

ARTICLE OPEN



Glass transition temperature prediction of disordered molecular solids

Kun-Han Lin¹✉, Leanne Paterson¹, Falk May² and Denis Andrienko¹✉

Glass transition temperature, T_g , is the key quantity for assessing morphological stability and molecular ordering of films of organic semiconductors. A reliable prediction of T_g from the chemical structure is, however, challenging, as it is sensitive to both molecular interactions and analysis of the heating or cooling process. By combining a fitting protocol with an automated workflow for forcefield parameterization, we predict T_g with a mean absolute error of $\sim 20^\circ\text{C}$ for a set of organic compounds with T_g in the $50\text{--}230^\circ\text{C}$ range. Our study establishes a reliable and automated prescreening procedure for the design of amorphous organic semiconductors, essential for the optimization and development of organic light-emitting diodes.

npj Computational Materials (2021)7:179; <https://doi.org/10.1038/s41524-021-00647-w>

INTRODUCTION

Amorphous molecular solids are materials composed of organic molecules with a disordered molecular packing. They are widely used in electronic and opto-electronic applications due to their flexibility, light-weight, solution-processability and tunability through versatile organic synthesis^{1–5}. For example, they serve as charge carrier transport materials^{6,7} and host materials^{8,9} in organic light-emitting diodes (OLEDs) and perovskite solar cells.

In addition to the electronic properties of these materials, the glass transition temperature, T_g , is especially important as it determines the morphological stability and hence the lifetime of electronic devices¹⁰. The search for high- T_g compounds is therefore crucial for all electronic applications. In this context, computer simulation protocols capable of reliable T_g -prediction are invaluable. Various efforts have been made in the past to achieve accurate T_g -prediction using statistical approaches, such as quantitative structure-property relationships^{11–14} and machine learning^{15–17}, as well as the direct extraction from molecular dynamics (MD) simulations^{18–21}. In the latter approach, reliable- T_g prediction has been achieved for a particular material through tailor-made forcefield parameters²². A systematic study of a dataset of conjugated small molecules using a deterministic approach is, however, lacking. This is presumably due to (1) the considerable computational cost and human effort involved in forcefield parameterization and molecular dynamics simulations and (2) the difficulty of finding a forcefield parameterization protocol that works for a dataset of molecules composed of different molecular building blocks.

In our recent work^{23,24}, we developed a multiscale computational protocol that successfully predicts relevant physical properties of OLED host materials, including solid-state ionization energy, density of states of charge carriers and the charge carrier mobility. Despite this success, the predicted T_g values are significantly off the experimental values. Since all torsional parameters were reparametrized with Lennard–Jones parameters from the OPLS database and Merz–Singh–Kollman partial charges^{25,26}, we conclude that the non-bonded terms^{24,27,28} should also be adjusted.

Here, we propose a computational methodology for reliable T_g -prediction of disordered organic molecular solids involving two

parts: (1) fitting protocol for a density–temperature plot and (2) non-bonded forcefield parameterization via atoms-in-molecule electron density partitioning^{29,30}. This methodology is tested on common organic semiconducting OLED hosts and hole-transporting materials, the chemical structures of which are shown in Fig. 1. The computed T_g values are in good agreement with the experimental T_g from our measurements and literature^{31–39}, with a mean absolute error (MAE) of 20.5°C , excluding two outliers.

RESULTS

Conventional bilinear fit

Conventionally, the simulated T_g is obtained via a bilinear fit, as an intersection of two lines in a density-temperature (ρ – T) plot, as shown in Fig. 2a. Ideally, two plateaus representing two linear regions should be observed in a slope–temperature plot. However, this is often not the case for data extracted from MD trajectories, as shown in Fig. 2b and Supplementary Fig. 1. For BCP, for example, a transition region II between regions I and III could be identified based on the change in the slope, as shown in Fig. 2b. The non-constant slope observed in Region I and III and the existence of Region IV may be because there is more than one mechanism responsible for the volume compression during the cooling step, and the cooling rate (100 K ns^{-1}) adopted in the simulations is too high. However, this non-ideal behavior persists even in the simulation with a 10 times slower cooling rate (10 K ns^{-1} , see Supplementary Fig. 2). Since MD simulations with a cooling rate of 10 K ns^{-1} and all-atom forcefields for 3000-molecule system (150,000–450,000 atoms) is already too costly for the computational screening, we aim at extracting T_g from a non-ideal ρ – T plot with a 100 K ns^{-1} cooling rate.

