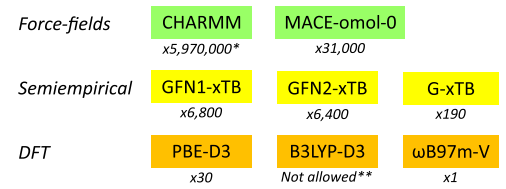


Figure 1. Overview of the methods benchmarked in the present cellulose study. Italicized values indicate the computational speed-up, relative to ORCA/ωB97M-V/def2-TZVPPd, based on the first 10 optimization steps. Timings are for the *chain-chain* dimer. All computations started from the same input geometry and were performed on a single compute node with 32 CPU cores. *For CHARMM, which was too fast for a reliable comparison over 10 steps, the speed-up was instead estimated from the number of optimization steps completed in one second. **Gaussian's license prohibits from publishing or sharing the performance data or benchmark times of Gaussian.

2. COMPUTATIONAL DETAILS

In this section, we first present details regarding the systems in focus, i.e., the cellobiose monomers and dimers (Subsection 2.1). Subsections 2.2−2.5 provide computational details regarding all the methods contained in Figure 1. Three categories of methods were used: (i) density functional theory (DFT), (ii) semiempirical density functional-based tight binding (SCC-DFTB), and (iii) atomistic force fields: physics-based (CHARMM36) and machine-learning (ML)-based (MACE). The systems were built either as periodic systems with the monomer or dimer residing in a large cell to avoid interactions between replicas, or in a nonperiodic setting, depending on which software was used. Calculations using the PBE-D3 and CHARMM36 methods were cast in periodic settings, and the rest as isolated monomers and dimers.

2.1. Monomer and Dimer Systems and Conformational Search

From a chemical point of view, cellulose is an organic polymer of D-glucose units linked by β-1,4-glycosidic ether bridges, forming cellobiose, the smallest repeating unit. Cellobiose in turn binds mainly through covalent bonds into long polysaccharide chains with hydrophilic (OH-rich) and hydrophobic (cellulose backbone) regions.[37] The hydrophilic OH-rich regions lead to additional stabilization through intra- and interchain hydrogen bonds. These regions also form flat sheets through the interchain hydrogen bonds. The sheets collect into crystals mainly through dispersion interactions and weak C−H···O bonds.[38] Native microfibrils consist of two crystal forms, $I\alpha$ and $I\beta$, described further in the literature.[2,38−43] Two different hydrogen-bonding patterns, A and B, have been proposed for each allomorph. Network A has been shown to dominate at normal conditions.[39,40] This paper deals with monomer and dimer systems only, although occasionally we relate the findings to structural aspects of the crystals.

Conformational variations in cellobiose can arise from several internal degrees of freedom. To describe the conformational differences, we use the dihedral angles of the glycosidic linkage Φ (H1−C1−O1−C4') and Ψ (C1−O1−C4'−H4') (see Figure 2). In vacuo, although some stable conformers have been identified,[44−46] cellobiose can exhibit a broad distribution of dihedral angles, making it difficult to point out one single lowest-energy structure.[44,47] Here, we instead investigate three different conformers to assess differences between methods.

A *trans* conformation (syn-Φ/syn-Ψ; Φ, 30°, Ψ, −20°) was constructed based on the atomic coordinates from the cellulose $I\beta$ crystal structure.[39] The terminal oxygen atoms were protonated to form the corresponding OH groups. A *cis* conformation (anti-Φ/syn-Ψ; Φ ≈ 180°, Ψ ≈ 0°) was derived starting from a DFT-optimized protonated cellobiose (cellobiose-H^+) structure,[48] by removing the protonating hydrogen. The *cis* and *trans* structures (Figure 2) are distinguished primarily by their glycosidic-linkage conformations, as characterized by the corresponding Φ and Ψ dihedral angles. For simplicity, we will use the terms *cis* and *trans* in this study for the *anti-syn* and *syn-syn* conformations, respectively, although this is not fully correct formally, as the *cis-trans* terminology refers to static, high-rotational-barrier conformations and the other terminology refers to flexible single-bond low-rotational-barrier conformations.

In addition to the *cis* and *trans* monomers, a conformational search was performed to identify the lowest-energy conformers at the GFN2-xTB level of theory with the Conformer−Rotamer Ensemble Sampling Tool (CREST)[49] (ver. 2.12). The GFN2-xTB level of theory was chosen due to the extensive computational cost for such simulations. The conformational search was performed with the iMDT-GC workflow to efficiently explore and generate low-energy conformers. The CREST simulations were performed in conjunction with the xTB package (ver. 6.6.1). The conformational search yielded a lowest-energy conformer, which will be labeled the min^0_{GFN2} conformer (Φ,Ψ) ≈ (160, 0°). The *cis*, *trans*, and min^0_{GFN2} conformers are the three primary monomer conformers investigated in this study. In addition, six further conformers (min^i_{GFN2}, i = 1, ..., 6) were found within $2k_BT_{300}$ of the min^0_{GFN2} global minimum monomer and were considered only for comparison of their relative energies.

The *cis*, *trans*, min^0_{GFN2}, and the six (min^i_{GFN2}, i = 1, ..., 6) monomers described above, optimized at the GFN2-xTB level, were subsequently used as starting geometries for optimization with the other methods. As stated in the introduction, all optimizations throughout this paper are performed at 0 K.

Two dimer conformers were extracted from the diffraction-determined $I\beta$(A) cellulose crystal structure,[39] which contains "infinite" parallel chains of connected cellobiose units, all in the *trans* configuration. The chains are H-bonded to each other in the $I\beta$(A) crystal, forming infinite sheets, which are arranged in a staggered fashion. Our two model dimers (Figure 3) were constructed as follows: Each dimer fragment was extracted directly from the crystal structure. The *chain−chain* dimer

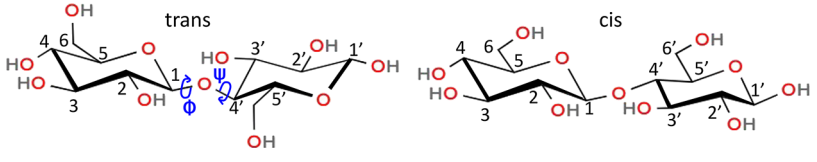


Figure 2. Structural representations of cellobiose in the *trans* (left) and *cis* (right) conformations.

Downloaded from pubs.acs.org/jctce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

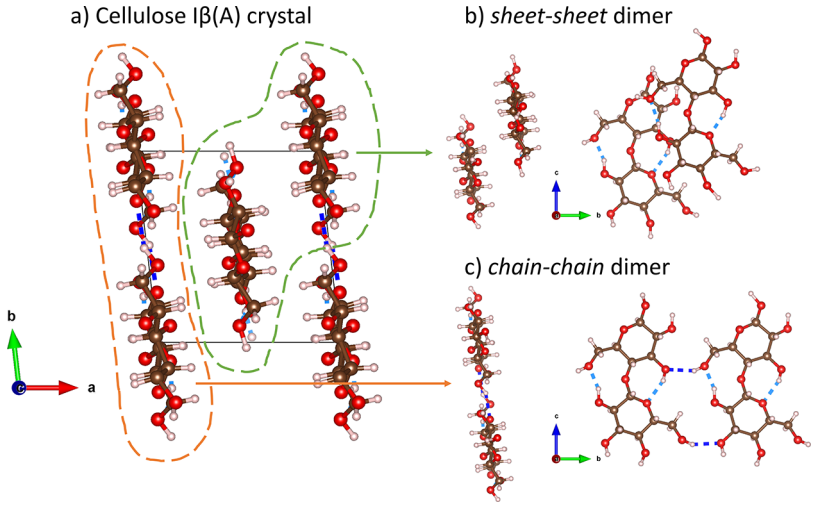


Figure 3. (a) The experimental Iβ(A) crystal structure and the two representative dimer motifs extracted from it: (b) the *sheet−sheet* dimer representing the intersheet stacking interactions, and (c) the *chain−chain* dimer representing the inter-chain hydrogen-bonding interactions.

contains a cellobiose unit from one chain and its H-bonded neighbor in the adjacent chain in the crystal. The *sheet−sheet* dimer instead contains a cellobiose unit from one sheet and its closest staggered neighbor in the adjacent sheet.

A CREST conformational search was performed also for the cellobiose dimer, with the same settings as for the monomer. A min^0_{GFN2} global minimum dimer was obtained. The *chain−chain*, *sheet−sheet*, and min^0_{GFN2} dimers are the three primary dimer structures investigated in this study. In addition, six min^i_{GFN2}, i = 1, ..., 6 dimer conformers were found within $2k_BT_{300K}$ of the min^0_{GFN2} global minimum dimer and were considered only for comparison of their relative energies.

The two *chain−chain* and *sheet−sheet* dimer fragments extracted from the experimental structures, as well as the min^0_{GFN2} and min^i_{GFN2}, i = 1, ..., 6 GFN2-xTB generated dimers, were subsequently used as starting geometries for 0 K optimization with all methods.

2.2. Density Functional Theory

DFT calculations were performed at three levels of theory (density functionals): the range-separated hybrid meta-GGA ωB97M-V functional, the hybrid B3LYP-D3 functional, and the PBE-D3 GGA functional. For some methods, we used one convergence criterion for energetic comparisons and a stricter convergence for structural and vibrational comparisons, called Level 1 and Level 2, respectively; see below.

ωB97M-V is a combinatorially optimized, range-separated meta-GGA hybrid functional with VV10 nonlocal correlation.[50] For these calculations, the def2-TZVPPD basis set was used as implemented in the ORCA 5.0.4 software.[51] The RIJCOSX approximation was used to accelerate the evaluation of the Coulomb and exchange integrals. Tight SCF convergence criteria were used. Geometry optimizations were performed using the Limited-memory Broyden−Fletcher−Goldfarb-Shanno (LBFGS) optimizer as implemented in the Atomic Simulation Environment (`ASE`).[52] A force criterion of 0.01 eV/Å (Level 1) and 0.002 eV/Å (Level 2) was used.

