Exploring Molecular Energy Landscapes by Coupling the DFTB Potential with a Tree-Based Stochastic Algorithm: Investigation of the Conformational Diversity of Phthalates

Valentin Milia, Nathalie Tarrat, Christophe Zanon, Juan Cortés, Mathias Rapacioli

Exploring molecular energy landscapes by coupling DFTB potential with a tree-based stochastic algorithm: Investigation of the conformational diversity of Phthalates

Valentin Milia,[†,‡] Nathalie Tarrat,[*,¶] Christophe Zanon,[†] Juan Cortés,[*,†] and Mathias Rapacioli[*,‡]

†*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France*

‡*Laboratoire de Chimie et Physique Quantiques LCPQ/FERMI, UMR 5626, Université de Toulouse (UPS) and CNRS, 118 Route de Narbonne, F-31062 Toulouse, France*

¶*CEMES, Université de Toulouse, CNRS, 29 Rue Jeanne Marvig, 31055 Toulouse, France*

E-mail: nathalie.tarrat@cemes.fr; juan.cortes@laas.fr; mathias.rapacioli@irsamc.ups-tlse.fr

Abstract

Exploring the global energy landscape of relatively large molecules at the quantum level is a challenging problem. In this work, we report the coupling of a non-redundant conformational space exploration method, namely the robotics-inspired Iterative Global exploration and LOcal Optimization (IGLOO) algorithm, with the quantum Density Functional Tight Binding (DFTB) potential. The application of this fast and efficient computational approach to three close-sized molecules of the phthalate family (DBP, BBP and DEHP) showed that they present different conformational landscapes. These differences have been rationalized making use of descriptors based on distances and dihedral angles. Coulomb interactions, steric hindrance and dispersive interactions have been found to drive geometric properties. A strong correlation has been evidenced between the two dihedral angles describing the side-chains orientation of the phthalate molecules. Our approach identifies low-energy minima without prior knowledge of the potential energy surface, paving the way for future investigations into transition paths and states.

Introduction

The theoretical prediction of physico-chemical properties of molecules such as chemical reactivity, ionisation energies, spectroscopy often require the knowledge of their low energy conformations. When the Born-Oppenheimer approximation can be applied, this involves the search for the most stable minima of the potential energy surface (PES) of the electronic ground state. In the case of flexible molecules (for instance biomolecules, polymers, chemicals, pollutants), the efficient exploration of their associated high-dimensional conformational space remains a challenging task, especially when a level of description of the potential close to an *ab initio* method is required. For this purpose, a cautious choice of the combination between an adequate exploration strategy (limiting the number of single point energy calculations to be performed, in particular over-sampling) and an appropriate level of description of the PES (compromise between the computational cost and requested accuracy) is mandatory.

A large majority of the PES exploration schemes[1,2] rely on either Monte Carlo[3] (MC) methods or Molecular Dynamics simulations[4] (MD). MC is a stochastic approach often performed within the Metropolis algorithm where random displacements are accepted as a function of a temperature, while MD propagates the nuclei positions by solving Newton's equations of motion. These methods are robust in exploring a conformational landscape and can provide a thermodynamic interpretation. They can also be combined with periodic local minimizations to locate the bottom of the PES wells. Various strategies have been implemented to improve the efficiency of MC- and MD-based methods. This includes for instance simulated annealing,[5] parallel tempering methods[6–8] or basin-hopping (BH) schemes,[9,10] used either in their standard[11–13] or improved versions.[14–16] A disadvantage of these schemes is that they do not keep track of the visited regions, which can lead to oversampling of certain areas of the PES to the detriment of exploration of others. In the case of free energy reaction path calculations, methods keeping a knowledge of the visited space (e.g. umbrella sampling or metadynamics) have been developed to increase the exploration efficiency.[17,18] These latter require however *a priori* knowledge of the reaction coordinates (collective variables), which prevents their use in a context of blind exploration of complex PES. In summary, there is still work to be done to develop efficient algorithms to discover potentially diverse energy basins, i.e. without prior knowledge of the system of interest and requiring little or no adaptation to a particular case study.

In recent years, methods inspired from robot motion planning algorithms have been proposed to efficiently explore the conformational space of molecular systems.[19–21] These methods construct data structures (trees or graphs) that encode the explored regions of the space, and avoid revisiting these regions. One of these algorithms is the Rapidly-exploring Random Trees (RRT),[22] which was subsequently extended to the exploration of energy landscapes aiming to find transition paths.[23–25] More recently, the RRT and BH algorithms have been combined to find energy minima on a PES.[26] The strategy applied in this work, called Iterative Global exploration and LOcal Optimization (IGLOO),[27] iterates RRT-based exploration and local minimization. The IGLOO algorithm will be detailed in the in Materials and Methods section.

Various levels of theory exist to compute the energy for the visited points of the PES. They range from high-level *ab initio* schemes with the wave function methods to lower-levels

such as force field approaches.[28] In between, Density Functional Theory[29] (DFT) is the most common method used to study systems with tens to hundreds of atoms. Unfortunately, the computational cost of DFT, although much lower than that of wave function methods, remains a bottleneck in the framework of exhaustive PES explorations, when millions of single point energy calculations are intended to be done. An alternative method, namely the Density-Functional based Tight-Binding approach (DFTB),[30–33] relying on several approximations of DFT, preserves the explicit quantum description of the electronic system while drastically reducing the computational cost thanks to the use of parameterized integrals and a minimal valence basis set. Its DFT ground usually makes it more transferable than force field models.

In the present work, we report the coupling of a non-redundant conformational space exploration approach, namely the robotics-inspired IGLOO method, with the quantum chemical DFTB potential. This enable efficient discovery of diverse energy basins while preserving a highly-accurate level of description of chemical systems. As an example of application, we considered three molecules of the phthalate family, whose energy landscapes present numerous degenerated basins.

Phthalates are commonly used in many consumer products such as PVC, coatings, adhesives, perfumes and cosmetics due to their plasticising properties.[34,35] For example, phthalates prevent nail varnish from chipping, make perfumes last longer, make tool handles stronger and more resistant, and increase the effectiveness of adhesives. Their global production is estimated at 3 million tonnes per year.[36] These molecules can be released into the environment during the production, use and disposal of products containing them, and these compounds can be found in water, air, soil and sediment. They have been associated with a variety of adverse effects on human health[37] depending on many factors, including dose, duration and route of exposure. In particular, they can act as endocrine disruptors by interfering with the natural hormones in the human body. Such effects were found in the premature development of breasts in Puerto Rican girls.[38] Some phthalates have been associated with reduced sperm quality in men, as well as birth defects in infants exposed in utero.[39] In this work, we focused on three representative phthalate molecules, namely the DEHP (di(2-ethylhexyl)) phthalate, which represents 50 % of the global phthalate production, the BBP (benzyl butyl) phthalate, and the DBP (dibutyl) phthalate which represent about one tenth of the DEHP production together. To our knowledge, no exhaustive exploration of their energy landscape has been reported in the literature.

The paper is organized as follows. In the next section, details about the IGLOO exploration scheme and the DFTB method are given together with a description of their coupling. In the results and discussion section, we first present the three investigated phthalate molecules and the geometric descriptors. The energy minima identified by the IGLOO-DFTB coupling are then described on the basis of their isomer spectra. Finally, the energy-structure relationships are investigated thanks to specifically developed analysis tools, giving insight in the intramolecular energy minimization mechanisms.

Materials and Methods

This section describes the implemented methods allowing the global exploration of the PES of molecular systems. First, the IGLOO exploration algorithm and the DFTB potential are briefly described. Then, we provide details about their combination in a computational framework.

Iterative Global exploration and LOcal Optimization

The Iterative Global exploration and LOcal Optimization (IGLOO) algorithm performs a global exploration of the conformational landscape of molecules to find the lowest energy representatives. IGLOO relies on an exploration strategy originally proposed to solve motion planning problems in robotics. More precisely, it applies a variant of the RRT algorithm,[22] adapted to the exploration of energy landscapes.[40,41] Similarly to other techniques that perform global optimization by iterating local searches, such as the Basin-Hopping algorithm,[9,10] the RRT-based exploration is coupled with a local optimization technique with the aim of descending into the energy basins. The IGLOO algorithm interleaves global explorations and local minimization stages in an iterative manner, also including a filtering stage to reduce the number of states considered in subsequent iterations. A more detailed description of these stages will be provided below, together with the implementation details. More in-depth explanations about IGLOO can also be found in a recent work,[27] which demonstrates the good performance of the algorithm compared to related methods for finding low-energy conformations of atomic or molecular systems. Note also that IGLOO was successfully applied to predict the structure of disaccharide molecules on metal surfaces.[42,43]

Density-Functional based Tight-Binding

The DFTB method is an approximated DFT scheme developed in the mid-90's[31,44] following the pioneering work of Foulkes and Haydock.[45] It is derived from a Taylor expansion of the Kohn-Sham effective potential energy with respect to the electronic density fluctuation where the molecular orbitals are expressed in a minimal valence basis set. At zero order, the algebra is equivalent to standard non-consistent semi-empirical Tight-Binding[46] and the formal derivation allows for the tabulation of the Kohn-Sham and overlap matrices elements in the atomic basis as diatomic terms from reference DFT calculations. The energy consists in the usual "band structure" terms and a short-range repulsive term.

This approach was further extended by Elstner et al.[32] to include second order terms in the Taylor expansion. At long distances, this correction accounts for the long-range electrostatic interactions between point charges and, at short distances, it also includes exchange-correlation contributions. Whereas the electronic problem can be solved with a single diagonalisation in the case of the zeroth-order DFTB, the introduction of second order contributions implies a self-consistent search for the electronic ground state density and energy. As the second order contribution is expressed depending on the atomic charges, the method is often referred to as Self-Consistent-Charge (SCC). The SCC extension allows DFTB to address problems for which the zeroth-order DFTB approach is not sufficient, in particular when atomic charges deviate from the neutral atoms reference and/or when

the Coulomb interaction between atomic charges has a decisive role in the determination of structural or energetic properties. More recently, DFTB has also been improved by including third order terms in the Taylor expansion, which introduces a charge dependence of the chemical hardness.[47,48] DFTB has been applied to compute various structural, energetic and thermodynamic properties as well as vibrational and electronic spectra, covering a wide range of systems like molecules, atomic or molecular clusters, extended materials or liquids.[33,49–57]

Implementation details

Our implementation of the coupled IGLOO-DFTB method is based on the interfacing of software developed in our laboratories: (*i*) the IGLOO algorithm implemented in the Molecular Motion Algorithms (MoMA) software suite (https://moma.laas.fr/) and (*ii*) the DFTB energy calculation implemented in the deMonNano code (http://demon-nano.ups-tlse.fr).

A schematic view of the coupling between IGLOO (MoMA) and DFTB (deMonNano) is provided in Fig. 1. MoMA is the master code, sending requests to and receiving data from deMonNano. The two programs communicate through a wrapper, allowing intercompatibility of software based on different programming languages. The protocol used for the communication is an INET socket[58] (a programming interface that enables applications to send and receive data over a network, either locally or via the Internet).

The interfaced software takes as input a set of parameters required to initialize the algorithms, as well as a chemical structure of a molecule. The main parameters of IGLOO are the number of (randomly sampled) initial states and the step size used for the first iteration of the algorithm, which aims to cover the conformational space roughly but globally. The step size is then self-adapted in subsequent iterations, shrinking as the exploration focuses in the low-energy basins. Other important parameters concern the stopping criteria. Several types of conditions are considered to determine the end of the iterative process performed by IGLOO. They are based on: (*i*) a maximum number of iterations; (*ii*) a limited computing time; (*iii*) estimated convergence, based on the evolution of the lowest energy value. These criteria are evaluated at the end of each iteration, and the first one to be satisfied stops the algorithm. In general, the parameters corresponding to the *i* and *ii* criteria are set to high values, so that the exploration iterates until estimated convergence is reached.

After the initialization, the algorithm iterates three successive stages (see[27] for deeper explanations on the method):

1. Exploration: At each iteration, IGLOO explores the conformational space using a stochastic process starting from a set of states. For the first iteration, these states are randomly sampled using a strategy inspired from the Poisson disk sampling process to ensure good dispersion. A variant of the RRT algorithm is then applied to explore reachable regions of the conformational space by growing random trees rooted at the initial states. New states are added to the tree if they are below an energy threshold, which is determined automatically and decreases with each iteration of the algorithm. These "single point" energy calculations are made by deMonNano, with MoMA providing the coordinates of a given conformation.

2. Local minimization: The explored conformations are minimized locally. Energy minimization is performed by deMonNano using a conjugated gradient technique. In order

to reduce computing time, several deMonNano executions can be performed in parallel, taking advantage of the independence of the calculations.

3. Filtering: This step enables dense areas to be cleaned up locally, with the aim to reduce the number of local minima from which the next iteration of the IGLOO algorithm is initialized.