The ambiguity in choosing the fitting ranges for Region I and Region III leads to a human bias, which is also common in the fitting process of charge carrier mobility⁴⁰. In Fig. 2b, for example, we could choose any fitting range lying between point A and point B for Region I. Similarly, we could choose any fitting range between point C and point D for Region III. To quantify this

¹Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany. ²Merck KGaA, 64293 Darmstadt, Germany. ✉email: link@mpip-mainz.mpg.de; denis.andrienko@mpip-mainz.mpg.de

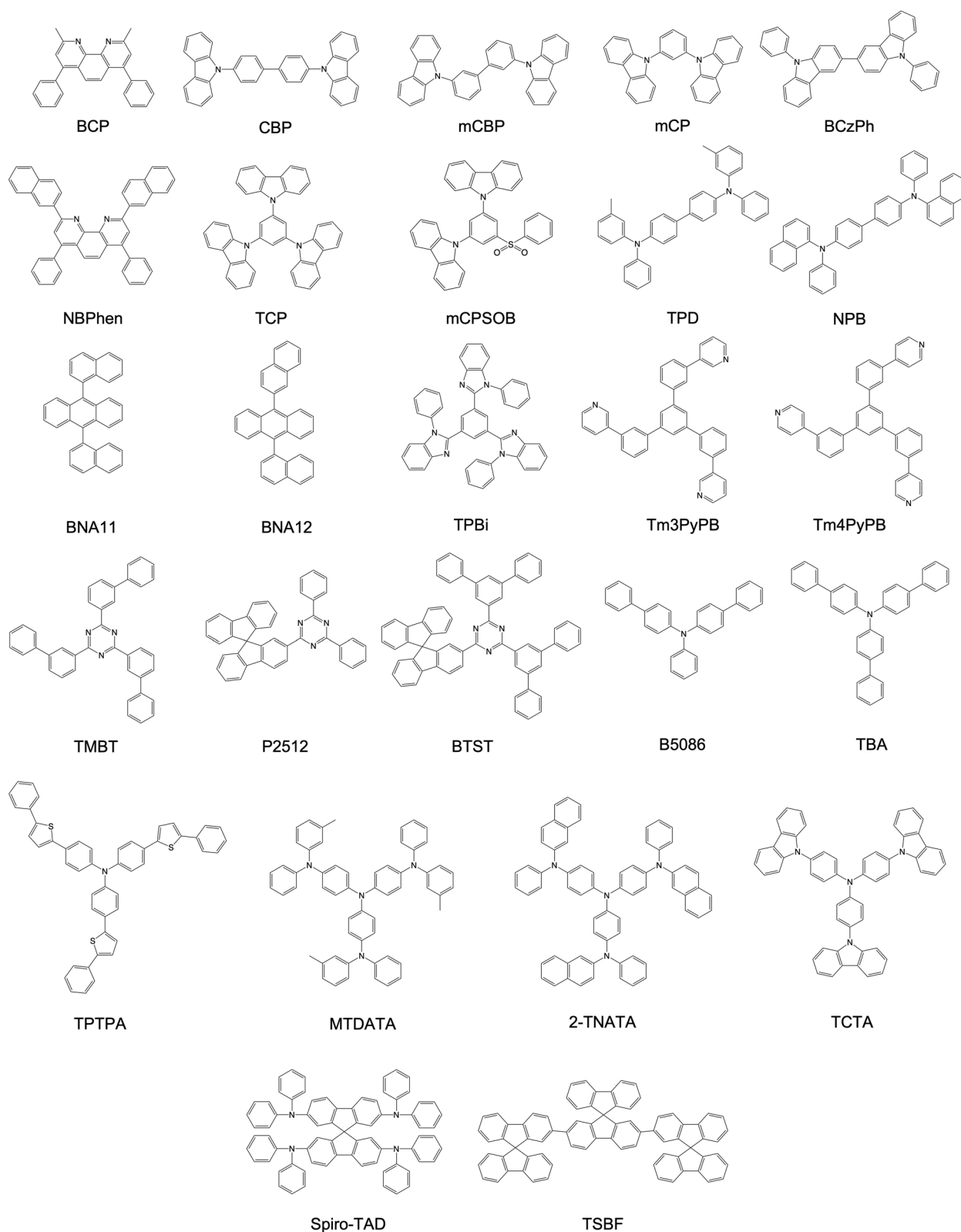


Fig. 1 Molecular structures of organic molecular solids. Chemical structures of 26 OLED host materials and charge transport materials investigated in this work.

variation, we show T_g calculated using four fitting ranges for Region I (A, B) and Region III (C, D) in Fig. 2c. For BCP this leads to variations from 51.2 °C to 147.7 °C, depending on the fitting ranges.

Determination of fitting ranges from a R^2 - T plot

To explore the best linear fitting ranges for Region I and III, we plotted the R^2 vs temperature (R^2 - T plot) with different sizes of the fitting range for BCP, as shown in Fig. 3. Starting from the range of