B3LYP-D3 is a commonly used hybrid density functional for molecular systems. For these calculations, the implementation in `Gaussian 16`[53] was used together with Dunning's double-ζ augmented aug-cc-pvdz[54] basis set and D3 dispersion correction.[30] Optimizations were performed using tight convergence criteria (maximum force: 1.5 × 10^{-5} Hartree/Bohr, RMS of forces: 10^{-5} Hartree/Bohr), an ultrafine integration grid, and otherwise default settings. The SCF convergence was set to very tight (10^{-8} Hartree), using the XQC convergence algorithm. For the dimer structures, symmetry constraints were disabled.

PBE-D3 is a GGA functional,[55] augmented with the D3 correction.[30] We used the Vienna Ab initio Simulation Package (`VASP`) 6.4.3[56−58] with periodic boundary conditions and plane-wave basis functions together with the projector augmented wave (PAW) pseudopotentials to describe the valence-core electron interactions.[59] The monomers and dimers were placed in simulation cells of 20 and 30 Å, respectively, with vacuum padding in all directions to ensure no periodic interactions.

A 500 eV plane-wave cutoff was used with Γ-point sampling. The electronic and ionic convergence criteria were set to 1 × 10^{-5} eV and 1 × 10^{-4} eV/Å (Level 1), and 1 × 10^{-7} eV and 1 × 10^{-8} eV/Å (Level 2). Optimizations of the monomers utilized the conjugate-gradient ionic optimizer, while the quasi-Newton optimizer was used for the dimers.

2.3. Semiempirical Density Functional Tight Binding

We herein investigate three semiempirical density functional tight-binding methods, namely GFN1-xTB, GFN2-xTB, and G-xTB. These are self-consistent tight-binding quantum-chemical methods.

Two GFN*n*-xTB variants were employed: GFN1-xTB and GFN2-xTB. These were designed for a cost-effective description of geometries, vibrational frequencies, and non-covalent interactions (GFN), while retaining the electronic information and accuracy. The GFN-xTB methods use minimal valence basis sets with minimal STO orbitals and element-specific parameters. Compared to GFN1-xTB,[29] GFN2-xTB[28] was updated with anisotropic multipole electrostatics and revD4 dispersion correction, and is reported to better describe polar and hydrogen-bonded systems, as well as reaction energies,[28] at only slightly increased computational cost. The GFN*n*-xTB calculations were performed using DFTB +[60] (ver. 23.1) with the xTB Hamiltonian.[28,29] The geometry optimizations utilized the LBFGS optimizer with a maximum force criterion of 1 × 10^{-7} eV/Å.

G-xTB is a novel, general-purpose addition to the xTB family, said to significantly improve the molecular description

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

Table 1. Overview of the Relation between Systems, Methods, Properties, and Figures Presented in This Work

	DFT			Semiempirical			Force fields	
systems/properties	ωB97M-V	B3LYP-D3	PBE-D3	G-xTB	GFN1-xTB	GFN2-xTB	CHARMM	MACE-OMol-0
Cellobiose monomers (*cis*, *trans*)								
Three optimized structures[a]	Figures 4/S1/S2	Figures 4/S1/S2	Figures 4/S1/S2	Figures 4/S1/S2	Figures 4/S1/S2	Figures 4/S1/S2	Figures 4/S1/S2	Figures 4/S1/S2
Stabilities for all investigated minima	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5
Harmonic frequencies for the three monomers	Figures 8/S7	Figures S8/S7	Figures S9/S7	Figures S10/S7	–	Figures S11/S7	–	Figures S12/S7
cellobiose dimers (*chain-chain*, *sheet-sheet*, and min_v^0)								
Three optimized dimer structures[b]	Figures 6/S4/S5	Figures 6/S4/S5	Figures 6/S4/S5	Figures 6/S4/S5	Figures 6/S4/S5	Figures 6/S4/S5	Figures 6/S4/S5	Figures 6/S5/S4
Stabilities for all the investigated minima	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5	Figure 5
Potential energy scans for the three dimers	Figure 7	Figure 7	Figure 7	Figure 7	Figure 7	Figure 7	Figure 7	Figure 7
Harmonic frequencies for two of the dimers	Figures 8/S7	Figures S8/S7	Figures S9/S7	Figures S10/S7	–	Figures S11/S7	–	Figures S12/S7

[a]Optimized monomers with each method, starting from the optimized *trans*, *cis*, and min^0_{GFN2} monomers, as obtained with the GFN2-xTB method. The min^0_{GFN2} monomer was obtained from the CREST-global minimum at the GFN2-xTB level. [b]Optimized dimers with each method, with *sheet–sheet* and *chain–chain* starting from the experimentally extracted dimers, and the min^0_{GFN2} dimer starting from the CREST-global minimum obtained with GFN2-xTB.

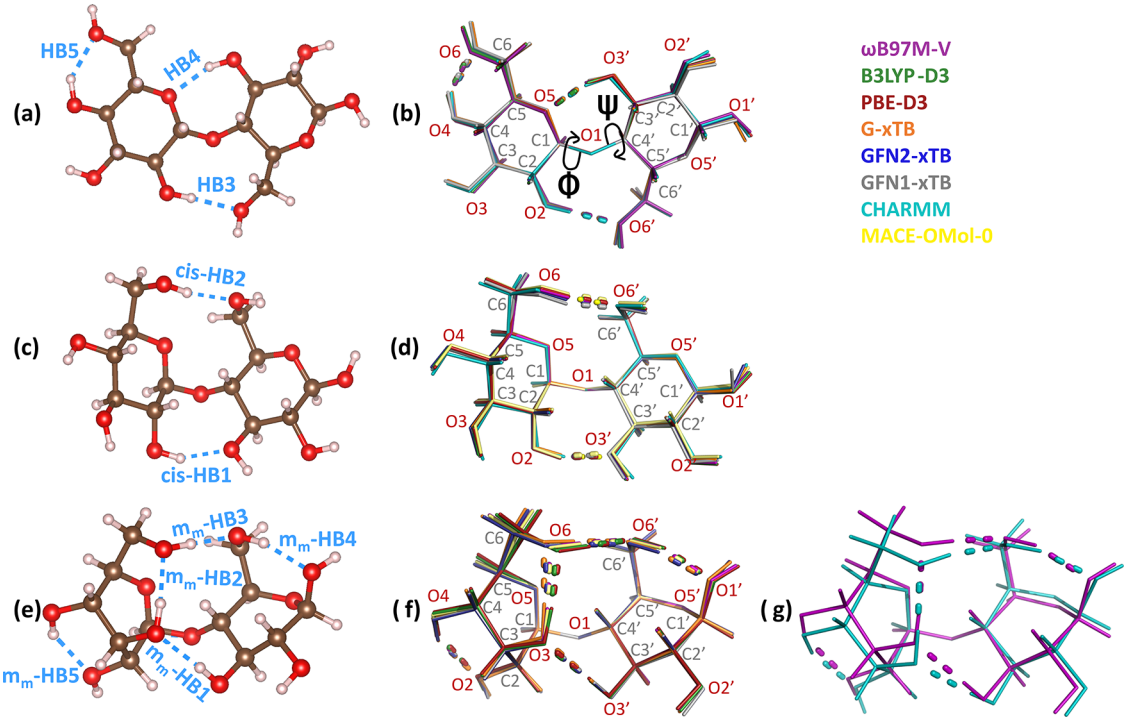


Figure 4. Comparison of the (a, b) *trans*, (c, d) *cis*, and (e–g) min^0_{GFN2} cellobiose conformers across levels of theory. (a, c, e) Ball-and-stick models of GFN2-xTB-optimized structures. (b, d, f, g) Superposition of structures optimized with the different methods, where (g) shows the comparison of the slightly deviating min^0_{GFN2}-CHARMM structure with the ωB97M-V reference. The optimized GFN2-xTB structure was used as the starting structure for optimizations in all methods. The top-right panel shows the color legend for the different methods.

compared to GFN*n*-xTB.[36] It has a more sophisticated charge-dependent q-vSZP basis set and incorporates range-separated exchange and higher-order charge effects, targeting ωB97M-V level accuracy, at the expense of increased computational cost compared to GFN2-xTB. The G-xTB calculations were carried out via the `g-xTB-ase`[61] interface. The geometry optimizations were performed using the `ASE` LBFGS optimizer with a force convergence criterion of 0.01 eV/Å.

2.4. Classical Force Field

We used Cellulose-Builder[62] to create the CHARMM36 topology file. CHARMM36-based structure minimizations and energy calculations were performed at 0 K using `LAMMPS`,[63] with a force convergence criterion of 1×10^{-8} eV/Å. Atomic positions were subsequently optimized within supercells of dimensions $22 \times 40 \times 31$ Å^3.

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223–12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

Table 2. R(H···O) Distances (Å) for O-H···O Hydrogen Bonds, Computed with Different Methods.[a] The atom notations for the monomers and dimers refer to those illustrated in Figure 4 and 6, respectively.