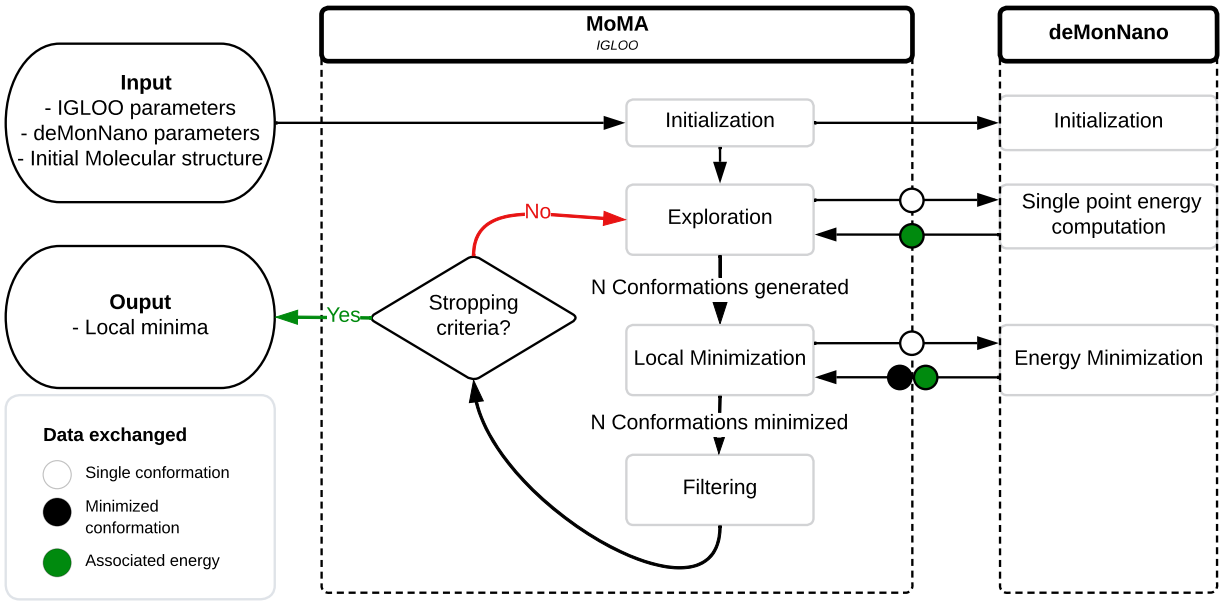


Figure 1: Schematic description of the IGLOO (MoMA) - DFTB (deMonNano) coupling.

In the present work, the degrees of freedom chosen for the IGLOO scheme are the phthalate dihedral (torsion) angles, used both to perform geometric displacements and to characterize similarities between different conformations. IGLOO was initialized with one hundred initial states for each molecule. This number was chosen to be large enough to ensure exhaustive exploration of the conformational space.

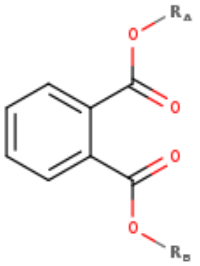
Energy computations were carried out using the third-order DFTB formalism (3ob parameters),[48,59] combined with an empirical dispersion correction.[60] A Fermi temperature of 100K was introduced to avoid convergence issues during the self-consistent scheme. Local minimizations were performed using a conjugated gradient technique. In order to validate the DFTB parameters and the dispersion correction, DFTB and DFT minima were compared on a structural and energetic basis (see Figures S5-S10 of the Supporting Information). The selected structures and the results of these comparisons are discussed below, in the section entitled "Structure-energy relationships". The output of the IGLOO-DFTB method is a set of files containing the Cartesian coordinates of each local minimum found.

Results and discussion

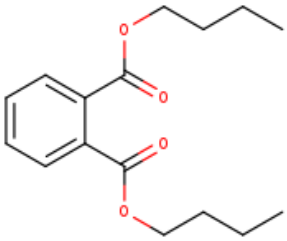
In this section, the three investigated phthalate molecules are introduced along with a presentation of the descriptors used in our analyses. Next, the isomer spectra of the molecules are studied, followed by an analysis of the conformational energy landscape, which shows a different behavior of the three molecules.

Geometric descriptors

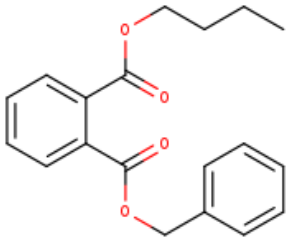
Phthalates are esters of phthalic acid composed of an aromatic benzene ring with two ester groups on ortho position. They differ between each other by their terminal groups. The generic form of phthalates is shown in Fig. 2-(a) together with the three phthalate molecules explored in the present study, namely BBP (Fig. 2-(b)), DBP (Fig. 2-(c)) and DEHP (Fig. 2-(d)). DBP and DEHP have two identical terminal groups, while BBP has two different ones, a 4-carbon alkyl and a phenyl-terminated chain.



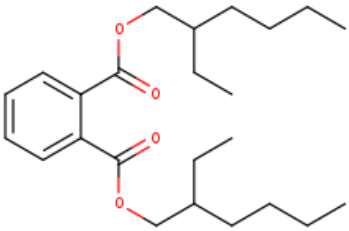
(a) Phthlate generic form



(b) DBP



(c) BBP

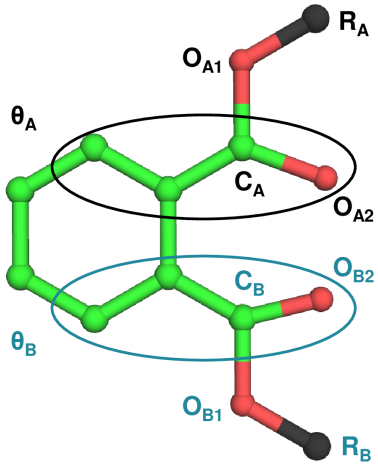


(d) DEHP

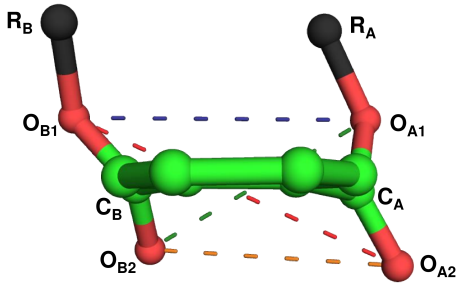
Figure 2: Generic form of phthalate (a) and the three molecules investigated in this work: dibutyl phthalate DBP (b), benzyl butyl phthalate BBP (c) and di(2-ethylhexyl) phthalate DEHP (d). Carbon skeleton in black and oxygen atoms in red. $R_{A,B}$ represent terminal groups of each side-chain.

The local minima obtained from the global exploration performed with the IGLOO-DFTB coupling were analysed on the basis of both energetic and structural descriptors. Descriptors characteristic of the relative orientation/organisation of the side-chains were defined from dihedral angles and interatomic distances (Fig. 3). Two dihedral angles were defined to describe the connection of the side-chains to the central aromatic ring (θ_{A} and θ_{B} in Fig. 3-(a)) as well as four oxygen-oxygen distances involving atoms belonging to two

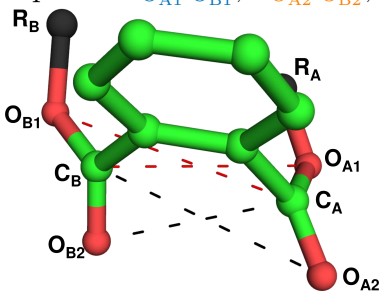
different chains: $d_{\mathrm{O_{A1}\text{-}O_{B1}}}$, $d_{\mathrm{O_{A2}\text{-}O_{B2}}}$, $d_{\mathrm{O_{A1}\text{-}O_{B2}}}$ and $d_{\mathrm{O_{A2}\text{-}O_{B1}}}$ (Fig. 3-(b)). The smallest value between these four distances defines the last descriptor, $dmin_{\mathrm{O\text{-}O}}$. Similarly, the $dmin_{\mathrm{C\text{-}O}}$ (Fig. 3-(c)) criterion was defined as the smallest distance between $d_{\mathrm{C_A\text{-}O_{B1}}}$, $d_{\mathrm{C_B\text{-}O_{A1}}}$, $d_{\mathrm{C_A\text{-}O_{B2}}}$ and $d_{\mathrm{C_B\text{-}O_{A2}}}$. In addition, we define (i) $d_{\mathrm{C\text{-}O_1}}$ as the smallest distance between $d_{\mathrm{C_A\text{-}O_{B1}}}$ and $d_{\mathrm{C_B\text{-}O_{A1}}}$ and (ii) $d_{\mathrm{C\text{-}O_2}}$ as the smallest distance between $d_{\mathrm{C_A\text{-}O_{B2}}}$ and $d_{\mathrm{C_B\text{-}O_{A2}}}$. Note that A and B are replaced by Bu (for butyl chain) and Be (for benzyl chain) for BBP because the chains are different and the atoms are not equivalent from one chain to the other.



(a) Dihedral descriptors: θ_{A} and θ_{B}



(b) O-O distance descriptors: $d_{\mathrm{O_{A1}\text{-}O_{B1}}}$, $d_{\mathrm{O_{A2}\text{-}O_{B2}}}$, $d_{\mathrm{O_{A1}\text{-}O_{B2}}}$ and $d_{\mathrm{O_{A2}\text{-}O_{B1}}}$



(c) C-O distance descriptors: $d_{\mathrm{C_A\text{-}O_{B1}}}$, $d_{\mathrm{C_B\text{-}O_{A1}}}$, $d_{\mathrm{C_A\text{-}O_{B2}}}$, $d_{\mathrm{C_B\text{-}O_{A2}}}$

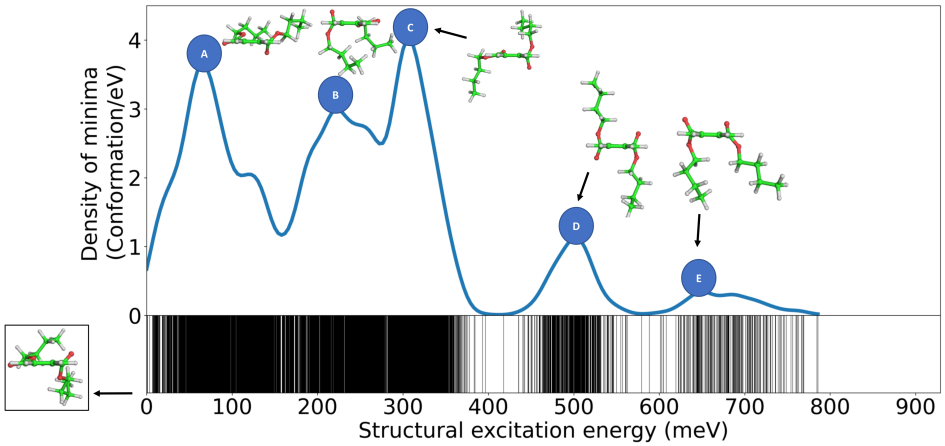
Figure 3: Structural descriptors: (a) dihedral angles and (b)-(c) interatomic distances. Carbon atoms are in green and oxygen atoms are in red. R balls in black represent the terminal group of each chain.

Structural excitation spectra

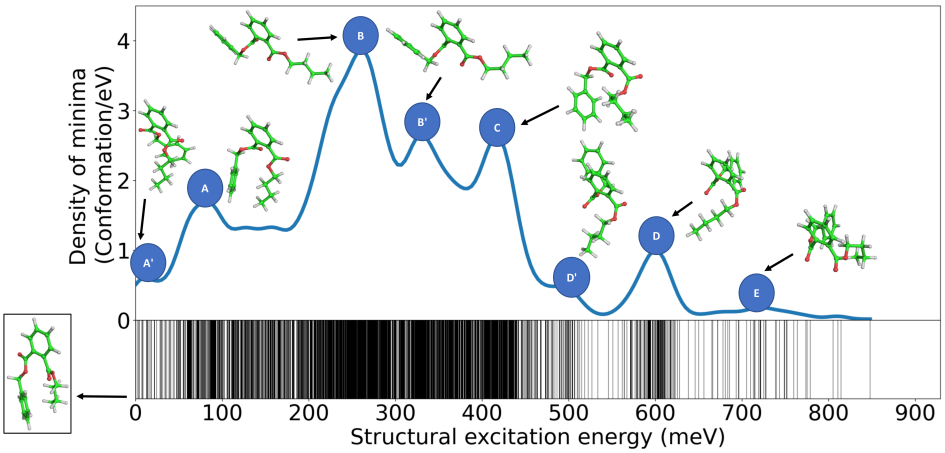
For each molecule, ten independent runs of the IGLOO-DFTB coupling were performed. Due of the stochastic nature of the exploration method, differences can be expected between the various runs. However, the algorithm showed very good reproducibility in the case of BBP and DBP, finding nearly the same low-energy minima at each run. In the case of DEHP, due to the higher dimensionality and the multiplicity of possible intramolecular interactions, several new minima were produced at each additional run. Consequently, the algorithm was applied ten more times (twenty in total). No new minimum was discovered during the last few runs, which is reassuring in terms of exploration convergence. The data analyzed below are the concatenated results of all runs of the algorithm for each molecule.

The structural excitation energy spectra obtained for the three investigated molecules are presented in Fig. 4, the zero energy reference being that of the global minimum. In the lower panels, each local minimum identified on the PES is represented by a bar. In the upper panels, these bar spectra are convoluted with a gaussian kernel in order to provide an estimation of the isomers density as a function of the energy. In the explorations, the highest energy isomers are located at 786 meV (DBP), 848 meV (BBP) and 929 meV (DEHP) above their respective global minimum. One should keep in mind that the present scheme aims at exploring the low energy regions of the PES, and the exploration of the highest energy regions is expected to be less exhaustive. As a result, the calculated isomers density at high energy is expected to be lower than the exact one. First, the general aspect of the isomer density distributions differ for the three molecules. These spectra can be characterized from the differences regarding their alternation of energy ranges exhibiting high/low isomer densities. For the DBP molecules (Fig. 4-(a)), the densest region extends from the global minimum energy to around 385 meV above it, with three peaks at 67 (peak **A**), 221 (peak **B**), and 308 meV (peak **C**). Another high-density peak is observed at 501 meV (peak **D**) surrounded by energy regions where almost no isomer was found. A final peak is identifiable at 642 meV (peak **E**), made up of few high-energy isomers. For the BBP molecule (Fig. 4-(b)), the isomer density is relatively low above the global minimum up to above 50 meV. In the 50-450 meV energy range in which the majority of isomers is found, three main high-density peaks are observed at around 81 meV, 260 meV, 418 meV (peaks **A**,**B**,**C**) and their variant (resulting from structures presenting small geometrical variations with respect to those dominating the main peak) noted by **'** at 20 meV and 328 meV (peaks **A'** and **B'**). At higher energies, the density is low except around 601 meV and 727 meV (peaks **D** and **E**) and the variant at 504 meV (peak **D'**) where other peaks are observed. For the DEHP molecule, the general aspect of the isomer density distribution (Fig. 4-(c)) is drastically different from the two previous ones. A unique and extended high density region is observed between 50 and 550 meV above the global minimum energy, the largest isomer density being located around 240 meV (peak **A**). A minor shoulder is also observed at 373 meV (peak **B**).