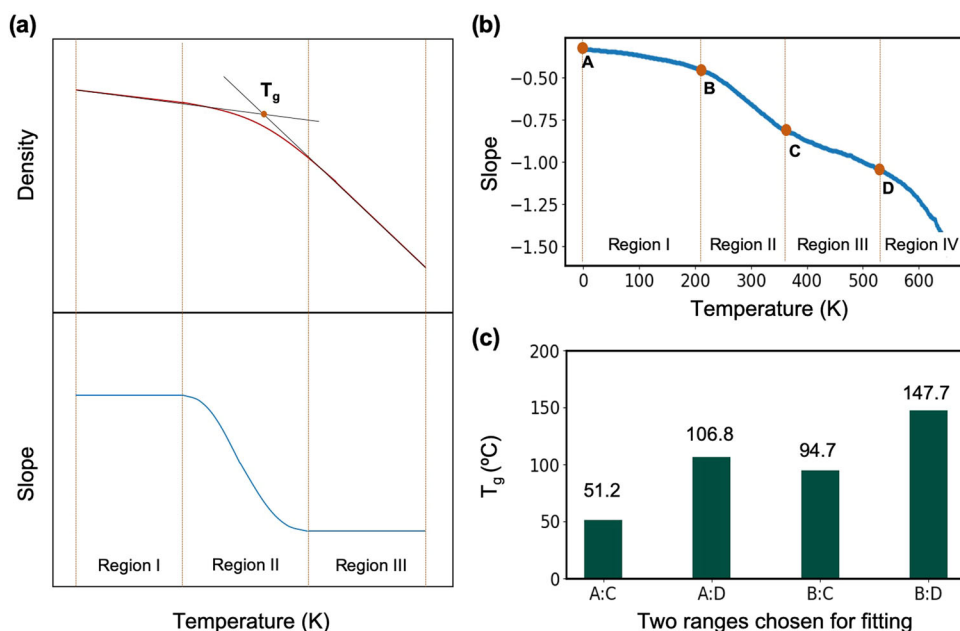


Fig. 2 Derivation of T_g via a bilinear fit. **a** A schematic illustration of a density–temperature and a slope–temperature plot characterizing a glass transition process. **b** Slope–temperature plot of BCP molecule. Each blue point on the plot represents a slope from a linear regression within the temperature range of $[T, T + 150]$. Four regions and corresponding edge points, A, B, C, D, are specified. **c** Fitted T_g using four different combinations of fitting ranges.

150 K, a clear valley region (Region II) that separates Region I from III can be observed. The decrease of R^2 in the valley region is a direct consequence of including a more non-linear transition region into the linear regression. A straightforward way of determining two optimal fitting ranges is then to pick two “hill tops”, corresponding to the maximum R^2 values, on the opposite sides of the valley. This protocol is generic, as is illustrated for all compounds in Supplementary Fig. 3, indicating that the R^2 – T plot does provide fitting ranges unambiguously.

The effect of the fitting ranges on T_g is shown in Supplementary Fig. 4. Overall, the variation of T_g due to the fitting range is much smaller than that observed in Fig. 2c. In most cases, the calculated T_g slightly rises with increasing fitting range. In some cases; however, the fitted T_g values exhibit complicated dependence on the fitting range (such as BTST and MTDATA). This originates from the synergistic effect of the aforementioned non-ideal ρ – T curve and the density fluctuation (see Supplementary Discussion for more details). This indicates that we cannot determine the optimal fitting range for each compound unbiasedly. Nevertheless, we can treat the fitting range as a parameter to be determined. Overall, the T_g calculated with the fitting ranges of 150 K and 200 K show better agreement with experimental values, as shown in Supplementary Fig. 4. Therefore, the fitting range of 200 K will be used for later comparison. The ρ – T plots and calculated T_g of the 26 compounds determined using the 200 K fitting range are shown in Supplementary Fig. 5.

Effect of VdW parameters

With the proposed fitting procedure at hand, we now study the impact of the forcefield on the predicted glass transition temperature. It has been shown that electronic properties, such as ionization energy, density of states and charge carrier mobility, of amorphous films simulated using the Merz–Singh–Kollman partial charges and OPLS VdW parameters are in good agreement with experimental values¹⁷. However, the T_g simulated using this forcefield, poorly agree with experiments, as shown in Fig. 4a and Supplementary Table 1. Assuming that the disagreement is due to non-bonded parameters, we reparametrized all non-bonded

parameters for each molecule employing electron density-based DDEC6 method^{18,19}. The predictions are also summarized in Fig. 4a. Compared to the MAE of 59.1 °C using the MK-based forcefield, DDEC6 gives a smaller MAE of 20.5 °C and a much higher R^2 (0.61 vs 0.87). The reported MAE excludes T_g of MTDATA and 2-TNATA, which are clear outliers for both methods (see Supplementary Table 1), and have a similar molecular structure, namely starburst dendrimers of triphenylamine (TPA)⁴¹. Since these two compounds are mostly composed of TPA blocks, the reason for the observed discrepancy may be the lack of non-local correlations in dihedral potentials of the TPA moiety, absent in both forcefields⁴². To correctly describe this behavior, an improved forcefield model that takes this correlation into account, either by a traditional fitting or a machine-learning approach³⁹, is required.