		ωB97M-V	B3LYP-D3	PBE-D3	G-xTB	GFN2-xTB	GFN1-xTB	CHARMM	MACE-OMol-0
trans monomer									
HB3	O_2–H···O_6'	1.936	1.935	1.904	1.904	1.935	1.971	1.956	1.937
HB4	O_3'-H···O_5	1.909	1.895	1.870	1.783	1.869	1.909	1.787	1.930
HB5	O_4–H···O_6	1.982	2.005	1.953	1.879	1.946	1.987	1.906	1.989
cis monomer									
cis-HB1	O_6–H···O_6'	1.937	1.933	1.882	1.852	1.951	2.035	1.853	1.949
cis-HB2	O_2–H···O_3'	1.966	1.954	1.933	1.947	2.002	2.015	1.976	1.974
min^0_{GFN2} monomer									
m_m-HB1	O_3'–H···O_3	1.916	1.870	1.816	1.880	1.981	2.006	2.032	1.904
m_m-HB2	O_3–H···O_6	1.828	1.783	1.708	1.750	1.841	1.870	1.872	1.847
m_m-HB3	O_6–H···O_6'	1.817	1.770	1.727	1.801	1.857	1.878	1.941	1.822
m_m-HB4	O_6'–H···O1'	1.874	1.847	1.799	1.776	1.837	1.867	1.908	1.883
m_m-HB5	O_4–H···O_2	1.946	1.950	1.890	1.839	1.925	1.957	1.922	1.955
chain–*chain* dimer									
HB3	O_2–H···O_6'	1.983	1.987	1.932	1.947	1.954	2.000	1.966	1.971
HB4	O_3'–H···O_5	1.796	1.779	1.769	1.706	1.802	1.837	1.760	1.811
HB5	O_4–H···O_6	1.942	1.967	1.928	1.836	1.904	1.952	1.916	1.982
HB1	**O_6'–H···O_3'**	**1.874**	**1.845**	**1.843**	**1.797**	**1.842**	**1.880**	**1.920**	**1.885**
HB2	**O_6–H···O_3**	**1.821**	**1.792**	**1.746**	**1.779**	**1.809**	**1.872**	**2.015**	**1.814**
HB3′	O_2–H···O_6'	1.797	1.770	1.731	1.749	1.775	1.827	1.892	1.829
HB4'	O_3'–H···O_5	1.946	1.952	1.922	1.829	1.897	1.944	1.809	1.960
HB5′	O_4–H···O_6	1.969	1.988	1.946	1.865	1.928	1.970	1.912	1.984
sheet–*sheet* dimer									
HB3	O_2–H···O_6'	1.824	1.810	1.780	1.803	1.852	1.888	1.890	1.806
HB4	O_3'–H···O_5	1.887	1.879	1.853	1.763	1.844	1.902	1.799	1.913
HB5	O_4–H···O_6	1.988	2.008	1.958	1.876	1.942	1.989	1.909	1.993
HB6	**O_6'–H···O_5'**	**1.869**	**1.843**	**1.829**	**1.750**	**1.880**	**1.916**	**2.138***	**1.886**
HB3′	O_2–H···O_6'	1.975	1.987	1.961	1.946	1.965	1.991	1.940	1.988
HB4'	O_3'–H···O_5	1.866	1.837	1.820	1.720	1.821	1.848	1.763	1.866
HB5′	O_4–H···O_6	1.994	2.016	1.959	1.876	1.944	1.990	1.911	2.004
min^0_{GFN2} dimer									
m_d-HB1	O_2–H···O_6'	1.755	1.742	1.724	1.754	1.819	1.849	1.821	1.751
m_d-HB2	**O_6–H···O1**	**1.819**	**1.785**	**1.757**	**1.705**	**1.785**	**1.815**	**1.752**	**1.826**
m_d-HB3	**O_6'–H···O_6**	**1.744**	**1.722**	**1.702**	**1.734**	**1.779**	**1.829**	**1.850**	**1.749**
m_d-HB4	**O_6'–H···O_2'**	**1.879**	**1.834**	**1.806**	**1.831**	**1.870**	**1.862**	**1.902**	**1.878**
m_d-HB5	**O_2'–H···O_5'**	**1.968**	**1.919**	**1.860**	**1.855**	**1.975**	**1.994**	**1.986**	**1.974**
m_d-HB6	**O_3'–H···O_6**	**1.788**	**1.751**	**1.700**	**1.771**	**1.803**	**1.829**	**1.850**	**1.794**
m_d-HB7	O_6–H···O_6'	1.791	1.771	1.762	1.788	1.854	1.875	1.888	1.797
m_d-HB8	O_2–H···O_3'	1.969	1.958	1.872	1.869	1.940	1.966	1.928	1.972

[a]Intermolecular distances are shown in bold. *Interaction not classified as a hydrogen bond by the applied criteria ($d < 2.1$ Å and $\angle O{-}H{\cdots}O > 130°$). See also Figures S4 and S5.

2.5. Machine-Learning Interatomic Potentials

As mentioned in the Introduction, MLIPs are machine-learned interatomic potentials that approximate quantum-mechanical potential energy surfaces (PESs) with far lower computational cost while retaining high accuracy. They serve as a bridge between expensive ab initio methods and fast but less advanced classical force fields. In this work, we use an MLIP built with the universal Message Passing Atomic Cluster Expansion (MACE) architecture. More specifically, we use the `MACE-OMol-0-extra-large-1024` foundation model,[11] which was trained on ωB97M-V data[64] and aimed to address the need for high-accuracy molecular structures. Geometry optimizations were performed using the LBFGS optimizer as implemented in `ASE`, with a force convergence criterion of 0.001 eV/Å.

2.6. Vibrational Properties

Harmonic vibrational calculations were performed for a selection of the methods employed here. For GFN2-xTB and G-xTB, we utilized the finite-difference approach implemented in the `ASE` Infrared module. Atomic displacements of 0.01 Å were applied, and the numerical derivatives were evaluated using two-sided finite differences.

With VASP, we utilized the phonopy[65] interface and phonopy-spectroscopy package[66] for vibrational analysis. Before the vibrational analysis, the systems were fully optimized to level 2 precision. The finite displacement was set to 0.01 Å. The single-point calculations were performed with a fast electronic minimization and an additional support grid to improve the evaluation of augmentation charges. Born-effective charges and the dielectric tensor were calculated to obtain IR intensities.

Downloaded from pubs.acs.org/jctce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

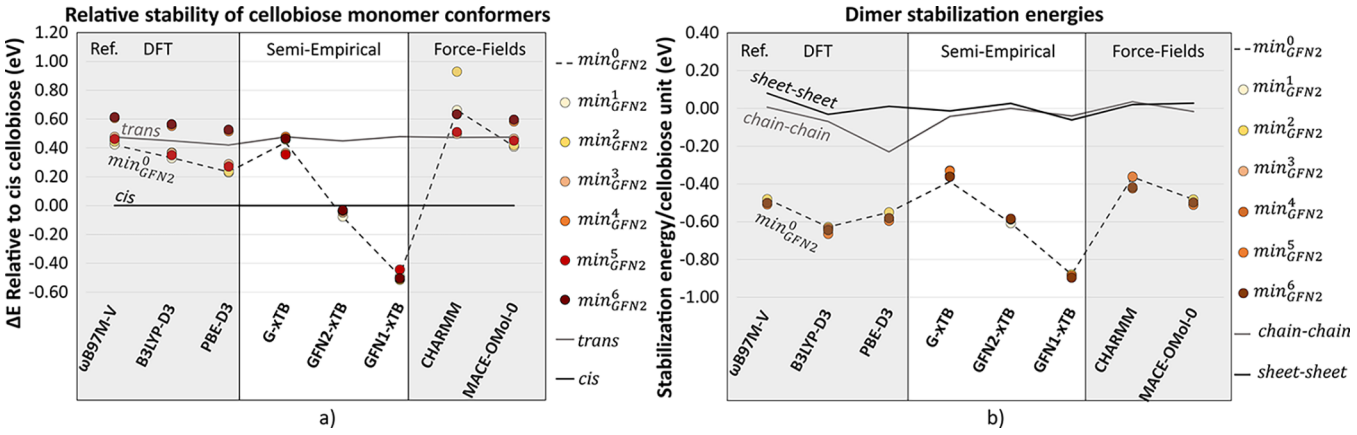


Figure 5. Results after geometry optimization with each method for the structures described in Section 2.1: (a) Relative monomer conformer energies referenced to the *cis* conformer and (b) dimer stabilization energies per cellobiose unit, defined as $E_{stabilization} = E_{dimer}/2 - E_{cis}$, where both the monomer and dimers are geometry optimized with the respective method. The colored circles denote the structures found within 2KbT from the min0GFN2 minimum with each method.

The B3LYP-D3 frequency calculations were performed with Gaussian's standard analytical harmonic vibrational analysis procedure.

The ωB97M-V vibrations were calculated using the ORCA numerical vibrational analysis (NumFreq) and default settings. Before the vibrational analysis, the systems were fully optimized to level 2 precision.

3. RESULTS AND DISCUSSION

The results are divided into three main parts: Section 3.1, Monomer structures and stabilities; Section 3.2, Cellobiose dimers; and Section 3.3, Vibrational signatures. The relations between systems, methods, properties, and figures in this work are summarized in Table 1.

Throughout the article, we define an O-H···O hydrogen bond as existing if R(H···O) is ≤ 2.1 Å and ∠ O-H···O is ≥ 130°.

3.1. Monomer Structures and Stabilities

The optimized *trans* and *cis* monomer structures, with three and two hydrogen bonds, respectively, closely resemble the corresponding starting structures and are qualitatively similar across the different methods (Figure 4a–d), indicating convergence to genuine local minima in each case. R(H···O) distances for O-H···O and C–H···O hydrogen bonds are shown in Table 2, and distance correlation curves are shown in Figures S1 and S2. MACE-OMol-0 outperforms all methods in comparison with the ωB97M-V results, not unexpectedly since it was trained on ωB97M-V data. The C–O covalent bonds are systematically underestimated in all xTB methods. The noncovalent distances vary the most between methods. Even among the DFT functionals, the H···O distances differ, with PBE-D3 showing larger deviations from the ωB97M-V reference than B3LYP-D3, of up to 0.12 Å.

Despite discussions and some inconsistencies in the computational literature, the *cis* conformer is now generally considered the more stable conformer in vacuo,[44,47,67] whereas *trans* is the most stable in the cellulose crystals and in liquid.[45,68] It has previously been acknowledged that the favoring toward the *trans* configuration in vacuo is associated with too small basis sets.[44] Reassuringly, all our results show that the *cis* conformer is lower in energy than *trans*. We find that the relative energies between the *trans* and *cis* types are similar (all are 0.4–0.5 eV) between methods (Figure 5a). The higher stability of the *cis* conformer (with its two hydrogen bonds) over the *trans* conformer (with three hydrogen bonds) suggests that factors beyond the number of hydrogen bonds influence the stability in cellulose, as previously stated by Wohlert et al.[8]

The lowest-energy conformation identified by the CREST conformational search at the GFN2-xTB level is a folded structure with *cis*-like glycosidic angles (Φ,Ψ) ≈ (0, −20°), puckered rings, and five hydrogen bonds (Figure 4e). As mentioned, we label this conformer min^0_{GFN2}. We also generated conformers within $2k_BT_{300K}$ from the global minimum with the CREST method, which will be discussed at the end of this section.

The GFN*n*-xTB methods overestimate the stability of the puckered CREST conformer (Figure 5a) compared to the ωB97M-V results. This overstabilization has been observed for other small ring structures.[36] We note that no such enhanced stability is obtained with G-xTB, which might be attributed to several key improvements implemented in G-xTB compared to the GFN*n*-xTB methods.