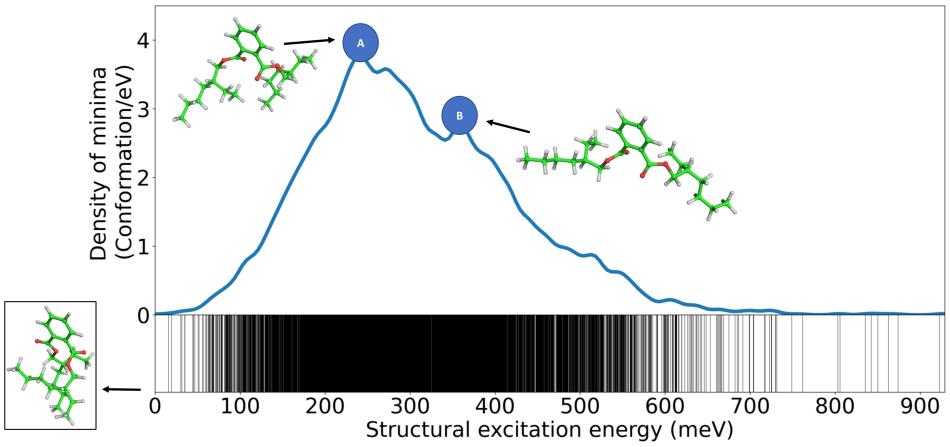
Fig. 4 also shows the lowest-energy structure and isomers located at the maximum of the density peaks for each molecule, pointing out the conformational variability. Nevertheless, one should be aware that they may not be representative of all the isomers making up the corresponding peak, since various structural patterns can contribute to a given peak. Deriving general trends would require a deeper analysis of the interplay between characteristic structural patterns and energies that will be addressed in the following sections.



(a) DBP



(b) BBP



(c) DEHP

Figure 4: Structural excitation energy spectra (bar and estimated density). For each molecule, main peaks are illustrated by their characteristic structures (multiple geometry could coexist in a peak). Global minima are depicted in the insets.

Structure-energy relationships

Distance based analysis

Although the three molecules are part of the same family, they exhibit different behaviours from an energetic point of view due to their terminal groups. Indeed, as the two alkyl chains are small in the DBP molecule, the interaction between the oxygen atoms is expected to drive the molecular energetics. The longer and ramified alkyl side-chains in DEHP would give rise to steric hindrance and to multiple possibilities for stabilisation through dispersive interactions. The interaction between the alkyl and aryl chains in the BBP molecule can generate structures stabilized by Coulomb interaction between the negatively charged aromatic carbon atoms and the positively charged hydrogen atoms of the butyl chain. In order to investigate relationships between the isomers energies and their structures, we first focus on the distances between oxygen-oxygen and carbon-oxygen atoms belonging to different side-chains.

On the pie charts of Fig. 5, the isomers are classified in four families depending on which of the $d_{\mathrm{O_{A1}\text{-}O_{B1}}}$, $d_{\mathrm{O_{A2}\text{-}O_{B2}}}$, $d_{\mathrm{O_{A1}\text{-}O_{B2}}}$ or $d_{\mathrm{O_{A2}\text{-}O_{B1}}}$ distances is the smallest one ($dmin_{\mathrm{O\text{-}O}}$). It appears that the molecules with small side-chains (DBP and BBP) exhibit similar distributions. The $d_{\mathrm{O_{A2}\text{-}O_{B2}}}$ is rarely the smallest one, i.e. the smallest distance always involves at least one of the side-chains connected oxygens $\mathrm{O_{A1}}$ or $\mathrm{O_{B1}}$, and the isomer population is equally shared in three groups by the attribution of $dmin_{\mathrm{O\text{-}O}}$ to either $d_{\mathrm{O_{A1}\text{-}O_{B1}}}$, $d_{\mathrm{O_{A1}\text{-}O_{B2}}}$ or $d_{\mathrm{O_{A2}\text{-}O_{B1}}}$. The pie chart of the DEHP molecule is different, with four almost similar quartiles. In this case, the smallest distance is attributed to $d_{\mathrm{O_{A2}\text{-}O_{B2}}}$ in one case out of five, this increase of occurrence being probably due to steric hindrance that prevents the oxygen atoms $\mathrm{O_{A1}}$ and $\mathrm{O_{B1}}$ from approaching each other.

On the pie charts of Fig. 6, the isomers are classified in two families depending on which of the $d_{\mathrm{C\text{-}O_1}}$ or $d_{\mathrm{C\text{-}O_2}}$ distances is the smallest one ($dmin_{\mathrm{C\text{-}O}}$). In the case of BBP, $d_{\mathrm{C\text{-}O_1}}$ and $d_{\mathrm{C\text{-}O_2}}$ are both split in two subfamilies because $\mathrm{R_A}$ and $\mathrm{R_B}$ are different. For the three molecules, the pie charts are dominated by the family $dmin_{\mathrm{C\text{-}O}} = d_{\mathrm{C\text{-}O_1}}$. This can be interpreted from the fact that O_1 is less negatively charged than O_2, reducing coulomb repulsion with the slightly negatively charged COO function. DBP appears to have a higher contribution of $d_{\mathrm{C\text{-}O_1}}$ than the other two molecules.

The scatter plots in Figs. 5 and 6 allow to correlate oxygen-oxygen and carbon-oxygen distances with energy distributions. It appears that for DBP and BBP, several distinct high density regions can be identified regarding the energy correlation with either the values of $dmin_{\mathrm{O\text{-}O}}$ and $dmin_{\mathrm{C\text{-}O}}$ or the density of minima per oxygen-oxygen subgroup. In the case of DBP, the five different regions appear clearly, and they can be easily identified by clustering method, as can be done for instance with a k-means method (this analysis is reported in Supporting Information). These five regions correspond to different dominant O-O and C-O interactions. As expected due to the presence of identical side-chains, a good superposition of the $dmin_{\mathrm{O\text{-}O}}$ cross-interaction curves is observed (red and green curves Fig. 5), which reassures us about the quality of the exploration. We remind that the lowest energy region of the structural excitation spectrum (Figs. 4 and 7-(a)) corresponds to the peak **A** (from

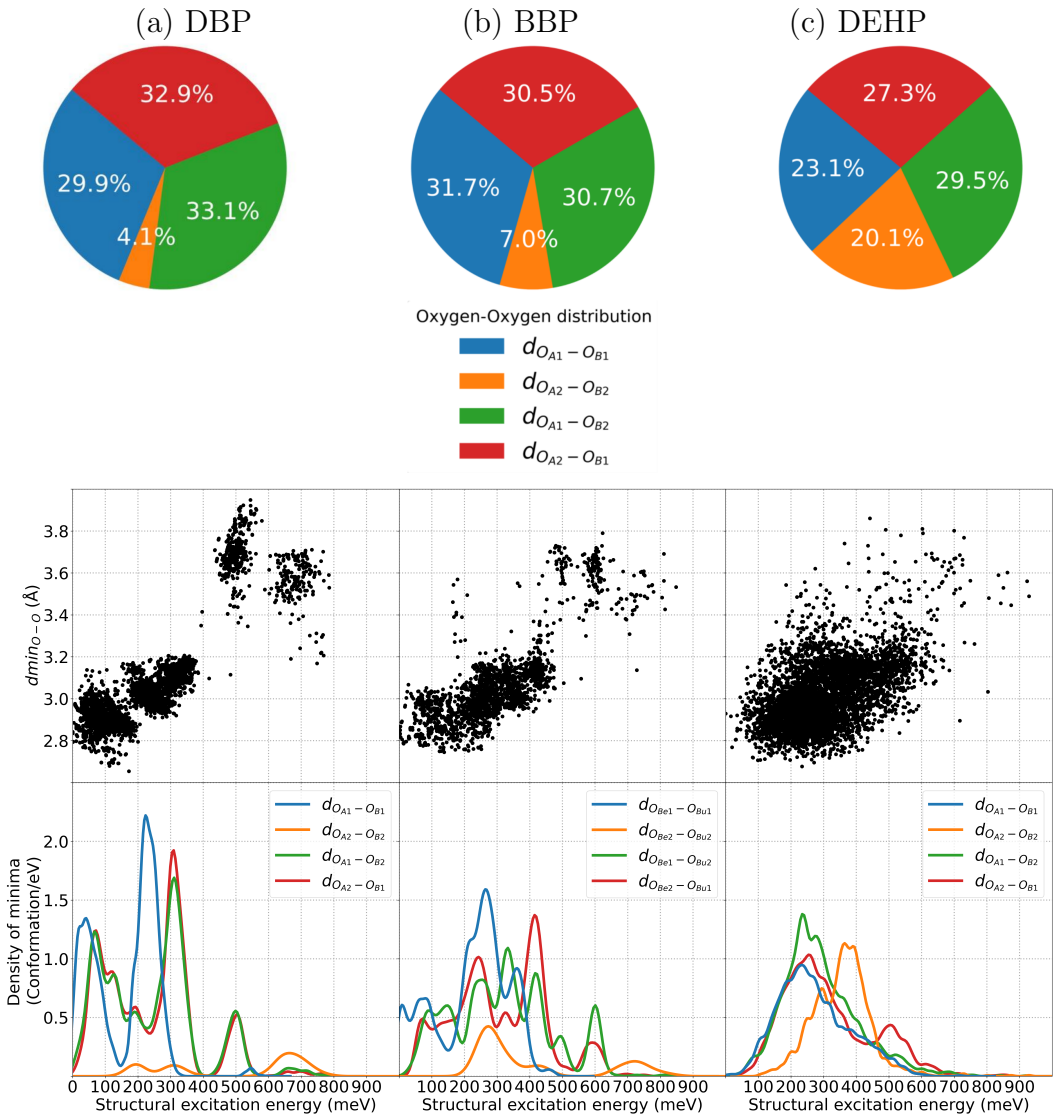


Figure 5: Global distribution of $dmin_{\text{O-O}}$ for all minima represented as a pie chart. The scatter plots report the $dmin_{\text{O-O}}$ values for each conformation and the curves represent the relative density of the conformations, both depicted as a function of the structural excitation energy. Isomers are classified according to the nature of $dmin_{\text{O-O}}$ (blue for $dmin_{\text{O-O}} = d_{\text{O}_{\text{A1}}\text{-O}_{\text{B1}}}$; orange for $dmin_{\text{O-O}} = d_{\text{O}_{\text{A2}}\text{-O}_{\text{B2}}}$; green for $dmin_{\text{O-O}} = d_{\text{O}_{\text{A1}}\text{-O}_{\text{B2}}}$ and red for $dmin_{\text{O-O}} = d_{\text{O}_{\text{A2}}\text{-O}_{\text{B1}}}$).

the most stable structure up to $\sim$ 150 meV). In these structures, the smallest distances between atoms from the COO functions always involve a positively charged carbon atom from one chain and a negatively charged oxygen atom O_1 from the other chain (Fig. 7-(a)), consequently, $dmin_{\text{C-O}}$ systematically involve an O_1 atom ($d_{\text{C}-\text{O}_1}$). This is consistent with the fact that the COO group is globally negatively charged (about -0.25 e) and that the charge carried by O_1 atom is about -0.35 e whereas that of the O_2 atom is about -0.56 e. For isomers belonging to peak **A**, the alkyl chains remains close even if they are from either side