In addition, we compare our results with those obtained from the protocol proposed by Patrone et al., which also features a reduction of human analysis variation⁴³. The T_g is determined as the intersection of high- and the low-temperature asymptotes of a hyperbolic fit of the whole range of the data. As shown in Fig. 4b, T_g calculated using Patrone’s protocol gives reasonably high R^2 and a slope close to one. If the transferability of the fitted linear model is good, we can also predict T_g using this model by adding the intercept on top of the simulated T_g values. Overall, our protocol performs better in terms of both MAE, 64.7 vs 20.5 °C and R^2 , 0.75 vs 0.87.

Efficiency of our computational protocol

Since our goal is to set up an automated prescreening procedure, we have analyzed the average computational cost and human effort required for obtaining the T_g for one compound. The specification of CPUs used for each computation is listed in the Methods section. Out of 6 h of human effort per compound, ca 70% is spent on generating the molecular topology and parameterizing bonded and non-bonded parameters. This motivates us to further develop an automatic/semi-automatic generator for these files. As for the computational cost, the total time spent on computing the T_g using the DDEC protocol amounts to ~850,000 cpu hours, as shown in Fig. 5. Around 50% of the

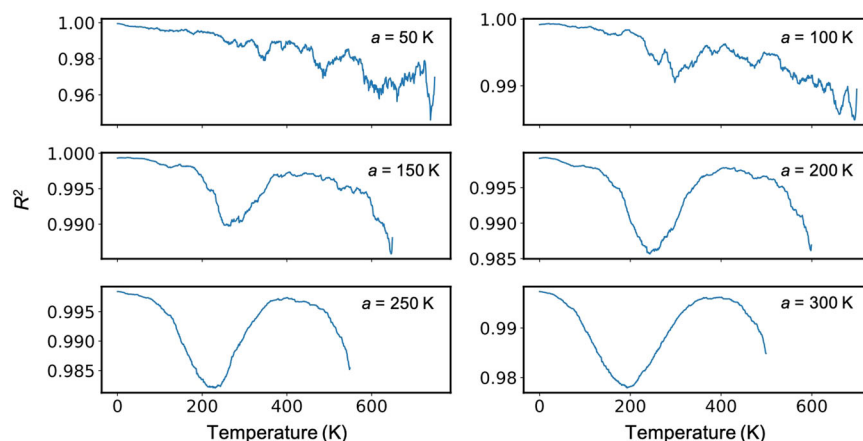


Fig. 3 The effect of fitting ranges on R^2 - T plot. The R^2 - T plot for BCP compound with six different fitting ranges. Each blue point in the plot represents an R^2 value fitted with a range $[T, T+a]$, where $a = 50, 100, 150, 200, 250,$ and 300 K.

resources are used in compressing and heating the as-generated low-density morphology. Nearly 30 and 20% of cpu hours are utilized for the MD cooling step and QM constrained dihedral scanning, respectively. The step for computing DDEC charge and LJ parameters is only 0.1% of the total computational cost. It is clear that further optimization of the protocol to achieve better efficiency is necessary for a large-scale screening purpose. One potential direction is to reduce the computational cost of compressing and heating in MD simulations by using a smaller system and denser initial morphology.

DISCUSSION

In summary, we propose a computational methodology for reliable T_g prediction, including a fitting procedure and an MD simulation protocol using DDEC6 non-bonded parameters. Our fitting procedure reduces the human analysis variation (~ 90 °C for BCP) resulting from different chosen fitting ranges for linear regression. In combination, the proposed fitting procedure ensures the reliability of T_g , while the DDEC6 forcefield parameters improve T_g predictions. The obtained values agree well with experimental values, with MAE = 20.5 °C, acceptable for in silico prescreening of organic hosts. The overestimation of T_g for two starburst TPA-based dendrimers encourages further forcefield development, as the TPA moiety is ubiquitous in organic semiconductors⁴⁴. Furthermore, we generated OPLS-DDEC forcefield parameters for 26 compounds, which are available in the open-access git repository. These forcefield parameters, which give reasonable T_g , could also be useful for multiscale simulations of complicated guest-host or interfacial systems in OLEDs and in silico deposition protocols. Finally, given that the molecular fragments included in this molecular dataset are the most common building blocks of organic semiconductors, we expect our protocol could have wide applicability in this field.

EXPERIMENTAL SECTION

For determination of the glass transition temperature (T_g) at Merck KGaA, Darmstadt, Germany, we used differential scanning calorimetry (DSC) analyzing samples of 10–15 mg in DSC 204/1/G Phönix from Netzsch. Samples were heated by 5 °C min^{-1} up to 370 °C then cooled by 20 °C min^{-1} to 0 °C and finally heated again by 20 °C min^{-1} to 370 °C, where T_g was determined by the kink in heat flow vs temperature using the temperature corresponding to half the drop in heat flux. Only for BCP and TMBT, this protocol did not yield a significant kink. TMBT was expected to be the lowest T_g

material from simulation, so we tried other protocols to measure T_g . We finally used a 5 mg TMBT sample in DSC Discovery from TA Instruments in nitrogen atmosphere and first heated by 20 °C min^{-1} up to 320 °C, then for cooling quenched the sample by liquid nitrogen and finally heated by 20 °C min^{-1} up to 320 °C, where the T_g was observed. Other protocols we tried for TMBT without the cooling quench did not lead to the observation of T_g .