In addition to the global minimum min^0_{GFN2} already discussed, we examine structures that should be reachable within a room-temperature MD simulation. Therefore, we end this section with a few words about these higher-energy structures found in our conformer search using the CREST program at the GFN2-xTB level. In Figure 5a, we have added structures found within $2k_BT_{300\ K}$ (0.052 eV) from the minimum (note that the dot marks for these conformers partially overlap in several places in the plot, but there are indeed six marks for each method). This corresponds to six unique conformations; they are shown in Figure S6. These six structures were subsequently used as inputs for geometry optimizations with the other considered methods. We find that these conformers in most cases relax to essentially a nearby local minimum structure in the different methods, but with much larger energy spread than the $2k_BT_{300\ K}$. These results further highlight the limitations of the GFN*n*-xTB and CHARMM methods for cellobiose, not only for the 0 K structures but also those that can appear in dynamic simulations. G-xTB seems to partially mitigate the problems, and the pattern of the higher-energy structures agrees reasonably quite well with the ωB97M-V pattern in Figure

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

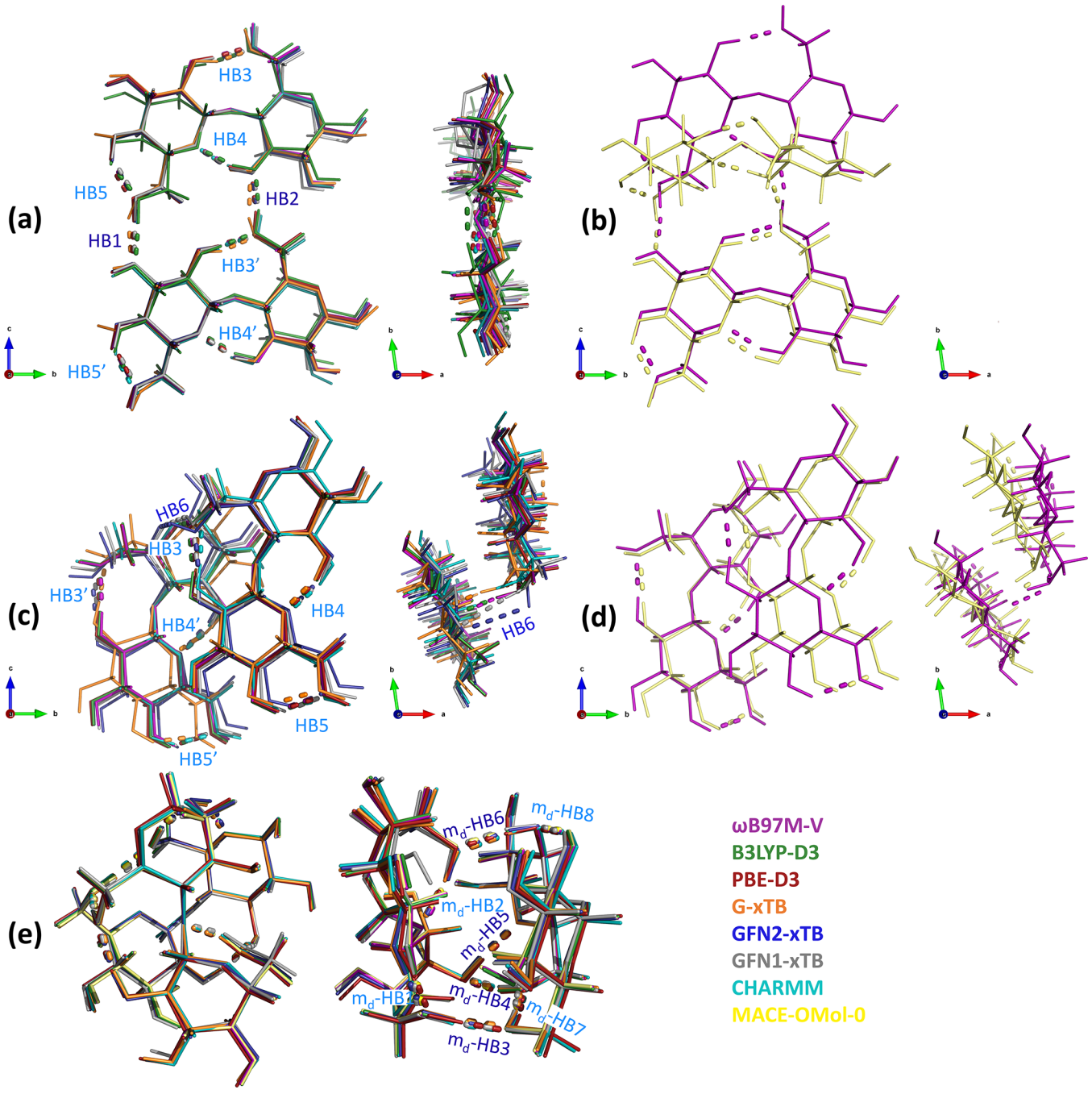


Figure 6. Comparison of the (a, b) *chain*−*chain*, (c, d) *sheet*−*sheet*, and (e) min^0_{GFN2} dimer structures across levels of theory. (a, c, e) Superposition of structures optimized with the different methods, excluding MACE-OMol-0 structures that deviate significantly from the reference (shown separately in panels b, d). (b, d) Comparison of the MACE-OMol-0 structure with the ωB97M-V reference for conformers where MACE-OMol-0 deviates. For (a−d), all optimizations were started from the experimentally extracted dimer structures. For (e), the GFN2-xTB structure obtained from CREST (min^0_{GFN2}) was used as the starting structure. The bottom-right panel shows the color legend for the different methods. The monomer conformations comprising the min^0_{GFN2} dimer structure are shown in S3.

5a. However, as seen in Figure S6, the energies are systematically shifted compared to the DFT reference. MACE-Omol-0 seems to be stable and provides reasonable energy and structural results comparable to ωB97M-V also for the broad set of cellobiose monomer conformers (see additional comparative plot in Figure S6).

3.2. Cellobiose Dimers

3.2.1. The Chain−Chain and Sheet−Sheet Dimers. Starting from the experimental-crystal-extracted *chain*−*chain* and *sheet*−*sheet* dimer structures, optimizations were performed with all methods used in this study. The resulting structures are compared in Figure 6a−d.

Both the *chain*−*chain* and *sheet*−*sheet* dimers adopt overall quite similar structures for all the methods except MACE-OMol-0, for which the *chain*−*chain* dimer is considerably different, and the *sheet*−*sheet* dimer is slightly different. In the *chain*−*chain* MACE structure, the two cellobiose units have "unfolded" from each other, forming a bent structure with a plane−plane angle not far from perpendicular. This is slightly

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

Table 3. Calculated Reorganization Energies for the Individual Monomer Fragments in the Dimers in Figure 6[a]

structure	ωB97M-V	B3LYP-D3	PBE-D3	G-xTB	GFN2-xTB	GFN1-xTB	MACE-Omol-0
min^0_{GFN2}: monomer 1 (*trans*)	+0.42	+0.43	+0.31	+0.20	−0.23	−0.48	+0.39
min^0_{GFN2}: monomer 2 (*cis*)	+0.47	+0.47	+0.37	+0.43	+0.32	+0.21	+0.45
chain−chain: monomer 1 (*trans*)	+0.03	+0.04	+0.06	+0.01	+0.05	−0.07	+0.05
chain−chain: monomer 2 (*trans*)	+0.05	+0.04	+0.06	+0.01	+0.03	−0.07	+0.03
sheet−sheet: monomer1 (*trans*)	+0.04	+0.03	+0.04	−0.02	−0.02	−0.11	+0.05
sheet−sheet: monomer 2 (*trans*)	+0.02	+0.04	+0.03	−0.05	+0.03	−0.05	−0.01

[a]For each method, the reorganization energy of a dimer is the sum of the reorganization energies of its two monomers. Each isolated monomer's single-point energy at its dimer geometry is compared to the optimized isolated monomer of the same conformer type (*cis* or *trans*); see the text for more details. Rows 1-2 refer to the dimer structures in Figure 6e, rows 3-4 refer to Figure 6a,b, and rows 5-6 refer to Figure 6c,d.

surprising given that MACE performs very well for all other systems and properties in this study and actually, perhaps also slightly surprisingly, all intra- and H-bond distances within the bent MACE structure conform very well with those of the unbent reference ωB97M-V structure (cf. the distance correlation plots in Figures S4 and S5). We have no straightforward explanation for the odd MACE structure bending, and have not delved deeper into this issue.

The *chain−chain* and *sheet−sheet* dimer stabilization energies (defined as $E_{stabilization} = E_{dimer}/2 - E_{cis}$, where both the monomer and dimers are geometry optimized with the respective method, as opposed to the interaction energy defined later) are of comparable magnitude (Figure 5b, given per cellobiose unit). This suggests that the dispersion-dominated interactions in the *sheet−sheet* dimer and the hydrogen bond-dominated interactions in the *chain−chain* dimer make comparable contributions to dimer stability. The *sheet−sheet* dimer forms an intermolecular O-H···O hydrogen bond (HB6) during the optimization. This is the only H-bond that, together with the dispersion interactions, holds the sheets together in this dimer.

One notes that many of the stabilization energies in Figure 5b are positive (not stabilizing). This peculiarity stems from the fact that the stabilization energies were calculated with respect to the lowest-energy monomer, i.e., the *cis* monomer, while the cellobiose moieties in the *chain−chain* and *sheet−sheet* dimers are all of the *trans* configuration, as mentioned above, as they were extracted from the experimental crystals. One can easily transform the stabilization energies in Figure 5b by adding the "$E_{trans} - E_{cis}$" value from Figure 5a for each computational method. The result will, in each case, be a negative value, confirming that using the isolated *trans* monomers as the reference when forming the optimized dimers is stabilizing, as must be. However, regardless of whether the *trans* or *cis* monomers are used as a reference, the figure displays the variation as a function of method, which is its main purpose. We find that all methods give stabilization energies close to the ωB97M-V method for the *chain-chain* and *sheet-sheet* dimers.

3.2.2. Global Minimum Structure and Conformational Search. Like we did for the monomer, we identify a min^0_{GFN2} global minimum dimer (Figure 5b) through a CREST conformational search at the GFN2-xTB level of theory. Similarly to our procedure for the monomers, this structure served as input for geometry optimizations with all methods. Inspection of the optimized structures (Figure 6e) shows that the optimized structures are all quite similar to the initial min^0_{GFN2} structure. MACE-OMol-0 gives the best agreement with the ωB97M-V bond distances (see the distance correlation plot in Figure S5c).