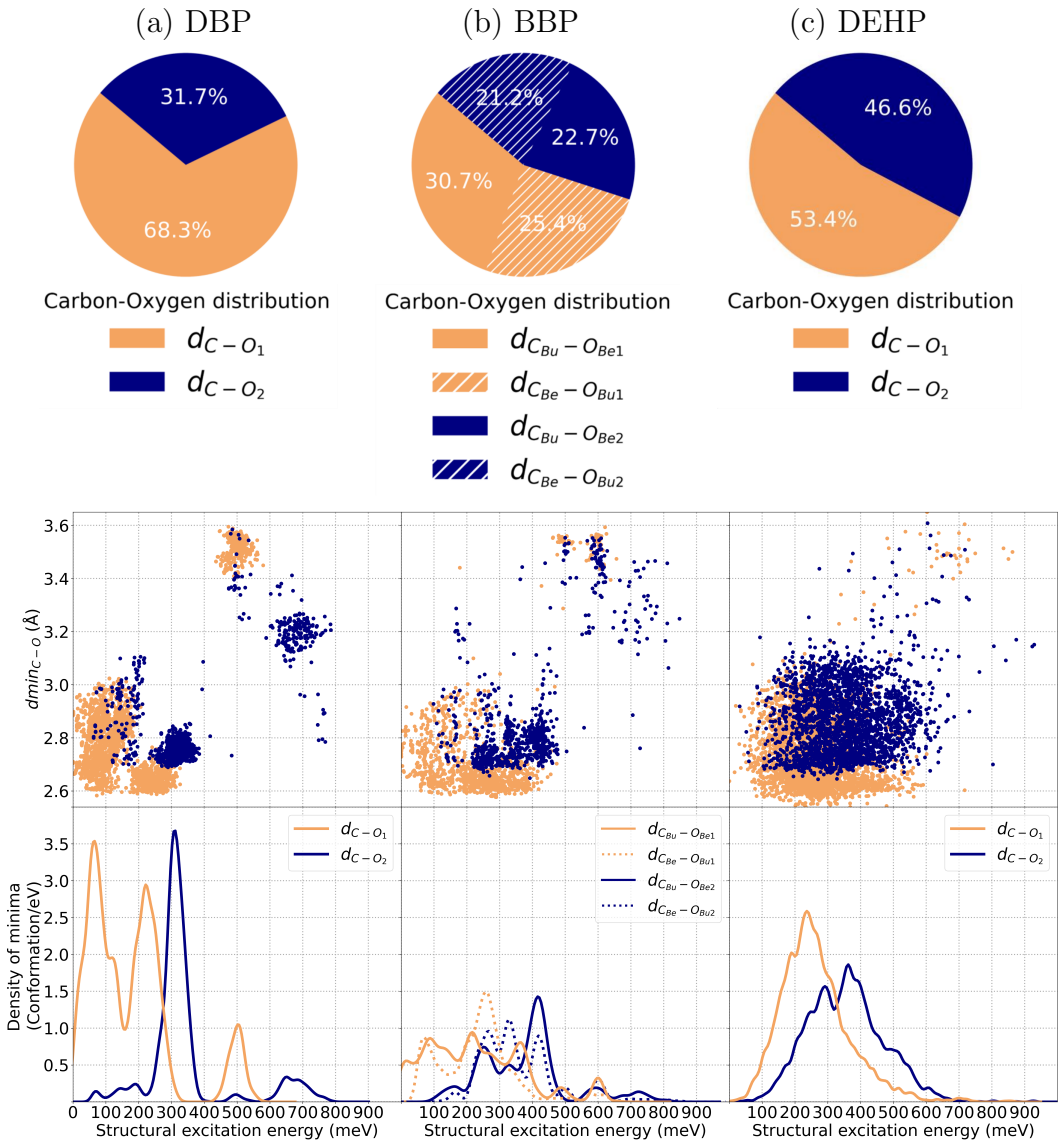
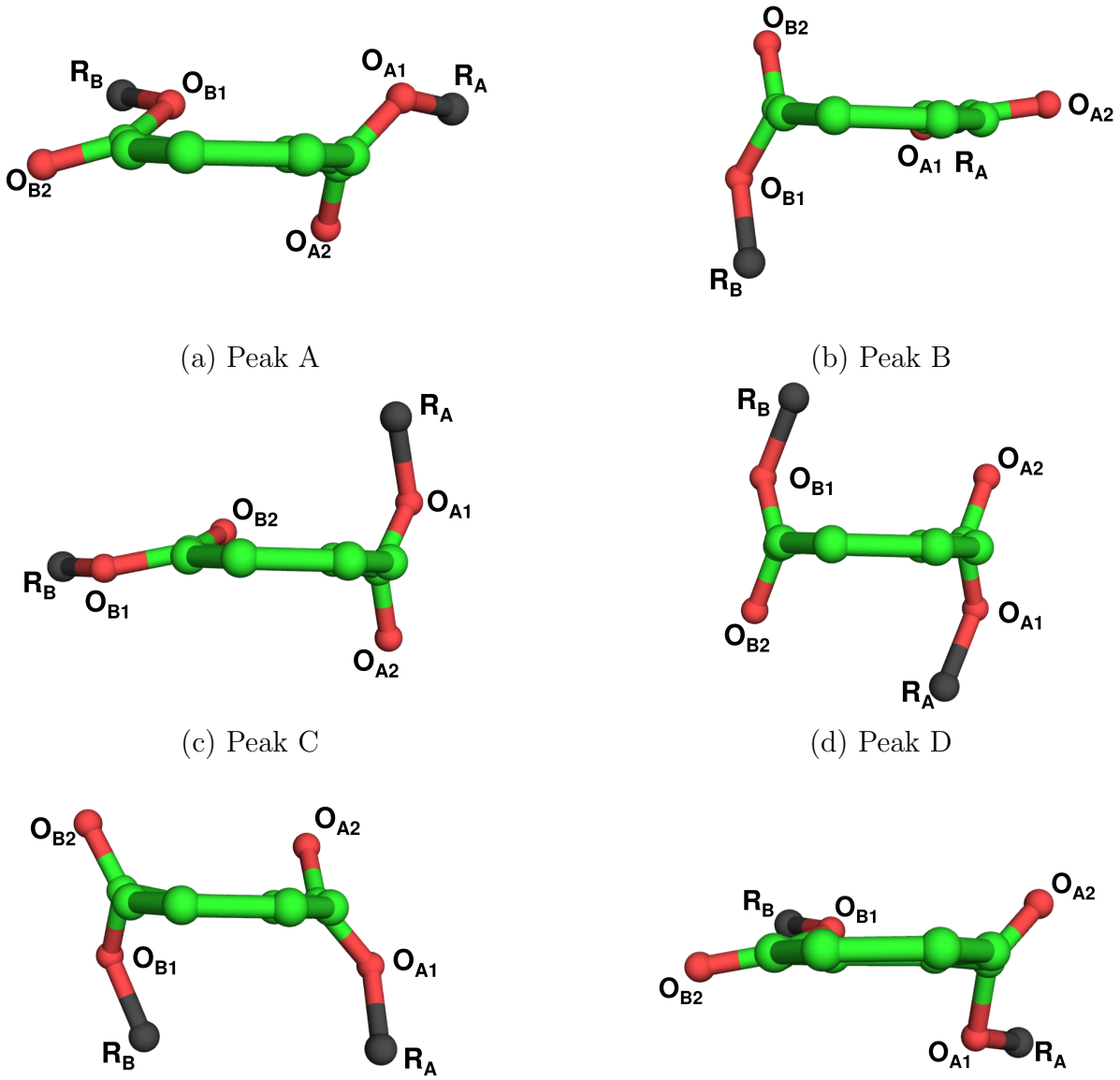


Figure 6: Global distribution of $dmin_{\text{C-O}}$ for all minima represented as a pie chart. The scatter plots report the $dmin_{\text{C-O}}$ values for each conformation and the curves represent the relative density of the conformations, both depicted as a function of the structural excitation energy. Isomers are classified according to the nature of $dmin_{\text{C-O}}$ (orange for $dmin_{\text{C-O}} = d_{\text{C-O}_1}$; blue for $dmin_{\text{C-O}} = d_{\text{C-O}_1}$).

of the central phenyl group. Moreover, structures containing two O_1 atoms pointing towards the carbon atom of the other chain COO group (highlighted by $dmin_{\text{O-O}} = d_{\text{O}_{\text{A}1}-\text{O}_{\text{B}1}}$) are particularly stable and gathered in the low energy region of the peak **A** (i.e. below ~ 50 meV, see Fig. 5-(a)). The second and third peaks (**B** and **C**) are mostly composed of structures where the two planes containing the COO groups are perpendicular to each others (Fig. 7-(b/c)). One COO group is parallel to the central phenyl and its O_1 (in peak **B**) or its O_2 (peak **C**) atom is involved in $dmin_{\text{C-O}}$. Again, the charge differences between O_1 and O_2 could explain the energy ranking between the two peaks. In addition, the induced relative



(a) Peak A

(b) Peak B

(c) Peak C

(d) Peak D

(e) Peak E

(f) Most stable DBP structure

Figure 7: Illustration of the ester groups relative orientation for characteristics structures of the DBP excitation spectrum main peaks (Fig. 4-(a)).

orientations of the side-chains in peak **C** result in less favourable interactions between the two alkyl chains (Fig. 7-(c)). In the last two peaks (**D** and **E**), the two planes containing the COO groups are perpendicular to the central phenyl, resulting in larger values for $dmin_{\text{C-O}}$ and $dmin_{\text{O-O}}$ than reported for the other peaks (Figs. 7-(d/e), 5-(a) and 6-(a)). These peaks differ by $dmin_{\text{O-O}} = d_{\text{O}_{\text{A1}}\text{-O}_{\text{B2}}}$ or $dmin_{\text{O-O}} = d_{\text{O}_{\text{A2}}\text{-O}_{\text{B1}}}$ in peak **D** or $dmin_{\text{O-O}} = d_{\text{O}_{\text{A2}}\text{-O}_{\text{B2}}}$ in peak **E**. This means that the side-chains are pointing in opposite (resp. similar) directions in peak **D** (resp. peak **E**). Although stabilizing interactions between the alkyl chains are almost absent in peak **D** structures whereas they are present in peak **E**, the steric hindrance

between these chains results in shortening the distances between O_{A2} and O_{B2} (as can be seen from $dmin_{\text{O-O}} = d_{\text{O}_{A2}\text{-O}_{B2}}$) and therefore increasing the coulomb repulsion between these oxygen atoms.

The structural excitation spectrum of the BBP molecule is trickier to interpret. The structures between the global minimum and 100 meV give birth to two peaks **A** and **A'** for which the $dmin_{\text{C-O}}$ distance always involves an O_1 atom, which minimizes the Coulomb repulsion (note that the charges of the oxygen atoms are similar for the three molecules). The COO functions are parallel to each other for peak **A'** structures (with $dmin_{\text{C-O}}$ involving the C of the butyl chain noted C_{Bu} and O_1 of the benzyl chain noted O_{Be1}) whereas peak **A** structures contain a mix of parallel and perpendicular structures (with a mix of $dmin_{\text{C-O}} = d_{\text{C}_{Bu}\text{-O}_{Be1}}$ and $dmin_{\text{C-O}} = d_{\text{C}_{Be}\text{-O}_{Bu1}}$). Bringing the less negatively charged oxygen atoms O_1 closer together, as in peak **A'**, maximizes chain interactions and minimizes oxygen-oxygen Coulomb repulsion. This explains the location of peak **A'** at lower energy than peak **A**. Note that, in addition to the peak **A'** features, the global minimum structure maximizes the interactions between the hydrogen atoms of its alkyl chain with the phenyl group of the other chain (Fig. 4-(b)). The structures present in peaks **B** and **B'** contain many various contributions, as can be seen from the $dmin_{\text{C-O}}$ and $dmin_{\text{O-O}}$ distribution plots (Fig. 5-(b) and Fig. 6-(b)) and it is therefore challenging to derive simple general trends. We note, however, that the smallest $dmin_{\text{C-O}}$ are obtained for $dmin_{\text{C-O}} = d_{\text{C-O}_1}$ and that $dmin_{\text{O-O}} = d_{\text{O}_{A2}\text{-O}_{B2}}$ is minority. Peak **C**, on the other hand, is similar to peak **C** in DBP, with $dmin_{\text{C-O}}$ involving an O_2 atom and a perpendicular orientation between the two COO functions with one of the latter contained in the plane of the central phenyl. $dmin_{\text{O-O}}$ consists solely of cross interactions with a slight preference for $dmin_{\text{O-O}} = d_{\text{O}_{Be2}\text{-O}_{Bu1}}$. Similar patterns are observed for the structures belonging to peaks **D** and **D'** and those of the peak **D** of DBP, i.e. with two COO planes perpendicular to the central phenyl and side-chains pointing in opposite directions. Structures of the peak **D'** are slightly more stable thanks to the stacking of the two phenyl groups, which is associated to $dmin_{\text{O-O}} = d_{\text{O}_{Be1}\text{-O}_{Bu2}}$. Finally, representative structures of peak **E** of BBP are very similar to those of peak **E** of DBP. The COO groups planes are perpendicular to the central phenyl and parallel to each other, and the two carbonyl oxygen are on the same side. In this group, the contribution of $dmin_{\text{O-O}} = d_{\text{O}_{A1}\text{-O}_{A2}}$ and $dmin_{\text{O-O}} = d_{\text{O}_{A2}\text{-O}_{A1}}$ is almost absent in BBP although a minor contribution was observed in the DBP case. We remind, however, that the exploration could be incomplete at such high energies because the IGLOO scheme is mostly designed to explore the low energy regions of the PES.

The structural excitation energy curve for DEHP (Fig. 4 bottom) is less structured than the two previous ones. The $dmin_{\text{C-O}}$ of the most stable structure involves an O_1 atom, with the COO functions perpendicular to each other, and one of them being in the plane of the central ring. In addition, the side-chains are close to each other and on either side of the phenyl plane. From $\approx$ 75 meV, the chains can be placed on the same side of the phenyl plane. Another structure appears from $\approx$ 87 meV with $dmin_{\text{C-O}} = d_{\text{C-O}_2}$. Above this energy, the curve is a single broad distribution up to 600 meV above the global minimum. At least two substructures, noted **A** at $\approx$ 240 meV and **B** at $\approx$ 380 meV, can be identified. The $dmin_{\text{C-O}}$ and $dmin_{\text{O-O}}$ distributions of DEHP in the in Fig. 5-(c) and Fig. 6-(c) are large and overlapping. The large number of possible interactions between the long alkyl chains results in a continuum of isomers over the wide energy range for each previously discussed struc-