METHODS

Computational section

Our protocol for T_g prediction involves three steps: (i) forcefield parameterization, (ii) classical MD simulations and (iii) fitting procedure, which are described as following.

(i) Forcefield parameterization: All bonded parameters apart from proper and improper dihedrals were taken from OPLS-AA forcefield^{45,46} and our previous work²⁴. The non-bonded parameters, atomic partial charges and Lennard-Jones parameters, were derived following the protocol proposed by Cole et al.²⁹. In short, the overlapping atomic electron densities were obtained via the density-derived electrostatic and chemical (DDEC6) electron density partitioning scheme³⁰. The atomic partial charges can then be obtained by integrating the corresponding atomic electron densities over the whole space. Additionally, the two parameters, A and B, in Lennard-Jones potential are then derived using Tkatchenko–Scheffler (TS) scheme⁴⁷, where the radius of the free atom in a vacuum (R_i^{free}) is taken from ref. 29. The electron density was obtained using Gaussian16⁴⁸ at $\omega\text{B97X-D}^{32}/6\text{-}311\text{G(d,p)}$ level and the DDEC6 computations were performed using Chargemol of version 09_26_2017³⁰. All molecules considered in this work were partitioned into several rigid fragments following the same procedure as our previous work²⁴. After non-bonded parameters were set, the dihedral potentials that connect these rigid fragments, which are usually missing in the OPLS-AA database, were parameterized using the constrained optimization scanning performed at $\omega\text{B97X-D3/def2-TZVP}$ level using ORCA 4.2.1⁴⁹. For more details of the parameterization of dihedral potentials, please refer to ref. 50.

(ii) Classical MD simulations: All classical MD simulations were performed using GROMACS of version 2020.3^{51,52}. For the long-range electrostatic interactions, the particle mesh Ewald (PME) method was employed with a 0.12 nm Fourier spacing. A cutoff of 13 Å was applied to all non-bonded interactions. The temperature and pressure control were accomplished using velocity rescaling with a stochastic term⁵³ ($\tau_T = 0.5$ ps) and an isotropic coupling for the pressure from a Berendsen barostat ($P_0 = 1$ bar, $\chi = 4.5 \times 10^{-5}$ bar $^{-1}$, and $\tau_P = 0.5$ ps).

For each compound, 3000 molecules were initially randomly placed in a simulation box with a low target density around $50\text{--}150$ kg m^{-3} using Packmol⁵⁴. The whole system was then heated up from 100 K to 300 K at a rate of 0.67 K ps^{-1} . It was then equilibrated at 300 K until the density reached a steady value. This step helps to prevent the system from exploding due to a high heating rate. Finally, the system was heated up from 300 K to 800 K at a rate of 0.5 K ps^{-1} , followed by an equilibration at

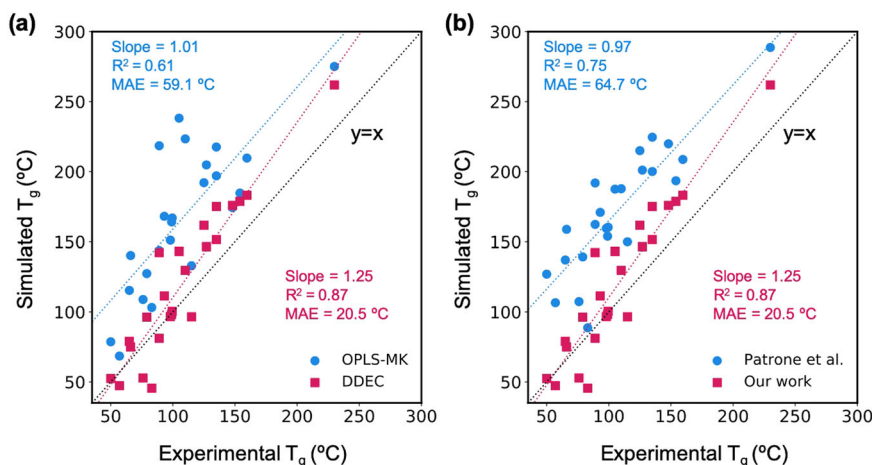


Fig. 4 Comparison between experimental and computational T_g values. The experimental and simulated T_g of 24 compounds (removing outliers MTDATA and 2-TNATA). **a** Simulated T_g calculated using different MD data (OPLS-MK vs DDEC6; see Methods) and the same fitting protocol (fitting size of 200 K). **b** Simulated T_g calculated using the same MD data (DDEC6) and different fitting protocols (Patrone et al. vs this work).