The min^0_{GFN2} dimer is significantly different from the *chain−chain* and *sheet−sheet* dimers and can best be described as consisting of strained monomers, one *cis* and one *trans*, in a compact structure forming an extensive hydrogen-bond network (the monomers are shown in Figure S3). The *cis-trans* combination enables four intermolecular hydrogen bonds, in addition to four intramolecular hydrogen bonds.

Starting from the min^0_{GFN2} dimer, the stabilities of the structures after optimization with the respective methods are shown in Figure 5b. In contrast to the monomer, all methods now consistently predict the puckered min^0_{GFN2} dimer structure to be energetically more favorable than the *chain−chain* and *sheet−sheet* dimers. One might naively attribute this stabilization to the presence of a cis conformer for one of the min^0_{GFN2} dimer units. However, it is striking that the GFN2-xTB versus ωB97M-V energy difference that was observed for the monomer (Figure 5a) is not reflected in the dimer energetics: while the GFN*n*-xTB methods substantially overstabilize the min^0_{GFN2} monomer conformer, the calculated stabilization energy of the corresponding min^0_{GFN2} dimer differs by only ∼0.1 eV from the ωB97M-V results.

To investigate the origin of this possible inconsistency, the *reorganization energy* for each monomer upon formation of the min^0_{GFN2} dimer was calculated for each method (Table 3). For each method, we define the reorganization energy for the min^0_{GFN2} dimer as

E(monomer 1 fixed at the min^0_{GFN2} dimer geometry)
− E(the optimized *trans* monomer)
+ E(monomer 2 fixed at the min^0_{GFN2} dimer geometry)
− E(the optimized *cis* monomer).

Monomer 1 is the unit to the left in Figure 6e, which is a distorted *trans* conformer, and monomer 2 is the unit to the right, which is a distorted *cis* conformer (see conformers in Figure S3). First, Table 3 shows that the reorganization energies of the min^0_{GFN2} dimers are substantially larger than those of the *chain−chain* and *sheet−sheet* dimers. Second, whereas all the DFT methods predict positive reorganization energies, as expected, the GFN*n*-xTB methods yield negative reorganization energies for the *trans* monomer associated with the min^0_{GFN2} dimer (min^0_{GFN2} monomer 1 (trans) in Table 3). This means that although the different methods predict similar stabilization energies for the min^0_{GFN2} dimer, they arrive at this result for fundamentally different physical reasons, as seen in Table 3. With the three DFT methods, the large reorganization penalty is compensated for by the substantial energy gain upon dimerization, as expected. In contrast, the stability of the GFN*n*-xTB methods originates partly from the energy release from the reorganization step itself, which for monomer 1 is energetically favorable rather than costly. We attribute this to the flatter PES of these methods, which gives rise to a richer set

Downloaded from pubs.acs.org/jctce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

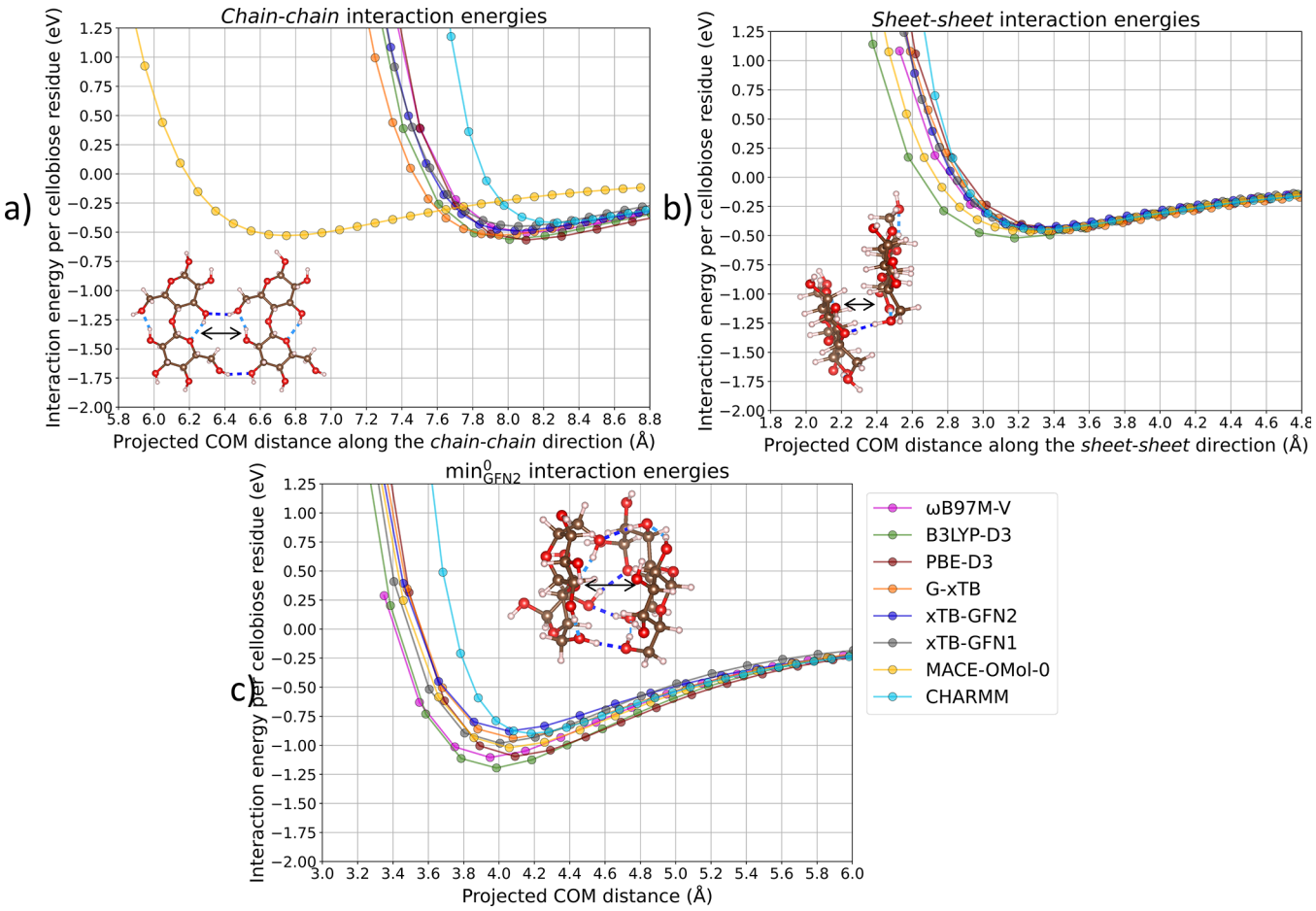


Figure 7. Single-point interaction energy scans for the (a) *chain*−*chain*, (b) *sheet*−*sheet*, and (c) min^0_{GFN2} dimers, generated by rigidly displacing one monomer perpendicular or parallel to the sheet plane (see insets*). The translation distance is the distance between the centers of mass (COMs) of the two cellobiose molecules projected onto the respective translation direction. The scans start from the dimer optimized with each of the six methods, with interaction energy referenced to the separated isolated monomer fragments with their geometries fixed. *Insets show the GFN2-xTB optimized dimer structures.

of metastable local minima. As a monomer distorts into its dimer geometry, it can therefore relax toward one of these nearby shallow minima instead of paying a strain penalty, which in itself is stabilizing.

Similarly to our procedure for the monomers, we also performed additional optimizations of the dimer conformers from the CREST conformational search at the GFN2-xTB level within $2k_BT_{300\ K}$ from the minima. Six additional dimer structures (min^i_{GFN2}, i = 1, ..., 6) were obtained (see configurations in Figure S6 and relative energies in Figure 5b). While the energy spans of the high-energy minima are quite similar, the xTB methods display different energy patterns compared to the DFT methods and MACE-OMol-0, showing smaller spread between the different min^i_{GFN2} local minima.

3.2.3. Potential Energy Scans for the Dimers. In this subsection, we explore aspects of the interchain and intersheet interactions using potential energy scans (defined as $E_{int} = (E_{dimer} - E_{monomer1} - E_{monomer2})/2$, where the dimer is geometry optimized, while monomers 1 and 2 are the respective separated, isolated monomer fragments from the dimer, with their geometries fixed). Starting from the *chain*−*chain*, *sheet*−*sheet* and min^0_{GFN2} dimers optimized with the respective method, i.e., the structures displayed in Figure 6a−e, the two cellobiose molecular fragments were translated rigidly along the double arrow shown in each of Figure 7a−c, covering intermolecular distances from small to large. A single-point energy calculation was performed at each step (with a 0.1 or 0.2 Å interval between the projected mass center of the molecules) and the interaction energy was calculated with reference to the separated isolated cellobiose fragments of the dimer. The resulting interaction energy curves (Potential energy curves) are shown in Figure 7.

The equilibrium distances agree well between methods, although the CHARMM results stick out, and of course the MACE results for the *chain*−*chain* dimer. The distorted MACE-OMol-0 *chain*−*chain* structure displays significant shifts in the distance between mass centers due to the distorted geometry, although interestingly, the energies are similar to the other methods, which we also saw in Figure 5b. We note again that the *chain*−*chain* and *sheet*−*sheet* interaction energies are of comparable magnitude.

Figures 5b and 7 both concern the interaction between two cellobiose molecules. Are the results in the two figures consistent? The answer is yes; it is only the reference cellobiose conformer that differs. While in Figure 5b the stabilization energies are given with respect to the fully optimized isolated **cis** monomers for each method, in Figure 7 the interaction energy is calculated against two separated reference monomers that have kept the geometries that they

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

had in the optimized dimer. Using the reorganization energies in Table 3, as well as the **cis**-to-**trans** energy difference readable from Figure 5a, will yield complete consistency between Figure 5b and the interaction energy at the minimum in Figure 7.

3.3. Vibrational Signatures

The scientific literature contains ample examples of experimental Infrared, Raman, and SFG vibrational studies of cellulose, from early publications such as that of Tsuboi in 1957[69] to today's vibrational characterization applications of advanced materials of, say, composites of nanocrystalline cellulose with redox-active inorganic nanoparticles.[70] However, for gas-phase cellobiose monomers and small aggregates, the experimental vibrational studies are much scarcer. In fact, we are not aware of any published vibrational spectroscopy studies of pure gas-phase cellobiose monomers or dimers. Gas-phase phenyl-cellobiose has been examined experimentally, and the authors claimed that the chromophore does not substantially affect the vibrational spectrum of cellobiose67,71. However, this is difficult to judge. The scarcity of experimental data makes modeling particularly valuable, and in fact also for systems where experimental data do exist, the assignment and interpretation of experimental spectra nowadays usually rely heavily on complementary information from calculations.