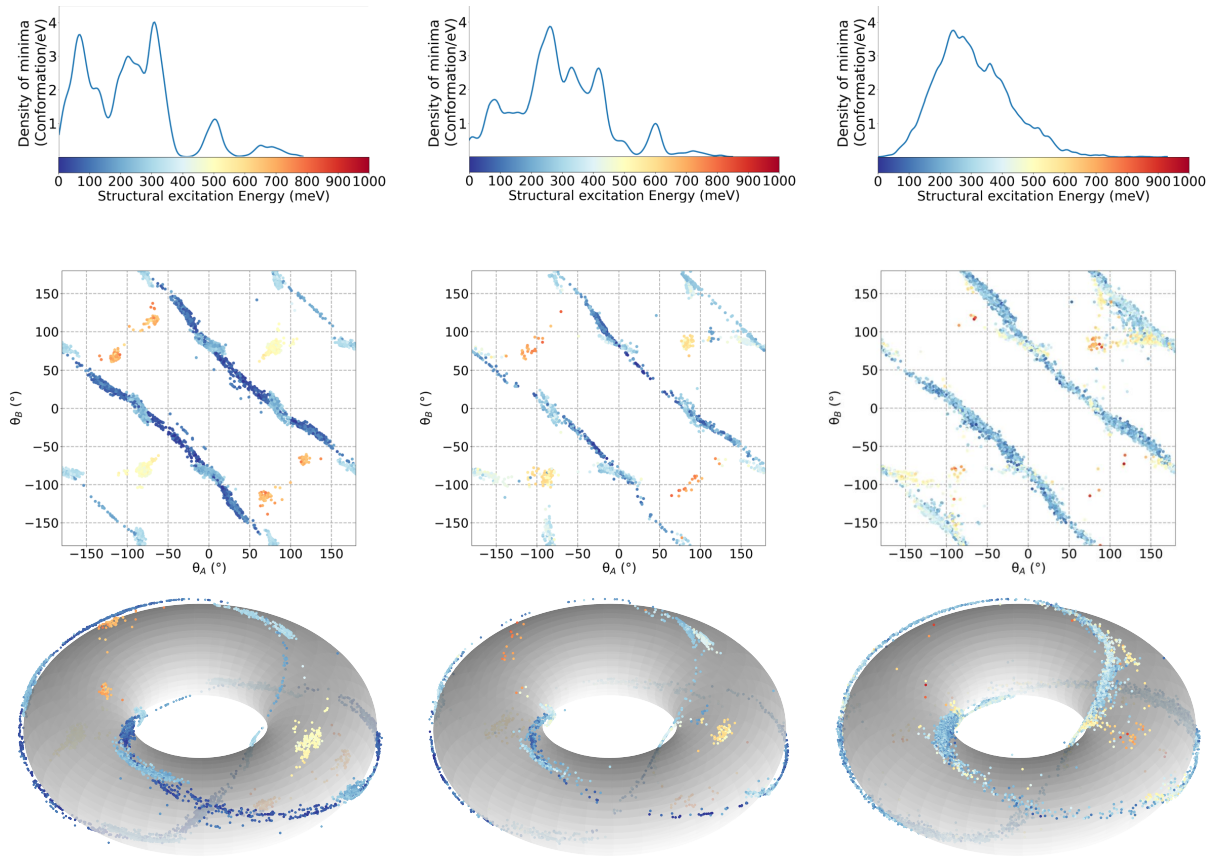
tural feature. While it was, to a large extend, easy to attribute a peak to a given structural characteristic for the two other molecules, the coexistence of structurally distinct isomers at a given energy hinders a detailed analysis of the DEHP structural excitation energy spectrum. Nevertheless, one can mention that peak **A** is dominated by structures exhibiting $dmin_{\text{C-O}} = d_{\text{C-O}_1}$ and an equal repartition between $dmin_{\text{O-O}} = d_{\text{O}_{\text{A1}}\text{-O}_{\text{B1}}}$, $dmin_{\text{O-O}} = d_{\text{O}_{\text{A1}}\text{-O}_{\text{B2}}}$ and $dmin_{\text{O-O}} = d_{\text{O}_{\text{A2}}\text{-O}_{\text{B1}}}$. Note that the slight difference between the $d_{\text{O}_{\text{A1}}\text{-O}_{\text{B2}}}$ and $d_{\text{O}_{\text{A2}}\text{-O}_{\text{B1}}}$ curves may be due to the difficulty of achieving complete exploration of high-dimensional conformational space. The presence of peak **B** is due to structures for which $dmin_{\text{C-O}} = d_{\text{C-O}_2}$ and $dmin_{\text{O-O}} = d_{\text{O}_{\text{A2}}\text{-O}_{\text{B2}}}$. This peak is higher in energy because it combines strong Coulomb repulsion (between the two most negatively charged O_2 atoms and between an O_2 atom and the COO group) and loss of dispersive stabilisation due to a large distance between the chains.

We performed additional DFT local minimizations (see section "Structural comparison between DFT and DFTB potentials" in Supporting Information for computational details) on the representative structures of the main peaks observed in Figure 4. Superimpositions of the corresponding DFTB and DFT structures are depicted in Figures S5, S6 and S7 for DBP, BBP and DEHP, respectively. Overall, the structural differences observed are minor except in the case of peak D of DBP, peak B of BBP and peak B of DEHP for which a slight modification of the orientation of the side-chains is observed. This is certainly due the existence of a multitude of very close minima in these zones of the DFTB and DFT potential energy surfaces, reflected by the continuum observed in the regions of these peaks in the DFTB structural excitation energy spectra (Figure 4). However, the structures concerned retain their main structural characteristics discussed above. A comparison of the DFTB and DFT energy ranking is given in Figures S8, S9 and S10 for DBP, BBP and DEHP, respectively. The hierarchy of minima is the same between the two methods, with only one inversion observed, between peaks B and C of the DBP. These results fully support our strategy of globally exploring the potential energy surfaces of phthalate molecules at the DFTB level.

Dihedral angle based analysis

The distribution of the conformations resulting from the IGLOO-DFTB exploration can also be visualized on a two-dimensional (2D) projection with respect to the dihedral angles θ_{A} and θ_{B} (see Fig. 3 for their definition). The projections[61] for the three molecules are presented in Fig. 8, where each conformation corresponds to a point colored as a function of its structural excitation energy. For clarity, the spectra on Fig. 4 are presented again on top of the color-bars in Fig. 8. The figure shows two types of projections for these 2D angular distributions. The first one, in the center of the figure, is a classical representation on a Euclidean plane. In the second one, at the bottom of the figure, the conformations are projected on the surface of a two-dimension torus. This type of representation is less usual but better suited to the visualization of angular values due to their periodicity.

An initial structural analysis of the molecules reveals symmetries in conformational space that should be found in the dihedral angle analysis. Changing the signs of both θ_{A} and θ_{B} is equivalent to performing a symmetric projection of the atom coordinates with respect to



(a) DBP (b) BBP (c) DEHP

Figure 8: Distribution of the conformations resulting from the IGLOO-DFTB exploration. Top: Structural excitation spectra (Fig. 4) with color-bar. Middle: Two-dimensional (2D) projection with respect to dihedral angles θ_{A} and θ_{B}. Bottom: Projection on the surface of a two-dimension torus. In these plots, each conformation corresponds to a point colored as a function of its structural excitation energy (upper panel color-bar).

a plane passing through the benzene ring. As this would lead to the same isomer, the 2D Euclidean projections in the central row of Fig. 8 should be symmetric with respect to the $y = x$ axis (ascending diagonal). In addition, when the two terminal groups are identical, exchanging the values of θ_{A} and θ_{B} also leads to the same structure and, as a consequence, symmetry with respect to the $y = -x$ axis (downward diagonal) should also appear. The difference between the two side-chains of BBP induces a loss of this second type of symmetry, particularly visible between the lower left and upper right bands in the corresponding 2D Euclidean plot. All the previously mentioned expected symmetries are recovered in Figure 8, which is reassuring regarding the quality of the global exploration. The figure shows similar angular distribution for the three molecules. All the low-energy conformations (colored in blue) are grouped within parallel bands in the Euclidean projection, or rings on the surface of the torus. Note that each ring is divided into two bands (one long and one short) on the Euclidean plane because periodicity is not taken into account. Note also that some isomers of higher energies (colored in yellow, orange and red) are located between these bands/rings.

The energetic grouping of DBP and BBP isomers shown in Fig. 4, can also be observed through well defined colored regions on the 2D projections of Fig. 8-(a-b). In the plot corresponding to DEHP (Fig. 8-(c)), energy basins are less clearly identifiable, reflecting, as previously mentioned, the complex competition between several weak stabilizing interactions.

The 2D projections clearly show that θ_{A} and θ_{B} are strongly correlated. Considering only the points in the blue bands/rings and using linear regression, we obtained $\theta_{\mathrm{B}} + \theta_{\mathrm{A}} = c$ with $|c|$ in the [85-92°] range and correlation coefficients larger than 0.97, the two bands/rings differing by the sign of c. This correlation between the two dihedral angles is illustrated in the animation provided as Supporting Information. The reason for this correlation is probably due to the fact that low-energy conformations tend to maximize the oxygen-oxygen distances between side-chains and, therefore, when one chain rotates, the other does so accordingly. In our previous analysis of the structural excitation spectrum, the high densities were mostly interpreted in the light of $dmin_{O-O}$ and $dmin_{C-O}$ values, the later being strongly linked to the θ_{A} and θ_{B} values.

Finally, we can imagine transitions between conformations projected onto the two low-energy bands/rings, passing through the high-energy yellow-red regions. However, finding these transitions would require a variant of the IGLOO algorithm, focused on sampling transition paths rather than low-energy basins.

Conclusion

In the present work, we have reported the coupling of a non-redundant conformational space exploration algorithm named IGLOO with the quantum chemical DFTB potential. This implementation enables rapid and efficient exploration of a molecule's energy landscape in vacuum, with a quantum potential.

This approach has been applied to the exploration of the conformational potential energy surface of three molecules representative of the phthalate family: butyl benzyl phthalate (BBP), dibutyl phthalate (DBP) and di-(2-ethylhexyl) phthalate (DEHP). This choice was motivated by their high impact on human health. The results show that BBP, DPB and DEHP, despite belonging to the same family and being close in size, present different conformational landscape properties. The general aspect of the structural excitation energy spectra shows different isomer density distributions for the three molecules. The DBP spectrum has well defined peaks while the DEHP one exhibits a continuum of close-energy states. The BBP spectrum is at the crossroads between these two previous behaviours. These differences have been rationalized making use of descriptors based on distances and dihedral angles.

DBP lower-energy structures are mostly governed by oxygen-oxygen coulomb interactions. In the case of BBP, original structures, where the positively charged hydrogen atoms of the butyl side-chain point toward the negatively charged aromatic carbon atoms, allow to maximize coulomb interactions stabilization. Finally, DEHP long and ramified side-chains induce steric hindrance and dispersive interactions, these latter being at the origin of competitions between plenty of isomers. These interactions drive the geometric properties of the investigated phthalate molecules leading either to peaks (DBP and BBP) or to a broad feature (DEHP) in characteristic O-O and C-O distances distribution plots and to a strong correlation between the two dihedral angles describing the side-chains orientation for the

three molecules.

One should note that the phthalate molecules have been studied here in the gas phase and that further research could provide a protocol for finding conformations that could exist under more realistic conditions. The effects of the environment could be incorporated through QM-MM explicit[62] or implicit[63] solvent scheme. The IGLOO-DFTB coupling implemented in this work, allowing the identification of low energy minima of a molecule with no a priori knowledge of its potential energy surface, could be extended in the future to the blind search of the minimum energy path between selected structures.

Data and Software Availability

Data presented in this article have been deposited on ZENODO:
https://zenodo.org/records/10040725.
As mentioned in the "Implementation details" section, the combined IGLOO-DFTB approach was implemented on the basis of the Molecular Motion Algorithms (MoMA) software suite (https://moma.laas.fr/) and the deMonNano code (http://demon-nano.ups-tlse.fr). Software binaries and user guidelines are available at:
https://gitlab.laas.fr/moma/binaries/igloo-dftb-coupling.

Acknowledgement

This work was supported by the French National Center for Scientific Research (CNRS) under the QUARTET 80|Prime grant. This work was granted access to the HPC resources of CALMIP supercomputing center under the allocation p19055 and p0059.

Supporting Information Available

Supporting information includes:

- Illustration of characteristic structures
  - Illustration of the side-chains relative orientation for characteristic structures of the main peaks of the DBP, BBP and DEHP structural excitation spectra.
  - DBP clustering by k-means method.
- Structural comparison between DFT and DFTB potentials
  - Computational details.
  - Comparison between characteristic structures of the main peaks of DBP, BBP and DEHP structural excitation spectra using DFT and DFTB potentials.
  - Comparison of DFT and DFTB energies.

References

(1) Schlegel, H. B. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. COMPUT. CHEM.* **2003**, *24*, 1514–1527.

(2) Paz Borbón, L. O. *Computational Studies of Transition Metal Nanoalloys*; SPRINGER BERLIN HEIDELBERG: Berlin, Heidelberg, 2011; pp 15–31.

(3) Hammond, B. L.; Lester, W. A.; Reynolds, P. J. *Monte Carlo Methods in Ab Initio Quantum Chemistry*; WORLD SCIENTIFIC, 1994.

(4) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *CURR. OPIN. CELL BIOL.* **2002**, *12*, 190–196.

(5) Selli, D.; Fazio, G.; Di Valentin, C. Modelling realistic TiO2 nanospheres: A benchmark study of SCC-DFTB against hybrid DFT. *J. CHEM. PHYS.* **2017**, *147*, 164701.

(6) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *PHYS. REV. LETT.* **1986**, *57*, 2607–2609.

(7) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *CHEM. PHYS. LETT.* **1999**, *314*, 141–151.

(8) Sugita, Y.; Okamoto, Y. Replica-Exchange Multicanonical Algorithm and Multicanonical Replica-Exchange Method for Simulating Systems with Rough Energy Landscape. *CHEM. PHYS. LETT.* **2000**, *329*, 261–270.

(9) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *PROC. NATL. ACAD. SCI.* **1987**, *84*, 6611–6615.

(10) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. PHYS. CHEM. A* **1997**, *101*, 5111–5116.

(11) Aktürk, A.; Sebetci, A. BH-DFTB/DFT calculations for iron clusters. *AIP ADV.* **2016**, *6*, 055103.

(12) Choi, T. H.; Liang, R.; Maupin, C. M.; Voth, G. A. Application of the SCC-DFTB Method to Hydroxide Water Clusters and Aqueous Hydroxide Solutions. *J. PHYS. CHEM.* **2013**, *117*, 5165–5179.

(13) Choi, T. H. Simulation of the (H2O)8 cluster with the SCC-DFTB electronic structure method. *CHEM. PHYS. LETT.* **2012**, *543*, 45–49.

(14) Yen, T. W.; Lai, S. K. Use of Density Functional Theory Method to Calculate Structures of Neutral Carbon Clusters Cn ($3 \leq n \leq 24$) and Study their Variability of Structural Forms. *J. CHEM. PHYS.* **2015**, *142*, 084313.