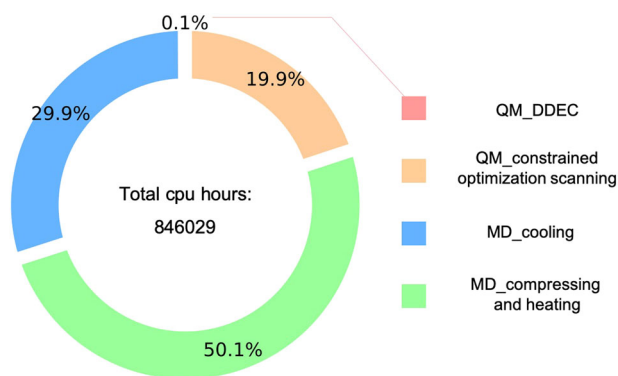


Fig. 5 Summary of the computational cost of our T_g protocol. Pie chart of the computational costs (in terms of cpu hours) of different steps summed over all compounds in our protocol.

800 K for 10 ns. The equilibration time is long enough to ensure a steady density of the system for all compounds discussed here. Finally, the system underwent a linear cooling procedure from 800 K to 0 K at a 100 K ns^{-1} cooling rate, where the density and temperature data were extracted for T_g fitting procedure.

The effect of cooling protocol (continuous vs stepwise) was also investigated. The cooling profiles for the stepwise cooling are shown in Supplementary Fig. 6. The density-temperature curves of mCP obtained using different conditions are very similar, as shown in Supplementary Fig. 7. For this reason, we put our main focus on other factors, such as non-bonded parameters and the fitting protocol. All T_g values shown in the main text are extracted from MD simulations with a continuous cooling protocol.

Furthermore, the effects of system size, initial configurations, barostats, cooling rates and density fluctuation were investigated for mCP (see Supplementary Discussion). We also evaluated our R^2 -based fitting protocol by considering the effect of the ideal/non-ideal ρ - T curve and the density fluctuation on our protocol. To conclude, our current computational workflow is useful for practical T_g prediction, reaching a balance of accuracy and the computational cost.

(iii) Fitting protocol: The simulated T_g is obtained via a bilinear fit, as an intersection of two lines in a density-temperature (ρ - T) plot. In order to obtain the optimal fitting range in a less biased way, we propose the following fitting strategy. We plotted the R^2 value as a function of temperature T for a series of linear regressions with fitting ranges $[T, T+a]$, where a is the fitting size (Fig. 3). In this plot, we can identify a clear valley

region (Region II) that separates Region I from III (Fig. 2). The fitting ranges for two linear regressions are chosen as the two “hill tops” (maximum R^2 value) on the opposite sides of the valley region. The effect of the fitting size is described in detail in the main text. The MK data shown in Fig. 1b are taken from our previous work²⁴, where the atomic charges are derived using Merz-Singh-Kollman scheme and VdW parameters are taken from OPLS forcefield database.

Finally, all QM computations (GAUSSIAN16 and ORCA) were performed using 20 CPUs of Intel Xeon Silver 4210 @ 2.2 GHz. All MD simulations were performed using (4 nodes; $4 \times 40 = 160$ CPUs) of Intel Skylake 6148 @ 2.4 GHz.

DATA AVAILABILITY

The input files of simulations, optimized forcefield parameters, the density-temperature data for all compounds are available in the Tg git repository, version 1.0: <https://gitlab.mpcdf.mpg.de/materials/tg>.

CODE AVAILABILITY

The python scripts for deriving VdW parameters and T_g -fitting are available in Tg git repository, version 1.0: <https://gitlab.mpcdf.mpg.de/materials/tg>.

Received: 28 June 2021; Accepted: 7 October 2021;

Published online: 08 November 2021

REFERENCES

- Kotadiya, N. B., Mondal, A., Blom, P. W. M., Andrienko, D. & Wetzelaer, G. J. A. H. A window to trap-free charge transport in organic semiconducting thin films. *Nat. Mater.* **18**, 1182–1186 (2019).
- Kotadiya, N. B. et al. Universal strategy for Ohmic hole injection into organic semiconductors with high ionization energies. *Nat. Mater.* **17**, 329–334 (2018).
- Lin, K.-H. et al. Multiarm and substituent effects on charge transport of organic hole transport materials. *Chem. Mater.* **31**, 6605–6614 (2019).
- Yin, X., Song, Z., Li, Z. & Tang, W. Toward ideal hole transport materials: a review on recent progress in dopant-free hole transport materials for fabricating efficient and stable perovskite solar cells. *Energy Environ. Sci.* **13**, 4057–4086 (2020).
- Chatterjee, T. & Wong, K.-T. Perspective on host materials for thermally activated delayed fluorescence organic light emitting diodes. *Adv. Opt. Mater.* **7**, 1–34 (2019).
- Jeong, J. et al. Pseudo-halide anion engineering for α -FAPbI₃ perovskite solar cells. *Nature* **592**, 381–385 (2021).
- Lin, K.-H., Prlj, A. & Corminboeuf, C. How does alkyl chain length modify the properties of triphenylamine-based hole transport materials? *J. Mater. Chem. C* **6**, 960–965 (2018).