With the experimental region of phenyl-β-cellobioside[71] being broader and individual peaks less distinct, direct comparison between the experimental and our computational peaks is difficult. We have therefore focused our analysis on the computational methods. Nevertheless, it can be noted that the OH stretching region with all methods we used lies above that of the experimental spectra of phenyl-β-cellobioside (see S7 and Figure 8). This is expected since the calculated vibrations are harmonic, while the experimental are anharmonic, especially the O-H bonds. These shifts are often corrected through scaling approaches for a better comparison to experiment, see for example ref 71. We have not applied any scaling factors to any of the calculated frequencies since our main analysis is between computational methods.

Vibrational patterns are fingerprints of the force constants pertaining to the (normal) modes of vibration. Here we present harmonic infrared spectra for our cellobiose monomers and dimers for some selected methods: the OH stretching region in Figure 8 (for ωB97M-V) and in Figures S8−S12 (for five of the other methods). We did not perform vibrational analysis for GFN1-xTB, as GFN2-xTB was considered the more appropriate xTB method based on the structural and energetic analyses presented above. For CHARMM36, vibrational analysis was likewise not performed. Instead, the stability of the optimized structures was assessed through MD simulations at 1, 15, and 298 K by monitoring their hydrogen-bond distributions (see S13).

The whole frequency range from 0 to 4000 cm^{-1} is displayed in the correlation plots in Figure S7. The various frequency regions seen in the correlation plots cover the following classes of vibrational modes, going from high to low frequencies: O−H stretching frequencies, C−H vibrations, and then for the lower frequency region, say below 1500 cm^{-1}, the spectral peaks have been assigned to bending, torsion, and shear in the literature, for example from CPMD simulations at the BLYP-D level by Pincu et al.[71] on a cellobiose derivative.

OH vibrations are sensitive probes of H-bonding, and our presentation will focus on this region. The vibrational modes are numbered 1 to 8 for the monomers in the three upper panels in Figure 8, and the main OH groups involved are numbered accordingly in the small structural figures inside the frequency strips.

3.3.1. Monomers. Starting with the *trans* and *cis* monomers (Figure 8), the vibrations numbered 1−5 lie high in frequency and correspond to the dangling (i.e., not H-bonded) OH groups. Modes No. 7−8 in *cis* and 6−8 in *trans* belong to OH groups that donate H-bonds, and for both the *trans* and *cis* monomers these modes are substantially downshifted (red-shifted) in frequency compared to the dangling OH groups. Furthermore, the IR intensities of the H-bonded OH groups are seen to be much higher. Both those effects are typical signatures of H-bonding. The three intramolecular H-bonds in the *trans* monomer correspond to modes No. 8 (HB3, 1.94 Å), No. 7 (HB4, 1.92 Å), and No. 6 (HB5, 1.98 Å); all three can be classified as typical medium-long R(H···O) hydrogen-bond distances. The OH group involved in mode No. 6 is H-bonded for the *trans* monomer but not so for the *cis* case, which is also reflected in the spectra. Modes 7 and 8 are strongly coupled into a symmetric and an antisymmetric mode; they are not symmetry-related but they are coupled by virtue of their similarity in strengths and their relative proximity.

The structure of the min^0_{GFN2} monomer is quite different from the *cis* and *trans* conformers, as we learned in Section 2.1. It has more and generally stronger (shorter) H-bonds than the *cis* and *trans* monomers, namely five intramolecular H-bonds in the R(H···O) range of 1.82−1.95 Å (ωB97M-V values, Table 2). They involve OH groups corresponding to modes No. 4−8 in the small figure and the frequencies are labeled accordingly in the spectrum. The lowest-frequency modes 7 (m_m-HB3, 1.817 Å) and 8 (m_m-HB2, 1.828 Å) correspond to the shortest hydrogen bonds. Mode 4 (m_m-HB5, 1.946 Å) refers to the longest bond in most methods, has a smaller angle (≈140°), and is weaker than the other H-bonds, explaining its higher wavenumber.

3.3.2. Dimers. Moving now to the dimers in Figure 8, we make several interesting observations. Starting with the *chain−chain dimer*, the strongest H-bonds appear to be the intramolecular H-bonds, but intermolecular H-bonds (labelled 1′ and 2) are also appreciable in strength and lie somewhere in the midst of the collection of spectral lines from H-bonded OH groups. A comparison between OH group No. 7' (in the molecular moiety shown to the left in the fourth panel of Figure 8) and No. 7 (in the right-hand moiety) reveals that their frequencies are hugely different, while the only apparent structural difference is that the OH group No. 7' both donates an H-bond and accepts one, while No. 7 is only a donor. This is crucial, as chains of H-bonds are known to strongly enhance the H-bonding of the "middle-most" molecule(s) through cooperative polarization effects, and we see such an example here.

In the *sheet−sheet dimer*, there is only one intermolecular H-bond (labelled 2′), with R(H···O) equal to 1.869 Å (ωB97M-V). It is best displayed in Figure 6c. The intermolecular interaction in the *sheet−sheet dimer* should be dispersion-dominated except for that hydrogen bond, and the two *trans* cellobiose moieties would be expected to maintain much of their original identities. Inspecting modes No. 1−6 and 1'−6', this seems to be the case. In mode 2', the O−H stretching along the intermolecular bond downshifts relative to the *trans* monomer, as expected. The downshift of modes 7 and 8' and

Downloaded from pubs.acs.org/jctce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

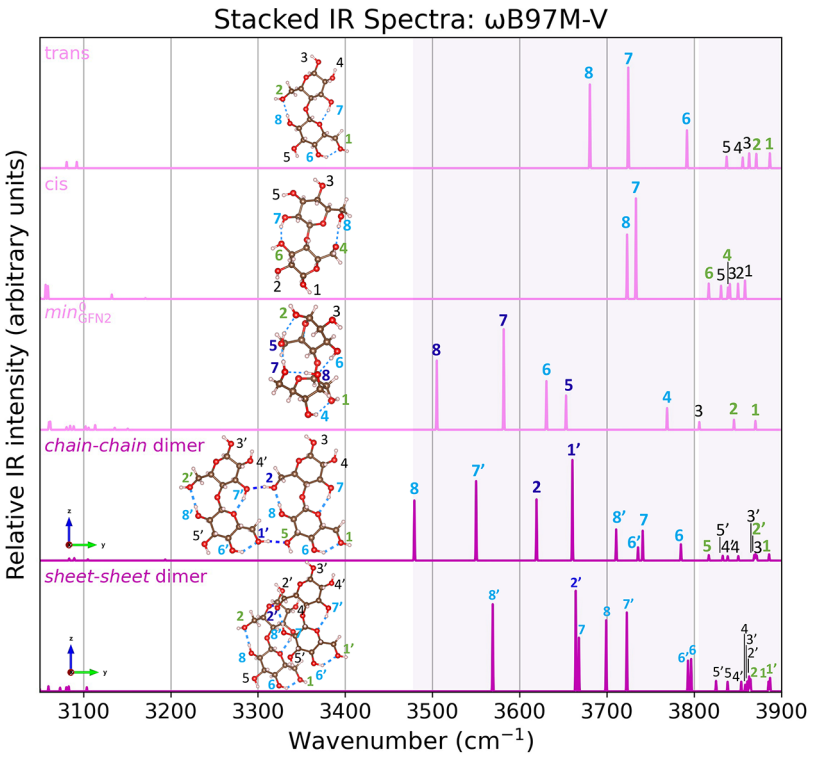


Figure 8. Computed IR spectra (ωB97M-V) of the trans, cis, and min^0_{GFN2} monomers and the *chain*−*chain* and *sheet*−*sheet* dimers.

Table 4. RMSD (cm^{-1}) of Harmonic Vibrational Frequencies for the *trans*, *cis*, and min^0_{GFN2} Monomers and the *Chain*−*Chain* and *Sheet*−*Sheet* Dimers, Relative to ωB97M-V (See Also Figure S7)[a]

	B3LYP-D3	PBE-D3	G-xTB	GFN2-xTB	MACE-OMol-0
trans monomer	20	70	58	106	9
cis monomer	18	68	54	103	9
min^0_{GFN2} monomer	26	86	54	105	10
chain−*chain* dimer	26	79	65	111	18
sheet−*sheet* dimer	20	69	72	103	12

[a]The frequencies (0−4000 cm^{-1}) were compared mode-by-mode in the order of vibrational assignment, assuming a one-to-one correspondence between modes without accounting for possible mode reordering.

the upshift of mode 8 follow the corresponding bond length changes upon dimerization (Table 2). HB3′ (Mode 8′) shortens from 1.936 to 1.824 Å, and HB4′ (Mode 7) shortens from 1.909 to 1.866 Å, consistent with strengthened, downshifted stretches. HB3 (Mode 8) lengthens from 1.936 to 1.975 Å, consistent with a weakened, up-shifted stretch. Mode 8 is largely unaffected compared to the *trans* monomer, because it lies farther from the intermolecular hydrogen bond and the neighboring molecule.

Next we compare the results from the different methods. We expect vibrational spectra to differ for two reasons. One, the structures relax to slightly different equilibrium geometries, as shown in Figures 4 and 6. Two, the shapes of the PESs, here especially their curvatures at equilibrium, differ.

The spectra for the OH stretching region from B3LYP-D3, PBE-D3, G-xTB, GFN2-xTB, and MACE-OMol-0 calculations are shown in Figures S8−S12 and can be compared with the ωB97M-V spectra (Figure 8), discussed in detail above. Starting with MACE-OMol-0, we note that this method gives the best agreement with ωB97M-V. This is also true over the whole 0−4000 cm^{-1} range as reported in the correlation analysis in Table 4, for which the MACE-OMol-0 vs ωB97M-V RMSD value is small, at most just 18 cm^{-1}. Again, the good agreement is comforting, although perhaps not surprising, as this MACE model was trained on ωB97M-V data, although not with respect to frequencies. Despite this, we note that especially for the *sheet*−*sheet* dimer, the frequency agreement in the O−H stretching region (Figures 8 and S12) between the two methods is only modest.