(15) Yen, T.-W.; Lim, T.-L.; Yoon, T.-L.; Lai, S. K. Studying the Varied Shapes of Gold Clusters by an Elegant Optimization Algorithm that Hybridizes the Density Functional Tight-Binding Theory and the Density Functional Theory. *COMPUT. PHYS. COMMUN.* **2017**, *220*, 143–149.

(16) Chen, X.; Zhao, Y.-F.; Wang, L.-S.; Li, J. Recent progresses of global minimum searches of nanoclusters with a constrained Basin-Hopping algorithm in the TGMin program. *COMPUT. THEOR. CHEM.* **2017**, *1107*, 57–65.

(17) Yang, M.; Yang, L.; Gao, Y.; Hu, H. Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface. *J. CHEM. PHYS.* **2014**, *141*, 044108.

(18) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WILEY INTERDISCIP. REV. COMPUT. MOL. SCI.* **2011**, *1*, 826–843.

(19) Al-Bluwi, I.; Siméon, T.; Cortés, J. Motion Planning Algorithms for Molecular Simulations: A Survey. *COMPUT. SCI. REV.* **2012**, *6*, 125–143.

(20) Gipson, B.; Hsu, D.; Kavraki, L.; Latombe, J.-C. Computational models of protein kinematics and dynamics: Beyond simulation. *ANNU. REV. ANAL. CHEM.* **2012**, *5*, 273–91.

(21) Shehu, A.; Plaku, E. A Survey of Computational Treatments of Biomolecules by Robotics-Inspired Methods Modeling Equilibrium Structure and Dynamic. *J. ARTIF. INTELL. RES.* **2016**, *57*, 509–572.

(22) LaValle, S. M.; Kuffner, J. J. Randomized Kinodynamic Planning. *INT. J. ROBOT. RES.* **2001**, *20(5)*, 378–400.

(23) Jaillet, L.; Corcho, F. J.; Pérez, J.-J.; Cortés, J. Randomized tree construction algorithm to explore energy landscapes. *J. COMPUT. CHEM.* **2011**, *32*, 3464–3474.

(24) Devaurs, D.; Molloy, K.; Vaisset, M.; Shehu, A.; Simeon, T.; Cortés, J. Characterizing Energy Landscapes of Peptides using a Combination of Stochastic Algorithms. *IEEE TRANS. NANOBIOSCIENCE* **2015**, *14*, 545–552.

(25) Estaña, A.; Molloy, K.; Vaisset, M.; Sibille, N.; Siméon, T.; Bernadó, P.; Cortés, J. Hybrid parallelization of a multi-tree path search algorithm: Application to highly-flexible biomolecules. *PARALLEL COMPUT.* **2018**, *77*, 84–100.

(26) Roth, C.-A.; Dreyfus, T.; Robert, C. H.; Cazals, F. Hybridizing rapidly exploring random trees and basin hopping yields an improved exploration of energy landscapes. *J. COMPUT. CHEM.* **2016**, *37*, 739–752.

(27) Margerit, W.; Charpentier, A.; Maugis-Rabusseau, C.; Schön, J. C.; Tarrat, N.; Cortés, J. IGLOO: An Iterative Global Exploration and Local Optimization Algorithm to Find Diverse Low-Energy Conformations of Flexible Molecules. *ALGORITHMS* **2023**, *16*.

(28) Harrison, J. A.; Schall, J. D.; Maskey, S.; Mikulski, P. T.; Knippenberg, M. T.; Morrow, B. H. Review of force fields and intermolecular potentials used in atomistic computational materials research. *APPL. PHYS. REV.* **2018**, *5*, 031104.

(29) Parr, R. G. Density Functional Theory. *ANNU. REV. PHYS. CHEM.* **1983**, *34*, 631–656.

(30) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *PHYS. REV. B* **1995**, *51*, 12947–12957.

(31) Seifert, G.; Porezag, D.; Frauenheim, T. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. *INT. J. QUANTUM CHEM.* **1996**, *58*, 185–192.

(32) 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.

(33) Spiegelman, F.; Tarrat, N.; Cuny, J.; Dontot, L.; Posenitskiy, E.; Martí, C.; Simon, A.; Rapacioli, M. Density-functional tight-binding: basic concepts and applications to molecules and clusters. *ADV PHYS-X* **2020**, *5*, 1710252.

(34) Staples, C. A.; Peterson, D. R.; Parkerton, T. F.; Adams, W. J. The environmental fate of phthalate esters: A literature review. *CHEMOSPHERE* **1997**, *35*, 667–749.

(35) Almeras, C.; Cancan, Y.; Gereec, V.; Millet, M. *Projet PERSAN Les Phtalates*; 2010.

(36) *Les Phtalates : Sources d'exposition et impregnation humaine*; Réseau Environnement Santé: Paris, France, 2011.

(37) Hauser, R.; Calafat, A. M. Phthalates and human health. *OCCUP. ENVIRON. MED.* **2005**, *62*, 806–818.

(38) Colón, I.; Caro, D.; Bourdony, C. J.; Rosario, O. Identification of phthalate esters in the serum of young Puerto Rican girls with premature breast development. *ENVIRON. HEALTH PERSPECT.* **2000**, *108*, 895–900.

(39) Kavlock, R. et al. NTP Center for the Evaluation of Risks to Human Reproduction: phthalates expert panel report on the reproductive and developmental toxicity of di(2-ethylhexyl) phthalate. *REPROD. TOXICOL.* **2002**, *16*, 529–653.

(40) Cortés, J.; Siméon, T.; Ruiz de Angulo, V.; Guieysse, D.; Remaud-Siméon, M.; Tran, V. A path planning approach for computing large-amplitude motions of flexible molecules. *BIOINFORMATICS* **2005**, *21*, i116–i125.

(41) Jaillet, L.; Corcho, F. J.; Pérez, J.-J.; Cortés, J. Randomized tree construction algorithm to explore energy landscapes. *J. COMPUT. CHEM.* **2011**, *32*, 3464–3474.

(42) Abb, S.; Tarrat, N.; Cortés, J.; Andriyevsky, B.; Harnau, L.; Schön, J. C.; Rauschenbach, S.; Kern, K. Carbohydrate Self-Assembly at Surfaces: STM Imaging of Sucrose Conformation and Ordering on Cu(100). *ANGEW. CHEM. INT. ED.* **2019**, *58*, 8336–8340.

(43) Abb, S.; Tarrat, N.; Cortés, J.; Andriyevsky, B.; Harnau, L.; Schön, J. C.; Rauschenbach, S.; Kern, K. Polymorphism in carbohydrate self-assembly at surfaces: STM imaging and theoretical modelling of trehalose on Cu(100). *R. SOC. CHEM. ADV.* **2019**, *9*, 35813–35819.

(44) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *PHYS. REV. B* **1995**, *51*, 12947–12957.

(45) Foulkes, W. M. C.; Haydock, R. Tight-binding models and density-functional theory. *PHYS. REV. B* **1989**, *39*, 12520–12536.

(46) Goringe, C. M.; Bowler, D. R.; Hernández, E. Tight-binding modelling of materials. *REP. PROG. PHYS.* **1997**, *60*, 1447.

(47) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction. *J. PHYS. CHEM.* **2007**, *111*, 10861–10873.

(48) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. CHEM. THEORY COMPUT.* **2011**, *7*, 931–948.

(49) Elstner, M. The SCC-DFTB method and its application to biological systems. *THEOR. CHEM. ACC.* **2006**, *116*, 316–325.

(50) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *PHYS. STATUS SOLIDI* **2000**, *217*, 357–376.

(51) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Carlo, A. D.; Suhai, S. Atomistic simulations of complex materials: ground-state and excited-state properties. *J. PHYS. CONDENS. MATTER* **2002**, *14*, 3015.

(52) Elstner, M.; Seifert, G. Density functional tight binding. *PHILOS. TRANS. ROYAL SOC. A* **2014**, *372*, 20120483.

(53) Gaus, M.; Cui, Q.; Elstner, M. Density functional tight binding: application to organic and biological molecules. *WIRES COMPUT. MOL. SCI.* **2014**, *4*, 49–61.

(54) Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. Validation of the density-functional based tight-binding approximation method for the calculation of reaction energies and other data. *J. CHEM. PHYS.* **2005**, *122*, 114110.

(55) Niehaus, T. A.; Suhai, S.; Della Sala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. Tight-binding approach to time-dependent density-functional response theory. *PHYS. REV. B* **2001**, *63*, 085108.

(56) Oliveira, A.; Seifert, G.; Heine, T.; duarte, H. Density-functional based tight-binding: an approximate DFT method. *J. BRAZ. CHEM. SOC.* **2009**, *20*, 1193–1205.

(57) Rapacioli, M.; Simon, A.; Dontot, L.; Spiegelman, F. Extensions of DFTB to investigate molecular complexes and clusters. *PHYS. STATUS SOLIDI* **2012**, *249*, 245–258.

(58) Xue, M.; Zhu, C. The Socket Programming and Software Design for Communication Based on Client/Server. 2009 Pacific-Asia Conference on Circuits, Communications and Systems. 2009; pp 775–777.

(59) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. CHEM. THEORY COMPUT.* **2013**, *9*, 338–354.

(60) Zhechkov, L.; Heine, T.; Patchkovskii, S.; Seifert, G.; Duarte, H. A. An Efficient a Posteriori Treatment for Dispersion Interaction in Density-Functional-Based Tight Binding. *J. CHEM. THEORY COMPUT.* **2005**, *1*, 841–847.

(61) Hill, C. *Learning Scientific Programming with Python*; Cambridge University Press, 2016.

(62) Yusef Buey, M.; Mineva, T.; Rapacioli, M. Coupling density functional based tight binding with class 1 force fields in a hybrid QM/MM scheme. *THEOR. CHEM. ACC.* **2022**, *141*, 16.

(63) Hou, G.; Zhu, X.; Cui, Q. An Implicit Solvent Model for SCC-DFTB with Charge-Dependent Radii. *J. CHEM. THEORY COMPUT.* **2010**, *6*, 2303–2314.

Supporting Information

"Exploring molecular energy landscapes by coupling DFTB potential with a tree-based stochastic algorithm: Investigation of the conformational diversity of Phthalates"

Valentin Milia,[†,‡] Nathalie Tarrat,[*,¶] Christophe Zanon,[†] Juan Cortés,[*,†] and Mathias Rapacioli[*,‡]

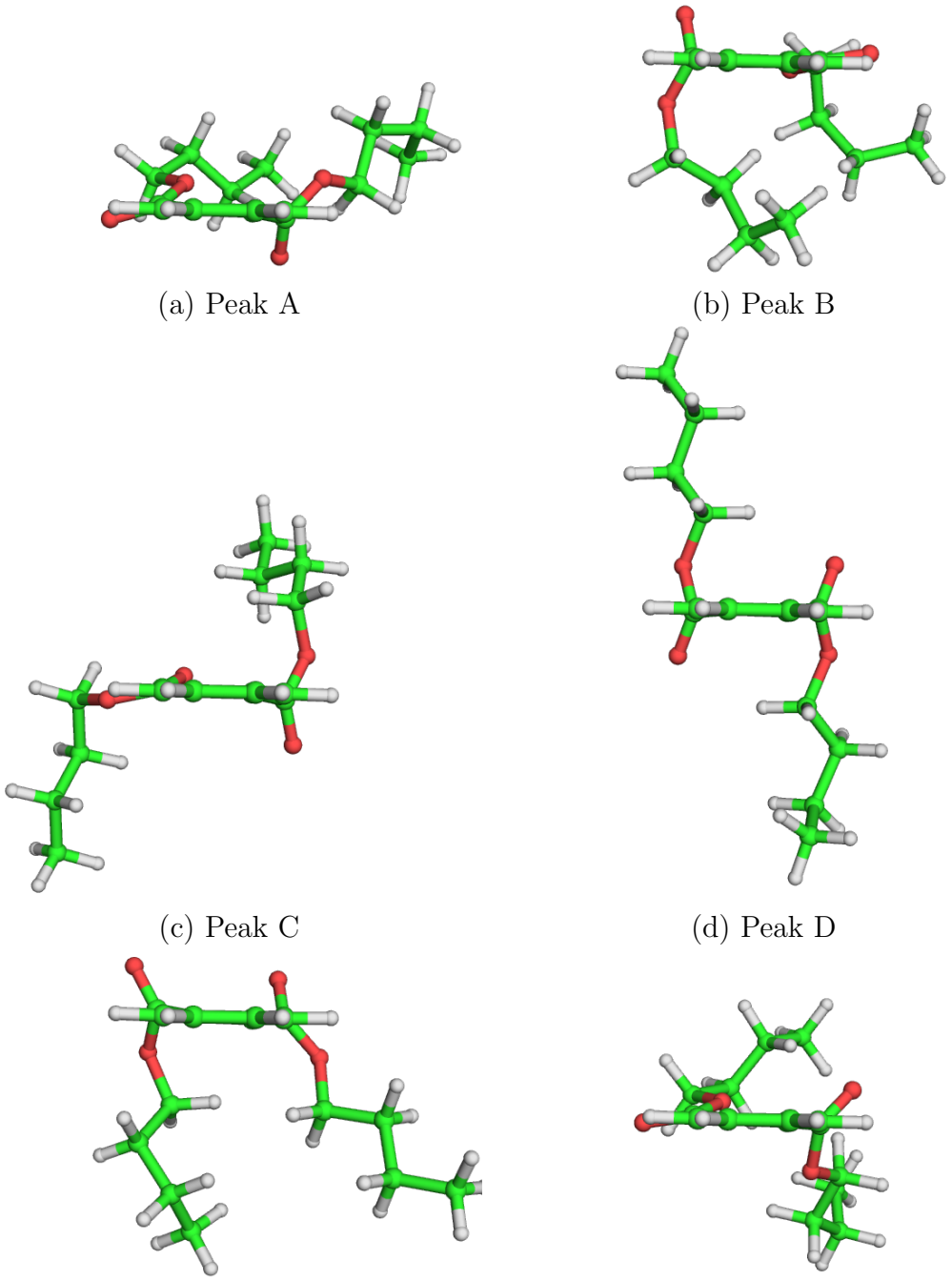
†*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France*

‡*Laboratoire de Chimie et Physique Quantiques LCPQ/FERMI, UMR 5626, Université de Toulouse (UPS) and CNRS, 118 Route de Narbonne, F-31062 Toulouse, France*

¶*CEMES, Université de Toulouse, CNRS, 29 Rue Jeanne Marvig, 31055 Toulouse, France*

E-mail: nathalie.tarrat@cemes.fr; juan.cortes@laas.fr; mathias.rapacioli@irsamc.ups-tlse.fr

Illustration of characteristic structures



(a) Peak A

(b) Peak B

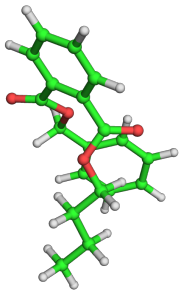
(c) Peak C

(d) Peak D

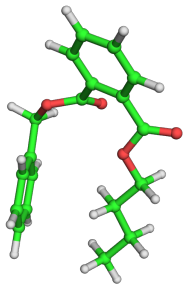
(e) Peak E

(f) Most stable DBP structure

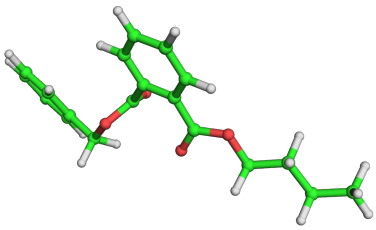
Figure S1: Illustration of the side-chains relative orientation for characteristic structures of the main peaks of the DBP structural excitation spectrum.