8. Jeon, S. O. et al. High-efficiency, long-lifetime deep-blue organic light-emitting diodes. *Nat. Photonics* **15**, 208–215 (2021).
9. Tao, Y., Yang, C. & Qin, J. Organic host materials for phosphorescent organic light-emitting diodes. *Chem. Soc. Rev.* **40**, 2943 (2011).
10. Chatterjee, T. & Wong, K.-T. Perspective on Host Materials for Thermally Activated Delayed Fluorescence Organic Light Emitting Diodes. *Adv. Opt. Mater.* **7**, 1800565 (2019).
11. Xie, R. et al. Glass transition temperature from the chemical structure of conjugated polymers. *Nat. Commun.* **11**, 4–11 (2020).
12. Yin, S., Shuai, Z. & Wang, Y. A quantitative structure–property relationship study of the glass transition temperature of OLED materials. *J. Chem. Inf. Comput. Sci.* **43**, 970–977 (2003).
13. Barbosa-da-Silva, R. & Stefani, R. QSPR based on support vector machines to predict the glass transition temperature of compounds used in manufacturing OLEDs. *Mol. Simul.* **39**, 234–244 (2013).
14. Xu, J. & Chen, B. Prediction of glass transition temperatures of OLED materials using topological indices. *J. Mol. Model.* **12**, 24–33 (2005).
15. Kim, C., Chandrasekaran, A., Huan, T. D., Das, D. & Ramprasad, R. Polymer genome: a data-powered polymer informatics platform for property predictions. *J. Phys. Chem. C* **122**, 17575–17585 (2018).
16. Zhang, Y. & Xu, X. Machine learning glass transition temperature of polyacrylamides using quantum chemical descriptors. *Polym. Chem.* **12**, 843–851 (2021).
17. Pilonia, G., Iverson, C. N., Lookman, T. & Marrone, B. L. Machine-learning-based predictive modeling of glass transition temperatures: a case of polyhydroxyalkanoate homopolymers and copolymers. *J. Chem. Inf. Model.* **59**, 5013–5025 (2019).
18. Afzal, M. A. F. et al. High-throughput molecular dynamics simulations and validation of thermophysical properties of polymers for various applications. *ACS Appl. Polym. Mater.* **3**, 620–630 (2021).
19. Odinokov, A., Freidzon, A. & Bagaturyants, A. Molecular dynamics simulation of the glass transition in 4,4'-N,N'-dicarbazolylbiphenyl. *Chem. Phys. Lett.* **633**, 41–46 (2015).
20. Root, S. E., Savagatrup, S., Pais, C. J., Arya, G. & Lipomi, D. J. Predicting the mechanical properties of organic semiconductors using coarse-grained molecular dynamics simulations. *Macromolecules* **49**, 2886–2894 (2016).
21. Lin, K.-H., Prlj, A. & Corminboeuf, C. A. Rising star: truxene as a promising hole transport material in perovskite solar cells. *J. Phys. Chem. C* **121**, 21729–21739 (2017).
22. McAlliley, J. H. & Bruce, D. A. Development of force field parameters for molecular simulation of polylactide. *J. Chem. Theory Comput.* **7**, 3756–3767 (2011).
23. Paterson, L., May, F. & Andrienko, D. Computer aided design of stable and efficient OLEDs. *J. Appl. Phys.* **128**, 160901 (2020).
24. Mondal, A. et al. Molecular library of OLED host materials—evaluating the multiscale simulation workflow. *Chem. Phys. Rev.* **2**, 031304 (2021).
25. Empereur-Mot, C. et al. Swarm-CG: automatic parametrization of bonded terms in MARTINI-based coarse-grained models of simple to complex molecules via fuzzy self-tuning particle swarm optimization. *ACS Omega* **5**, 32823–32843 (2020).
26. Stankevych, A. et al. Density of states of OLED host materials from thermally stimulated luminescence. *Phys. Rev. Appl.* **15**, 044050 (2021).
27. Bejagam, K. K., Singh, S., An, Y. & Deshmukh, S. A. Machine-learned coarse-grained models. *J. Phys. Chem. Lett.* **9**, 4667–4672 (2018).
28. Bejagam, K. K., Singh, S., An, Y., Berry, C. & Deshmukh, S. A. PSO-assisted development of new transferable coarse-grained water models. *J. Phys. Chem. B* **122**, 1958–1971 (2018).
29. Cole, D. J., Vilseck, J. Z., Tirado-Rives, J., Payne, M. C. & Jorgensen, W. L. Biomolecular force field parameterization via atoms-in-molecule electron density partitioning. *J. Chem. Theory Comput.* **12**, 2312–2323 (2016).
30. Manz, T. A. & Limas, N. G. Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv.* **6**, 47771–47801 (2016).
31. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615 (2008).
32. Li-Ying, S. et al. Improving the efficiency of blue organic light-emitting diodes by employing Cs-derivatives as the n-dopant. *Acta Phys.—Chim. Sin.* **28**, 1497–1501 (2012).
33. Wang, Z. et al. Phenanthro[9,10-d]imidazole as a new building block for blue light emitting materials. *J. Mater. Chem.* **21**, 5451 (2011).