Compared to ωB97M-V, the frequencies yielded by all the other electronic structure methods (Table 4 and Figure S7) lie (a bit or more) too low, especially the C−H and O−H stretches. In PBE-D3, the systematic downshift of about 100 cm^{-1} for O−H vibrations is a well-known feature of GGA-type functionals.

4. CONCLUSIONS

In this work, we systematically benchmarked the semiempirical xTB methods (GFN1-xTB, GFN2-xTB and G-xTB), the MACE-OMol-0 machine-learning interatomic potential (MLIP), and the classical CHARMM36 force field against high-level DFT reference methods. The aim has been to identify computational approaches that provide an optimal balance between accuracy and efficiency for modeling cellulose systems. As model systems, we considered several conformers of cellobiose monomers and dimers.

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

Even at the DFT level, we observe that although the investigated functionals (ωB97M-V, B3LYP-D3, and PBE-D3) provide qualitatively consistent results, noticeable quantitative differences remain for the investigated properties. Using ωB97M-V as the reference, the PBE-D3 results exhibit systematically larger deviations from the reference than B3LYP-D3. The H···O hydrogen-bond distances are systematically shorter with PBE-D3, with differences up to 0.12 Å relative to ωB97M-V. Similarly, the relative stabilities of the monomers follow the same ordering, ωB97M-V > B3LYP-D3 > PBE-D3, with differences between ωB97M-V and PBE-D3 at most 0.2 eV. For the dimers, we find no such trend. The interaction energies show smaller, but noticeable, differences in the energy−distance potential curves. The vibrational frequencies exhibit larger RMSDs for PBE-D3 than for B3LYP-D3, compared to ωB97M-V. Nevertheless, PBE-D3 remains a practical choice for larger and periodic systems, such as bulk cellulose structures, where computational cost becomes a limiting factor.

The investigated xTB methods are approximately one to two orders of magnitude faster than DFT calculations. However, we find that xTB suffers from some distortions in the potential energy surface that favor puckered conformations, particularly for the GFN-xTB models. This limitation should be considered when studying clustered cellulose configurations, including small cellulose assemblies and surface models, where weaker intermolecular packing constraints allow individual chains to adopt such puckered geometries. Moreover, we find that these energetic discrepancies propagate into intermolecular interaction energies, artificially improving the energetic agreement with DFT. Without the conformational analysis, this discrepancy would likely have gone unnoticed. G-xTB partially mitigates these shortcomings and improves the agreement with the ωB97M-V reference, albeit at an increased computational cost. Harmonic vibrational frequencies are generally underestimated with the xTB methods compared to the three DFT methods, especially in the high-frequency region associated with O−H and C−H stretching modes, although the overall spectral features remain qualitatively consistent with the DFT results.

Overall, the MACE-OMol-0 MLIP model shows remarkably good agreement with the ωB97M-V reference while offering several orders of magnitude lower computational cost. However, in two of the dimer cases (Figure 6b and d), it fails to reproduce the hybrid DFT structures, and consequently also the energy-distance potential curve, especially of the *chain−chain* dimer. This highlights the importance of system-dependent model validation and fine-tuning where needed.

We also benchmarked CHARMM36, a widely used and computationally efficient classical force field for large-scale cellulose simulations. Our study shows that its structural and energetic accuracy compared to ωB97M-V is quite reasonable, but that force field parameter refinement would be necessary to achieve consistently reliable predictions.

Overall, judging from these small-system explorations only, the MACE-OMol-0 model appears to be a very promising choice for accelerating cellulose simulations while maintaining near-DFT accuracy. The semiempirical methods also show promise for applications where qualitative accuracy is sufficient, particularly G-xTB, which was found to alleviate some of the potential energy distortions identified in the GFN*n*-xTB models.

Altogether, our study highlights both successes and issues that call for caution when using approximate computational methods to accelerate energy and structure calculations for cellulose systems. Further validation and benchmarking, including condensed-matter cellulose systems, are needed to establish the transferability of these findings to realistic cellulose environments. Extending this work to crystalline and hydrated cellulose structures would be a valuable next step, as hydrogen bonding and competing intermolecular interactions in these systems may pose additional challenges for approximate computational methods. For the semiempirical quantum-mechanical methods and the advanced Machine-learned interatomic potentials, such as those from the MACE family, performing targeted validation and fine-tuning on cellulose-specific configurations could improve their accuracy compared to DFT and enhance their transferability to crystalline and hydrated environments. However, the small cellulose systems that were treated in this study already allowed us to discuss both H-bond-dominated and dispersion-dominated cellulose systems and to scrutinize results that were found to deviate between the methods in spite of the fact that, at first glance, the resulting optimized structures appeared to be reassuringly similar. This study allowed us to present results from some interesting new approximate methods and assess them in relation to methods that are already well established in the cellulose research community. Hopefully our findings can offer a guide in the selection process among computationally efficient and adequately accurate methods for cellulose simulations.

■ ASSOCIATED CONTENT

Data Availability Statement

The data in this study are available in the published article, in Supporting Information, and at https://github.com/feliciaandersson/Intra-and-intermolecular-interactions-in-cellobiose-benchmarking-accelerated-computational-methods. The scripts used to obtain the results with different methods are available at https://github.com/feliciaandersson/Intra-and-intermolecular-interactions-in-cellobiose-benchmarking-accelerated-computational-methods.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.6c01354.

Additional structural analyses of cellobiose monomers and dimers, including conformational motifs identified by CREST; comparisons of covalent bond lengths and hydrogen-bond geometries across computational methods; relative conformer energies; vibrational frequency correlations with the ωB97M-V reference; and computed infrared spectra obtained with DFT, GFN2-xTB, G-xTB, and MACE-OMol-0 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Felicia Andersson − *Department of Chemistry-Ångström, Uppsala University, S-75231 Uppsala, Sweden;* orcid.org/0009-0006-5782-0992; Email: felicia.e.andersson@kemi.uu.se

Kersti Hermansson − *Department of Chemistry-Ångström, Uppsala University, S-75231 Uppsala, Sweden;*

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

orcid.org/0000-0003-2352-0458; Email: kersti@kemi.uu.se

Peter Broqvist − *Department of Chemistry-Ångström, Uppsala University, S-75231 Uppsala, Sweden;* orcid.org/0000-0002-9842-4332; Phone: +46 184716274; Email: peter.broqvist@kemi.uu.se

Authors

Shokirbek Shermukhamedov − *Department of Chemistry-Ångström, Uppsala University, S-75231 Uppsala, Sweden;* orcid.org/0000-0002-3429-7493

Valentina Basoli − *Medical Additive Manufacturing Research Group (Swiss MAM) Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*

Fabiana Arduini − *Department of Chemical Science and Technologies, University of Rome Tor Vergata, 00133 Rome, Italy;* orcid.org/0000-0001-8990-6922

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.6c01354

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the National Academic Infrastructure for Supercomputing in Sweden (NAISS) for providing the computer resources used in this project. We also acknowledge financial support from the Swedish Research Council (Vetenskapsrådet), the National Strategic e-Science Program eSSENCE, and the Origami Paper-based tecHnology fOr the innovativE aNd sustaInable Organ-on-Chip devices (Phoenix-OoC) project, funded under the European Union's Horizon Europe EIC 2023 Pathfinder Open programme, grant agreement No 101130395.

■ ABBREVIATIONS

DFT	density functional theory
MLIP	machine-learning interatomic potential
ML	machine-learning
MD	molecular dynamics
SCC	self-consistent charge
DFTB	density functional tight binding
xTB	extended tight binding
CREST	Conformer−Rotamer Ensemble Sampling Tool
LBFGS	limited-memory Broyden−Fletcher−Goldfarb−Shanno
ASE	atomic simulation environment
VASP	Vienna Ab initio simulation package
PAW	projector augmented wave
PES	potential energy surrface
MACE	Message Passing Atomic Cluster Expansion

■ REFERENCES

(1) Peter, Z. Order in cellulosics: Historical review of crystal structure research on cellulose. *Carbohydr. Polym.* **2021**, *254*, No. 117417.

(2) Uto, T.; Yui, T. DFT Optimization of Isolated Molecular Chain Sheet Models Constituting Native Cellulose Crystal Structures. *ACS Omega* **2018**, *3*, 8050−8058.

(3) Belcastro, L.; Arduini, F. Recent advances in electrochemical paper-based analytical devices for drug analyses. *Electrochim. Acta* **2025**, *541*, No. 147084.

(4) Arduini, F. Paper as a Sustainable Material for Smart Electrochemical (Bio)sensors with Unprecedented Features: A Perspective. *Anal. Chem.* **2025**, *97*, 10126−10138.

(5) Tong, Y.; Cai, Y.; Wu, Y.; Zhuo, W.; Liao, J. From Design to Application: Advanced Cellulose Scaffolds for Engineered Tissue Regeneration. *Polymers* **2026**, *18*, No. 614.

(6) Afshang, A.; Jalali, S.; Amiryaghoubi, N. Cellulose-Based Scaffolds for Tissue Engineering. *Iran. J. Chem. Chem. Eng.* **2024**, *43*, 499−521.

(7) Vázquez-Rivas, E.; Desales-Guzmán, L. A.; Pacheco-Sánchez, J. H.; Burillo-Amezcua, S. G. Cellulose-Based Hybrid Hydrogels for Tissue Engineering Applications: A Sustainable Approach. *Gels* **2025**, *11*, No. 438, DOI: 10.3390/gels11060438.

(8) Wohlert, M.; Benselfelt, T.; Wågberg, L.; Furó, I.; Berglund, L. A.; Wohlert, J. Cellulose and the role of hydrogen bonds: not in charge of everything. *Cellulose* **2022**, *29*, 1−23.

(9) Mazumder, S.; Golbabaei, M. H.; Zhang, N. Advances in Computational Modeling and Machine Learning of Cellulosic Biopolymers: A Comprehensive Review. *Biomimetics* **2025**, *10*, No. 802.

(10) Kalita, B.; Gokcan, H.; Isayev, O. Machine learning interatomic potentials at the centennial crossroads of quantum mechanics. *Nat. Comput. Sci.* **2025**, *5*, 1120−1132.

(11) Batatia, I.; Kovács, D. P.; Simm, G. N. C.; Ortner, C.; Csányi, G. M. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. https://www.proceedings.com/068431-0830.html.