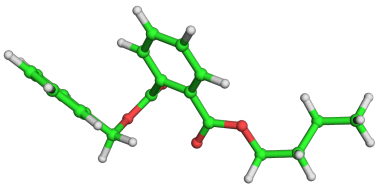
(a') Peak A'



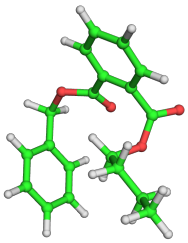
(a) Peak A



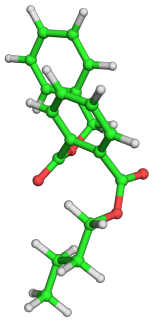
(b) Peak B



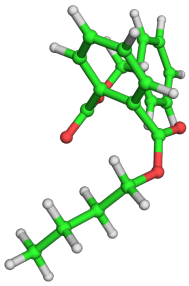
(b') Peak B'



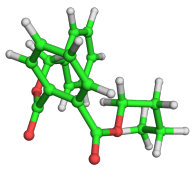
(c) Peak C



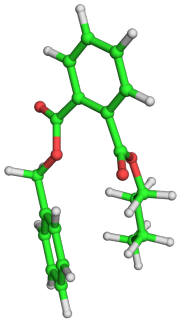
(d') Peak D'



(d) Peak D

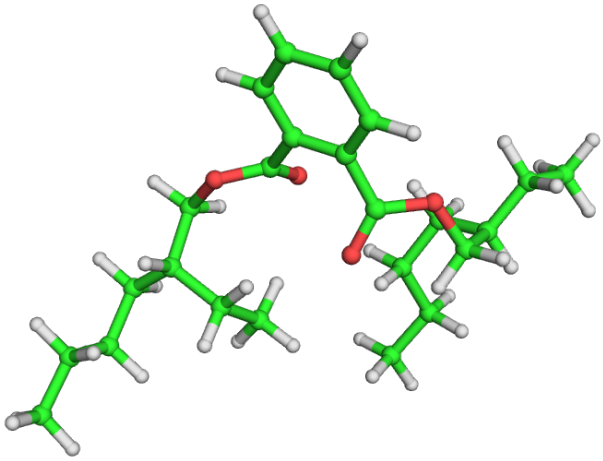


(e) Peak E

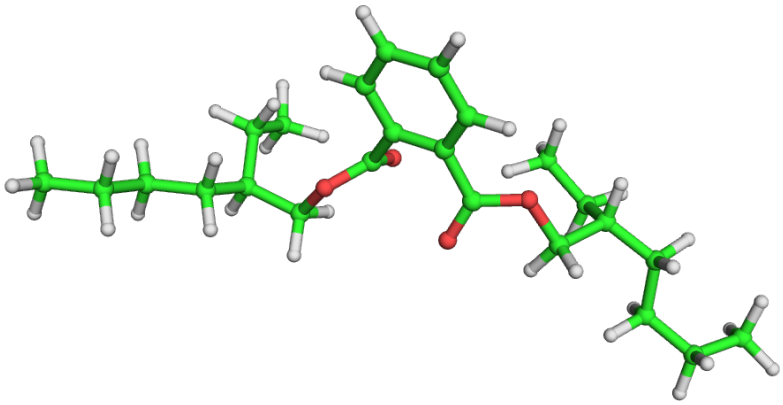


(f) Most stable BBP structure

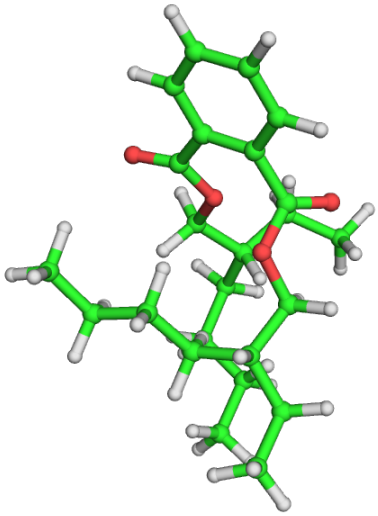
Figure S2: Illustration of the side-chains relative orientation for characteristic structures of the main peaks of the BBP structural excitation spectrum.



(a) Peak A



(b) Peak B



(c) Most stable DEHP structure

Figure S3: Illustration of the side-chains relative orientation for characteristic structures of the main peaks of the DEHP structural excitation spectrum.

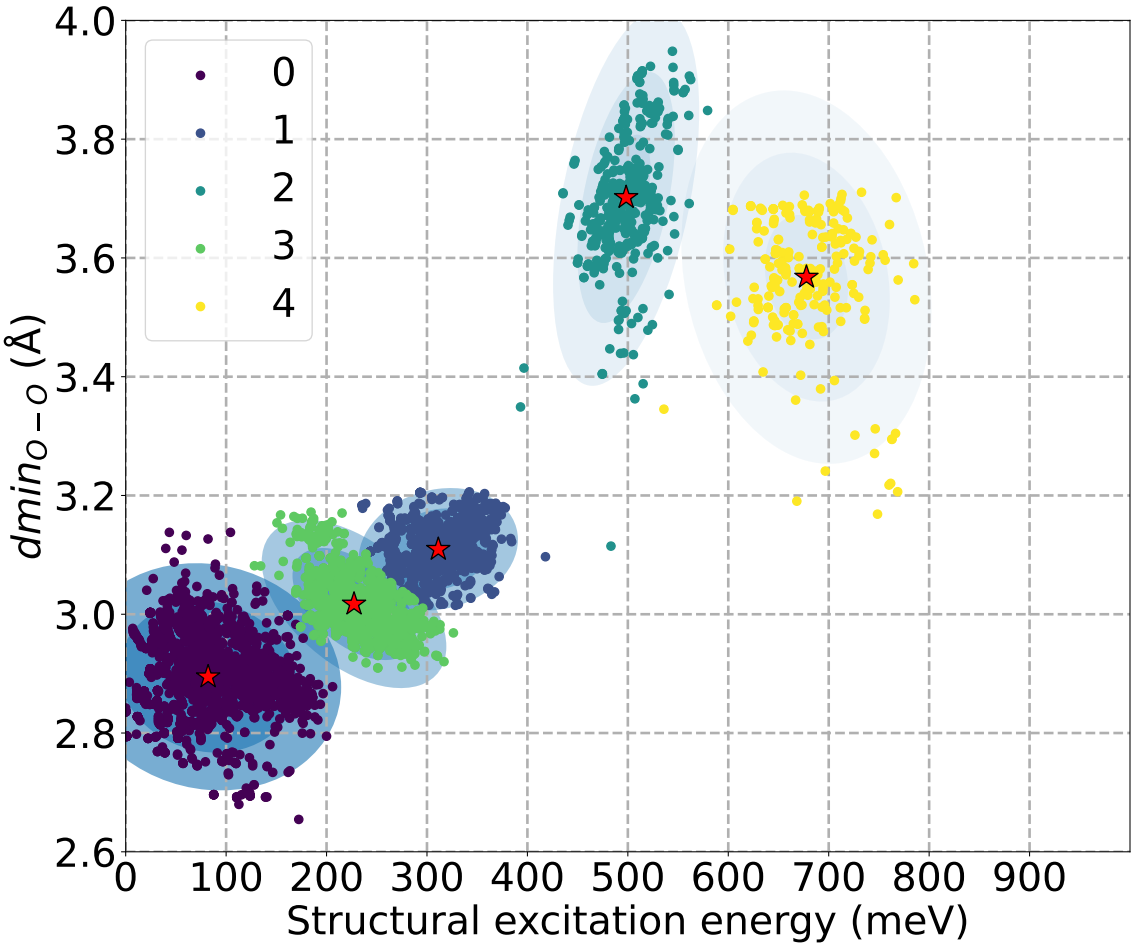
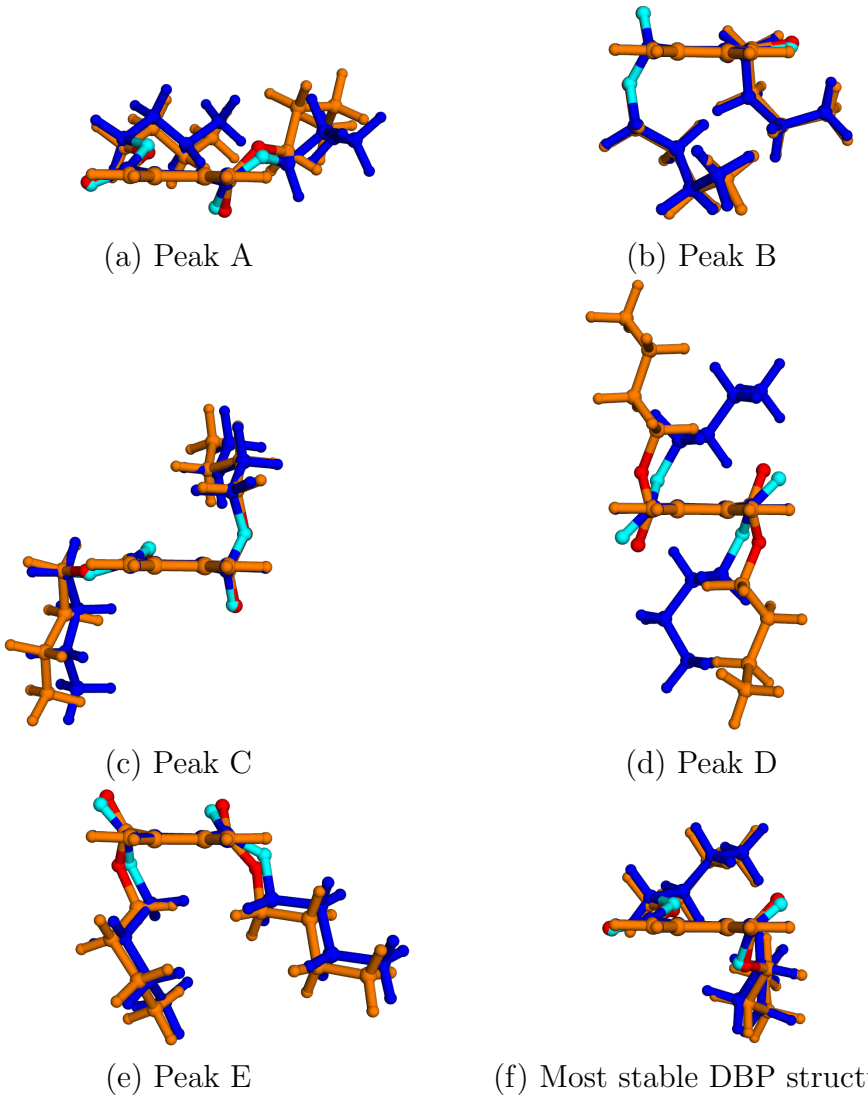


Figure S4: Clustering by k-means method of the point cloud corresponding the plot of the $dmin_{\text{O-O}}$ distance of the DBP molecule as a function of its structural excitation energy. Red stars and blue ovals represent the center and the covariance of each cluster, respectively.

Structural comparison between DFT and DFTB potentials

Density functional theory (DFT) computational details: DFT calculations were performed using the Gaussian 16 set of programs.[1] The high-nonlocal and hybrid meta exchange-correlation M06–2X functional[2] was used together with a 6-311++G(d,p) basis set. This combination was chosen as it has been previously shown to describe phthalates energetics with satisfactory performance.[3] Structural parameters result from full geometry optimization in the gas phase, with no imposed constraint. Default SCF and geometry optimization criteria were used.
DFTB computational details: See the Implementation details section in the manuscript.

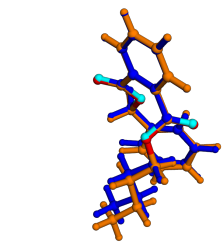


(a) Peak A (b) Peak B

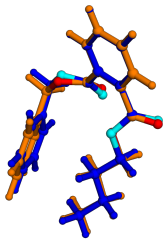
(c) Peak C (d) Peak D

(e) Peak E (f) Most stable DBP structure

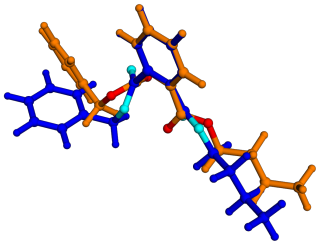
Figure S5: Superposition of the representative structures of the main peaks observed in the structural excitation spectrum of DBP after local minimization at DFTB and DFT level. DFTB: carbon and hydrogen in orange and oxygen in red. DFT: carbon and hydrogen in blue and oxygen in cyan.