34. Park, K. N., Cho, Y.-R., Kim, W., Park, D.-W. & Choe, Y. Raman spectra and current-voltage characteristics of 4,4',4''-tris(2-naphthylphenylamino)triphenylamine thin films. *Mol. Cryst. Liq. Cryst.* **498**, 183–192 (2009).
35. Gaj, M. P., Fuentes-Hernandez, C., Zhang, Y., Marder, S. R. & Kippelen, B. Highly efficient Organic Light-Emitting Diodes from thermally activated delayed fluorescence using a sulfone–carbazole host material. *Org. Electron.* **16**, 109–112 (2015).
36. Wong, K.-T. et al. Ter(9,9-diarylfuorene)s: highly efficient blue emitter with promising electrochemical and thermal stability. *J. Am. Chem. Soc.* **124**, 11576–11577 (2002).
37. Kageyama, H., Ohishi, H., Tanaka, M., Ohmori, Y. & Shirota, Y. High-performance organic photovoltaic devices using a new amorphous molecular material with high hole drift mobility, Tris[4-(5-phenylthiophen-2-yl)phenyl]amine. *Adv. Funct. Mater.* **19**, 3948–3955 (2009).
38. Shahnowaz, S. et al. Hole-transporting materials for organic light-emitting diodes: an overview. *J. Mater. Chem. C* **7**, 7144–7158 (2019).
39. Su, S.-J., Takahashi, Y., Chiba, T., Takeda, T. & Kido, J. Structure-property relationship of pyridine-containing triphenyl benzene electron-transport materials for highly efficient blue phosphorescent OLEDs. *Adv. Funct. Mater.* **19**, 1260–1267 (2009).
40. Blakesley, J. C. et al. Towards reliable charge-mobility benchmark measurements for organic semiconductors. *Org. Electron.* **15**, 1263–1272 (2014).
41. You, J., Li, G. & Wang, Z. Starburst dendrimers consisting of triphenylamine core and 9-phenylcarbazole-based dendrons: synthesis and properties. *Org. Biomol. Chem.* **10**, 9481 (2012).
42. Friederich, P., Konrad, M., Strunk, T. & Wenzel, W. Machine learning of correlated dihedral potentials for atomistic molecular force fields. *Sci. Rep.* **8**, 2559 (2018).
43. Patrone, P. N., Dienstfrey, A., Browning, A. R., Tucker, S. & Christensen, S. Uncertainty quantification in molecular dynamics studies of the glass transition temperature. *Polym. (Guildf.)* **87**, 246–259 (2016).
44. Agarwala, P. & Kabra, D. A review on triphenylamine (TPA) based organic hole transport materials (HTMs) for dye sensitized solar cells (DSSCs) and perovskite solar cells (PSCs): evolution and molecular engineering. *J. Mater. Chem. A* **5**, 1348–1373 (2017).
45. Jorgensen, W. L. & Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl Acad. Sci.* **102**, 6665–6670 (2005).
46. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
47. Tkatchenko, A. & Scheffler, M. Accurate molecular Van Der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **102**, 073005 (2009).
48. Frisch, M. J. et al. *Gaussian 16*, Revision B.01. (Gaussian, Inc., 2016).
49. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).
50. Poelking, C. et al. Characterization of charge-carrier transport in semicrystalline polymers: electronic couplings, site energies, and charge-carrier dynamics in poly(bithiophene-alt-thienothiophene) [PBTTT]. *J. Phys. Chem. C* **117**, 1633–1640 (2013).
51. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
52. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
53. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
54. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).

ACKNOWLEDGEMENTS

D.A. acknowledges the BMBF Grant InterPhase (No. FKZ13N13661) and the European Union Horizon 2020 Research and Innovation Program 'Widening Materials Models' under Grant Agreement No. 646259 (MOSTOPHOS). This research has been supported by the King Abdullah University of Science and Technology (KAUST), via the Competitive Research Grants (CRG) Program. D.A. acknowledges KAUST for hosting his sabbatical. DFG is acknowledged for financial support through the collaborative research center TRR 146. K.-H. L. acknowledges the financial support from the Swiss NSF Early Postdoc Mobility fellowship (grant no. P2ELP2_195156).

AUTHOR CONTRIBUTIONS

K.-H.L. conceived the study, performed forcefield parameterization, MD simulations, data analysis, and prepared the online git repository and the first draft of the manuscript. L.P. performed forcefield parameterization, discussed results, and

contributed to writing. F.M. performed forcefield parameterization, provided the measured values, discussed results, and contributed to writing. D.A. supervised the work, discussed results, and contributed to writing.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-021-00647-w>.

Correspondence and requests for materials should be addressed to Kun-Han Lin or Denis Andrienko.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021