(12) Khodayari, A.; Hirn, U.; Spirk, S.; Ogawa, Y.; Seveno, D.; Thielemans, W. Advancing plant cell wall modelling: Atomistic insights into cellulose, disordered cellulose, and hemicelluloses - A review. *Carbohydr. Polym.* **2024**, *343*, No. 122415.

(13) Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; Gunsteren, W. F. v.; Hünenberger, P. H. Conformational and Dynamical Properties of Disaccharides in Water: a Molecular Dynamics Study. *Biophys. J.* **2006**, *90*, 4337−4344.

(14) Stortz, C. A.; Johnson, G. P.; French, A. D.; Csonka, G. I. Comparison of different force fields for the study of disaccharides. *Carbohydr. Res.* **2009**, *344*, 2217−2228.

(15) Wang, D.; Ámundadóttir, M. L.; van Gunsteren, W. F.; Hünenberger, P. H. Intramolecular hydrogen-bonding in aqueous carbohydrates as a cause or consequence of conformational preferences: a molecular dynamics study of cellobiose stereoisomers. *Eur. Biophys J.* **2013**, *42*, 521−537.

(16) Bazooyar, F.; Momany, F. A.; Bolton, K. Validating empirical force fields for molecular-level simulation of cellulose dissolution. *Comput. Theor. Chem.* **2012**, *984*, 119−127.

(17) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W. I.; MacKerell, A. D. J. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162−3180.

(18) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.

(19) Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8*, 4681−4690.

(20) Elahi, A.; Yan, X.; Chaudhuri, S. Enhancing the OPLS-AA force field for cellulose Iβ: structural stability and surface functionalization capability with the CM5 charge model. *Carbohydr. Polym.* **2025**, *360*, No. 123572.

(21) Matthews, J. F.; Beckham, G. T.; Bergenstråhle-Wohlert, M.; Brady, J. W.; Himmel, M. E.; Crowley, M. F. Comparison of Cellulose Iβ Simulations with Three Carbohydrate Force Fields. *J. Chem. Theory Comput.* **2012**, *8*, 735−748.

(22) Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr CHARMM Additive All-Atom Force Field for Aldopentofuranoses, Methyl-

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

aldopentofuranosides, and Fructofuranose. *J. Phys. Chem. B* **2009**, *113*, 12466−12476.
(23) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D., Jr. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543−2564.
(24) Momany, F. A.; Schnupf, U. DFTMD studies of β-cellobiose: conformational preference using implicit solvent. *Carbohydr. Res.* **2011**, *346*, 619−630.
(25) Murillo, J. D.; Moffet, M.; Biernacki, J. J.; Northrup, S. High-temperature molecular dynamics simulation of cellobiose and maltose. *AIChE J.* **2015**, *61*, 2562−2570.
(26) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(27) Frauenheim, T.; Seifert, G.; Elsterner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. A Self-Consistent Charge Density-Functional Based Tight-Binding Method for Predictive Materials Simulations in Physics, Chemistry and Biology. *Physica Status Solidi (b)* **2000**, *217*, 41−62.
(28) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(29) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1−86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.
(30) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, No. 154104.
(31) Caldeweyher, E.; Mewes, J.-M.; Ehlert, S.; Grimme, S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8499−8512.
(32) Wittmann, L.; Gordiy, I.; Friede, M.; Helmich-Paris, B.; Grimme, S.; Hansen, A.; Bursch, M. Extension of the D3 and D4 London dispersion corrections to the full actinides series. *Phys. Chem. Chem. Phys.* **2024**, *26*, 21379−21394.
(33) Pinto, A.; Llanos, A.; Gomila, R. M.; Frontera, A.; Rodríguez, L. Ligand and Gold(I) Fluorescein-AIEgens as Photosensitizers in Solution and Doped Polymers. *Inorg. Chem.* **2023**, *62*, 7131−7140.
(34) Zhang, Q. H.; Xu, N.; Jiang, Z. N.; Liu, H. F.; Zhang, G. A. Chitosan derivatives as promising green corrosion inhibitors for carbon steel in acidic environment: Inhibition performance and interfacial adsorption mechanism. *J. Colloid Interface Sci.* **2023**, *640*, 1052−1067.
(35) Ha, N. T. T.; Nguyen Thi Kim, G.; Nguyen Ngoc, H.; Lan, P. T. Understanding the interaction in cellulose-chitosan composite and its adsorption ability for Nickel (II): a theoretical investigation. *Mol. Simul.* **2023**, *49*, 1303−1310.
(36) Froitzheim, T.; Müller, M.; Hansen, A.; Grimme, S. g-xTB: A General-Purpose Extended Tight-Binding Electronic Structure Method For the Elements H to Lr (Z = 1−103) 2025 https://chemrxiv.org/doi/10.26434/chemrxiv-2025-bjxvt.
(37) Barrios, N.; Parra, J. G.; Iza, P.; Venditti, R. A.; Pal, L. Density Functional Theory Study of the Cellobiose Model System for Unveiling Cellulose-Water Interactions to Enhance Dewatering and Drying. *J. Phys. Chem. A* **2025**, *129*, 5018−5033.
(38) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. Neutron Crystallography, Molecular Dynamics, and Quantum Mechanics Studies of the Nature of Hydrogen Bonding in Cellulose Iβ. *Biomacromolecules* **2008**, *9*, 3133−3140.
(39) Nishiyama, Y.; Langan, P.; Chanzy, H. Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2002**, *124*, 9074−9082.
(40) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. Crystal Structure and Hydrogen Bonding System in Cellulose Iα from Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2003**, *125*, 14300−14306.
(41) Atalla, R. H.; VanderHart, D. L. Native Cellulose: A Composite of Two Distinct Crystalline Forms 1984 https://www.science.org/doi/10.1126/science.223.4633.283.
(42) Sugiyama, J.; Vuong, R.; Chanzy, H. Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* **1991**, *24*, 4168−4175.
(43) VanderHart, D. L.; Atalla, R. H. Studies of microstructure in native celluloses using solid-state carbon-13 NMR. *Macromolecules* **1984**, *17*, 1465−1472.
(44) Strati, G. L.; Willett, J. L.; Momany, F. A. Ab initio computational study of β-cellobiose conformers using B3LYP/6−311++G**. *Carbohydr. Res.* **2002**, *337*, 1833−1849.
(45) Bosma, W. B.; Appell, M.; Willett, J. L.; Momany, F. A. Stepwise hydration of cellobiose by DFT methods: 1. Conformational and structural changes brought about by the addition of one to four water molecules. *THEOCHEM* **2006**, *776*, 1−19.
(46) Yan, S.; Yao, L. DFT application in conformational determination of cellobiose. *Carbohydr. Res.* **2015**, *404*, 117−123.
(47) French, A.; Johnson, G. Roles of starting geometries in quantum mechanics studies of cellobiose. *Mol. Simul.* **2008**, *34*, 365−372.
(48) Pincu, M.; Brauer, B.; Gerber, R. B. When a proton attacks cellobiose in the gas phase: ab initio molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15382.
(49) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST—A program for the exploration of low-energy molecular chemical space. *J. Chem. Phys.* **2024**, *160*, No. 114110.
(50) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, No. 214110.
(51) Neese, F. Software update: the ORCA program system, version 5.0. *WIRES Comput. Mol. Sci.* **2022**, *12*, No. e1606.
(52) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.:Condens. Matter* **2017**, *29*, No. 273002.
(53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.et al. *Gaussian 16 Revision C.01*; Gaussian, Inc.: Wallingford, CT, 2016.
(54) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(55) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(56) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **1994**, *49*, 14251−14269.
(57) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(58) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169−11186.
(59) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758−1775.
(60) Hourahine, B.; Aradi, B.; Blum, V.; et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.
(61) grimme-lab Release v1.0.0 · grimme-lab/g-xtb. https://github.com/grimme-lab/g-xtb/releases/tag/v1.0.0.

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

(62) Gomes, T. C. F.; Skaf, M. S. Cellulose-Builder: A toolkit for building crystalline structures of cellulose. *J. Comput. Chem.* **2012**, *33*, 1338−1346.
(63) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, No. 108171.
(64) Levine, D. S.; Shuaibi, M.; Spotte-Smith, E. W. C.et al. The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models,arXiv:2505.08762 [physics.chem-ph]. 2026 http://arxiv.org/abs/2505.08762.
(65) Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **2015**, *108*, 1−5.
(66) Skelton, J. M.; Burton, L. A.; Jackson, A. J.; Oba, F.; Parker, S. C.; Walsh, A. Lattice dynamics of the tin sulphides SnS2, SnS and Sn2S3: vibrational spectra and thermal transport. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12452−12465.
(67) Cocinero, E. J.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. The Building Blocks of Cellulose: The Intrinsic Conformational Structures of Cellobiose, Its Epimer, Lactose, and Their Singly Hydrated Complexes. *J. Am. Chem. Soc.* **2009**, *131*, 11117−11123.
(68) Pincu, M.; Gerber, R. B. Hydration of cellobiose: Structure and dynamics of cellobiose−$(H_2O)n$, n = 525. *Chem. Phys. Lett.* **2012**, *531*, 52−58.
(69) Tsuboi, M. Infrared spectrum and crystal structure of cellulose. *J. Polym. Sci.* **1957**, *25*, 159−171.
(70) Kameneva, S. V.; Popkov, M. A.; Kozlova, T. O.; Luginina, A. A.; Yapryntsev, A. D.; Baranchikov, A. E.; Ivanov, V. K. UV-shielding photochromic composite films based on nanocrystalline cellulose modified with CeO2 and WO3 nanoparticles. *Sci. Rep.* **2026**, *16*, No. 2516.
(71) Pincu, M.; Cocinero, E. J.; Mayorkas, N.; Brauer, B.; Davis, B. G.; Gerber, R. B.; Simons, J. P. Isotopic Hydration of Cellobiose: Vibrational Spectroscopy and Dynamical Simulations. *J. Phys. Chem. A* **2011**, *115*, 9498−9509.

Downloaded from pubs.acs.org/jctcce/article-pdf/22/22/12223/6911670/acs.jctc.6c01354.pdf by UNIV DEGLI STUDI DI TOR VERGATA user on 25 November 2026

https://doi.org/10.1021/acs.jctc.6c01354
J. Chem. Theory Comput. 2026, 22, 12223−12238