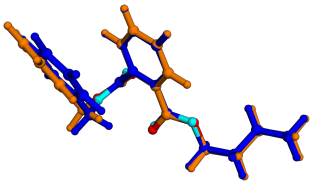
(a') Peak A'



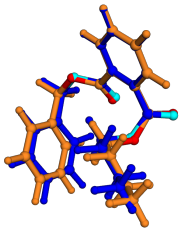
(a) Peak A



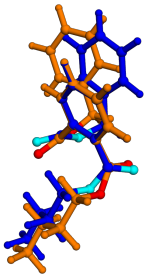
(b) Peak B



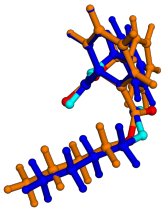
(b') Peak B'



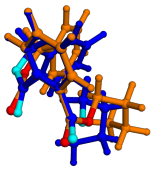
(c) Peak C



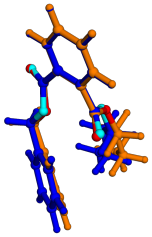
(d') Peak D'



(d) Peak D

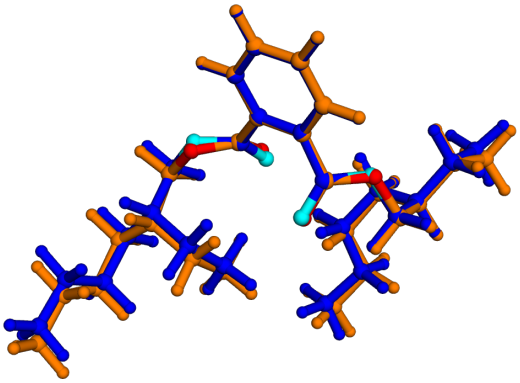


(e) Peak E

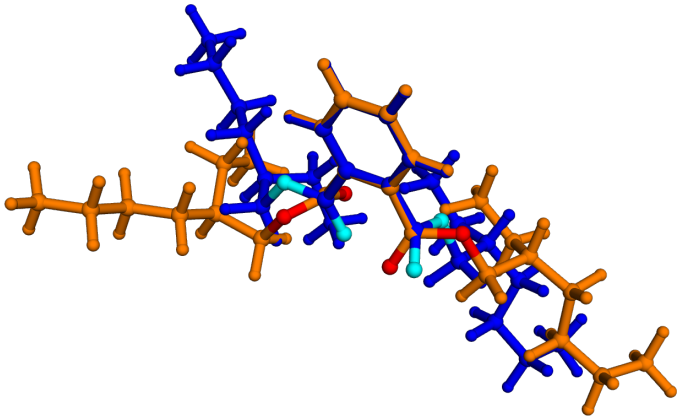


(f) Most stable BBP structure

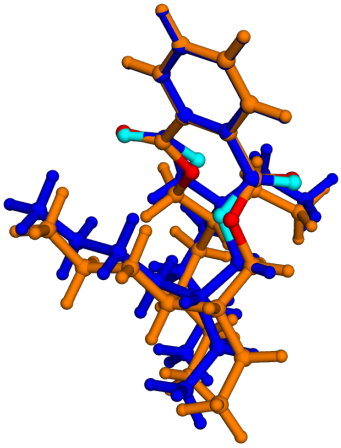
Figure S6: Superposition of the representative structures of the main peaks observed in the structural excitation spectrum of BBP after local minimization at DFTB and DFT level. DFTB: carbon and hydrogen in orange and oxygen in red. DFT: carbon and hydrogen in blue and oxygen in cyan.



(a) Peak A



(b) Peak B



(c) Most stable DEHP structure

Figure S7: Superposition of the representative structures of the main peaks observed in the structural excitation spectrum of DEHP after local minimization at DFTB and DFT level. DFTB: carbon and hydrogen in orange and oxygen in red. DFT: carbon and hydrogen in blue and oxygen in cyan.

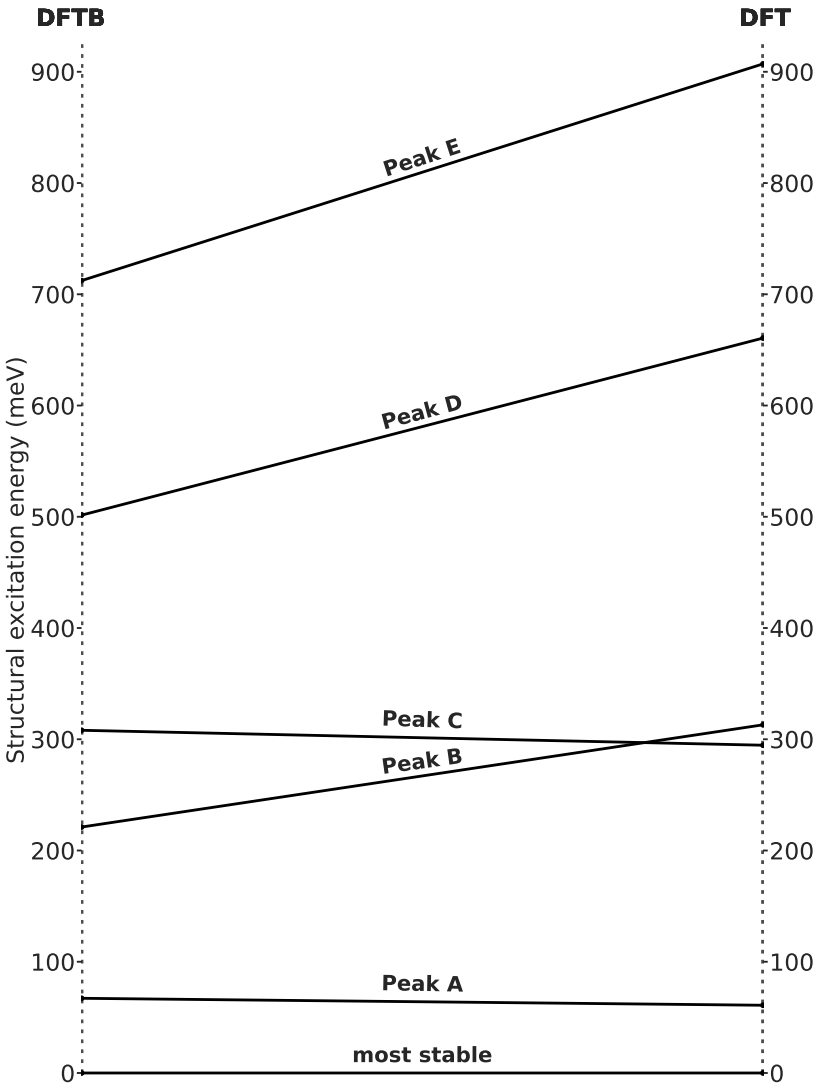


Figure S8: Comparison of DFT and DFTB energies of the characteristic DBP structures of the main peaks: lines connect the DFTB (left) and DFT (right) structural excitation energies (in meV) of the main peaks structures identified in figure 4. The DFTB(resp. DFT) structural excitation energy reference correspond to the lowest-energy structure computed at the DFTB(resp. DFT) level.

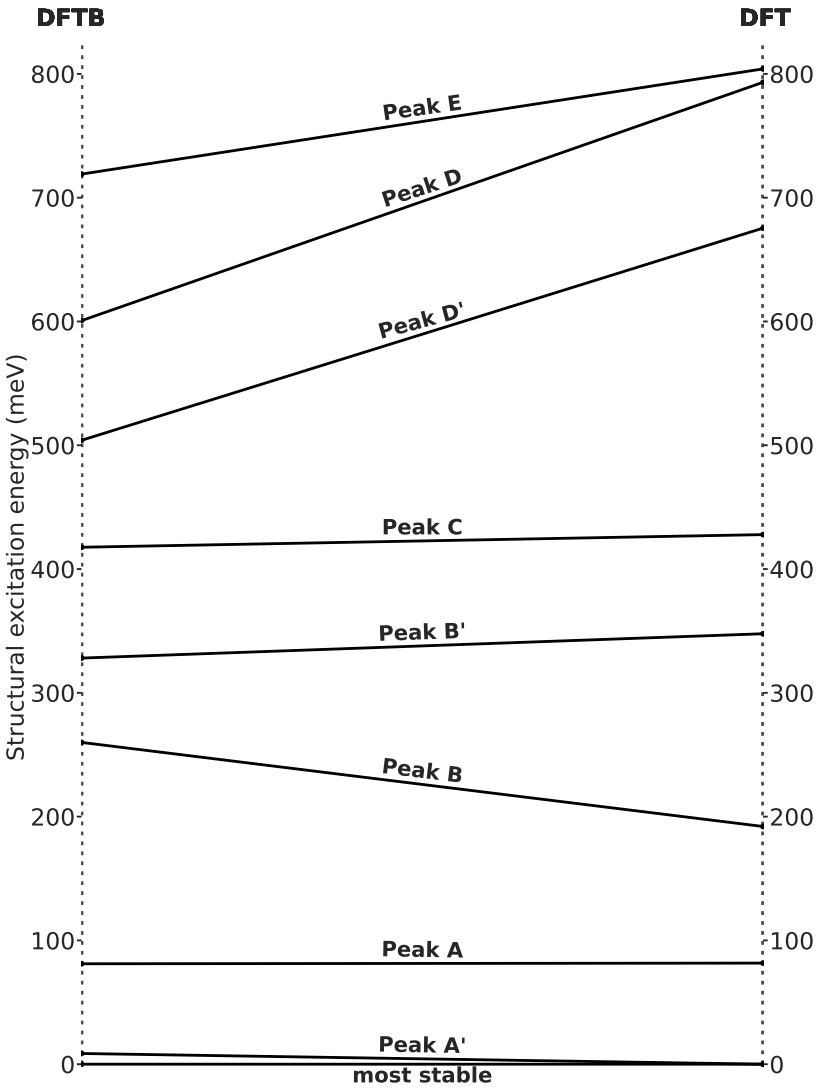


Figure S9: Comparison of DFT and DFTB energies of the characteristic BBP structures of the main peaks: lines connect the DFTB (left) and DFT (right) structural excitation energies (in meV) of the main peaks structures identified in figure 4. The DFTB(resp. DFT) structural excitation energy reference correspond to the lowest-energy structure computed at the DFTB(resp. DFT) level.

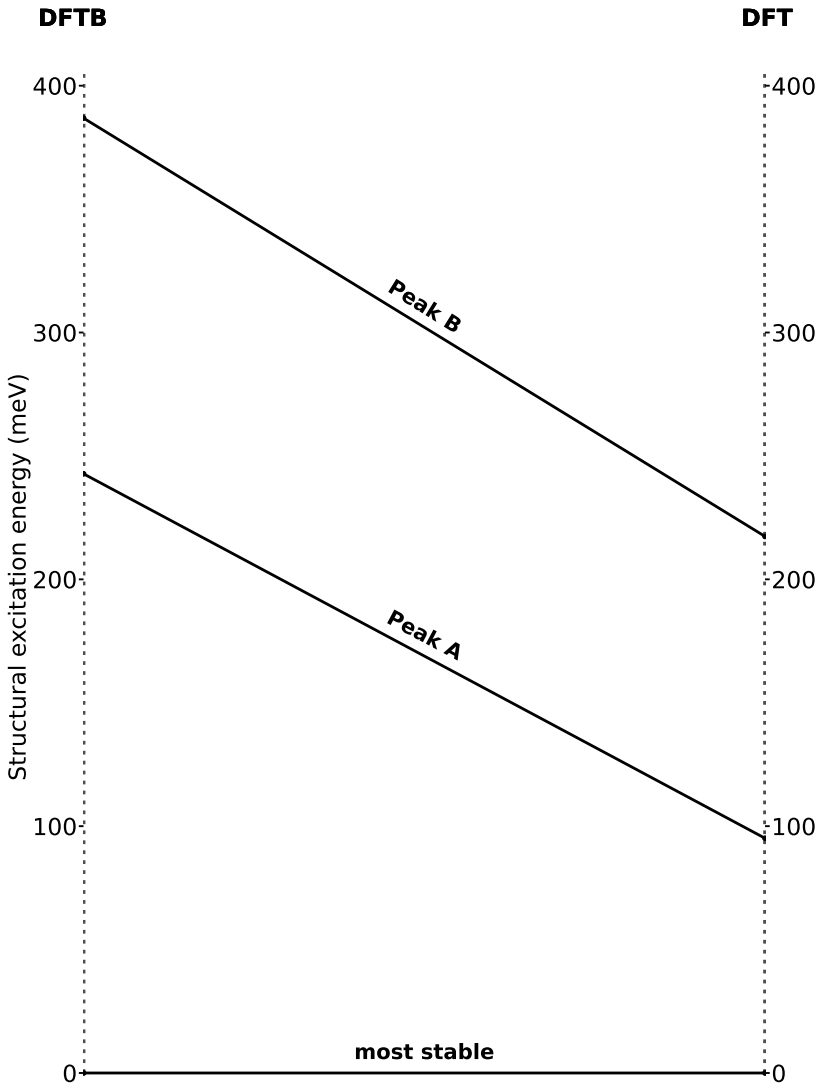


Figure S10: Comparison of DFT and DFTB energies of the characteristic DEHP structures of the main peaks: lines connect the DFTB (left) and DFT (right) structural excitation energies (in meV) of the main peaks structures identified in figure 4. The DFTB(resp. DFT) structural excitation energy reference correspond to the lowest-energy structure computed at the DFTB(resp. DFT) level.

References

(1) Frisch, M. J. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(2) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(3) Poskrebyshev, G. A. The standard thermochemical properties of the p-Benzylphenol and Dimethyl phthalate, and their temperature dependencies. *Comput. Theor. Chem.* **2021**, *1197*, 113146.