



OPEN

Ab initio determination of crystal stability of di-*p*-tolyl disulfide

Xuan Hao^{1,2,7}, Jinfeng Liu^{3,7}, Imran Ali^{2,7}, Hongyuan Luo², Yanqiang Han², Wenxin Hu⁴, Jinyun Liu⁵✉, Xiao He^{4,6}✉ & Jinjin Li²✉

With the rapid growth of energy demand and the depletion of existing energy resources, the new materials with superior performances, low costs and environmental friendliness for energy production and storage are explored. Di-*p*-tolyl disulfide (*p*-Tol₂S₂) is a typical lubricating material, which has been applied in the field of energy storage. The conformational properties and phase transformations of *p*-Tol₂S₂ have been studied by pioneers, but their polymorphs and the polymorphism induced crystal structure changes require further analysis. In this study, we perform the crystal structural screening, prediction and optimization of *p*-Tol₂S₂ crystal with quantum mechanical calculations, i.e., density functional theory (DFT) and second-order Møller–Plesset perturbation (MP2) methods. A series of crystal structures with different molecular arrangements are generated based on the crystal structure screening. As compared to long-established lattice energy calculation, we take an advantage of using more accurate technique, which is Gibbs free energy calculation. It considers the effects of entropy and temperature to predict the crystal structures and energy landscape. By comparing the Gibbs free energies between predicted and experimental structures, we found that phase α is the most stable structure for *p*-Tol₂S₂ crystal at ambient temperature and standard atmospheric pressure. Furthermore, we provide an efficient method to discriminate different polymorphs that are otherwise difficult to be identified based on the Raman/IR spectra. The proposed work enable us to evaluate the quality of various crystal polymorphs rapidly.

Organic disulfide has been one of the major interesting research materials due to its superior lubricating property¹ and the extensive applications in storing energy². Di-*p*-tolyl disulfide (*p*-Tol₂S₂, with the formula of C₁₄H₁₄S₂), a typical representative molecule of a large group of diphenyl disulfides, has been explored in many aspects^{3–5}. A number of analyses have examined the interplay of conformational transformation and the thermodynamic stability of the *p*-Tol₂S₂, demonstrating that the change in pressure^{6–9} or recrystallization^{10–12} would result in structural phase transition. Previous studies have primarily concentrated on the pressure effect on the conformational qualities of *p*-Tol₂S₂ and its molecular aggregation in crystalline state^{13–17}, but the polymorphs and polymorphism induced difference either in the crystal structure or the physical and chemical properties have not been studied. The present paper focuses on the polymorphs of *p*-Tol₂S₂, studying the stability of different forms and providing an effective spectral method to distinguish them. There are several different polymorphs of *p*-Tol₂S₂ with different conformations and molecular aggregations. The two main structures are phase α and phase β . Phase α is an independent molecule ($Z' = 1$), which has an asymmetric conformation, with the monoclinic unit cell of space group P2₁. At higher pressures (> 1.6 GPa), phase α transform into a steric hindrance and aggregated phase β with reduced lattice parameters and volume. Structure β is a triclinic phase with two independent molecules ($Z' = 2$) per unit cell and has a space group of P₁ (see Fig. 1). The structures of *p*-Tol₂S₂ were determined by X-ray diffraction and its cell parameters can be found in the Cambridge Structural Database (CSD).

¹School of Chemistry and Molecular Engineering, Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, East China Normal University, Shanghai 200062, China. ²Key Laboratory for Thin Film and Microfabrication of Ministry of Education, Department of Micro/Nano-Electronics, Shanghai Jiao Tong University, Shanghai 200240, China. ³State Key Laboratory of Natural Medicines, Department of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 210009, China. ⁴School of Computer Science and Software Engineering, The Computer Center, East China Normal University, Shanghai 200062, China. ⁵Anhui Provincial Engineering Laboratory for New-Energy Vehicle Battery Energy-Storage Materials, Anhui Laboratory of Molecule-Based Materials, College of Chemistry and Materials Science, Anhui Normal University, Wuhu 241002, Anhui, China. ⁶NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China. ⁷These authors contributed equally: Xuan Hao, Jinfeng Liu and Imran Ali. ✉email: jyliu@ahnu.edu.cn; xiaoh@phy.ecnu.edu.cn; lijijin@sjtu.edu.cn

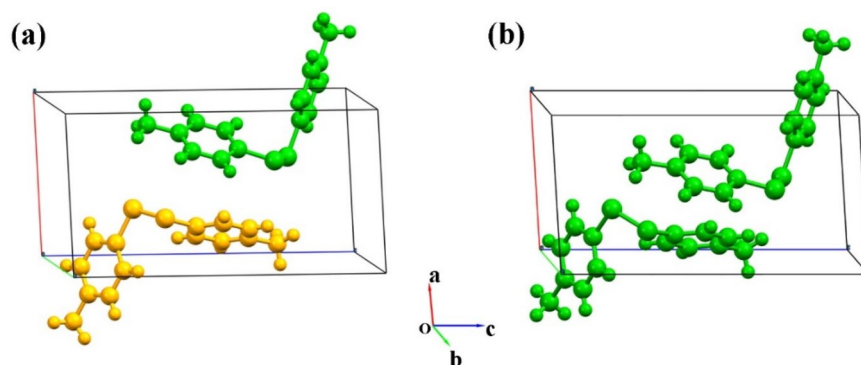


Figure 1. Crystal structures of p -Tol₂S₂. Phase α (a) has a space group of $P2_1$, while phase β (b) has a space group of P_1 . The molecules are colored by the symmetry operation, where the unit cells of phase α and phase β contain one and two independent molecules, respectively.

Polymorphisms, differing in their physical properties, have different solubility and dissolution rates. Therefore, discrimination of polymorphisms is important for all molecular crystal products and industrial production. Crystal structure prediction (CSP) is an emerging tool to help crystallographers determine various thermodynamically favorable crystal packings which are independent of the kinetics of crystallization. However they depend on a wide range of controllable or less easily controllable crystallization condition^{18–20}. The advancement and development of CSP have been widely performed on different polymorphisms^{21–23}. In this paper, to obtain more reliable structures, we use different CSP tools of MOLPAK^{24,25} (MOlecular PAcKing) and USPEX^{26–28} to perform crystal structure screening, prediction and optimization of p -Tol₂S₂. The MOLPAK and USPEX totally produce 9000+ possible hypothetical crystal structures which lie in the most common space groups with different molecular arrangements. The 38 candidates (20 from MOLPAK and 18 from USPEX) with the lowest lattice energies are selected to perform further optimization, where the crystal structure optimization is calculated at the DFT level using the ω B97XD/6-31G* method, and the single-point electronic energy is calculated at the MP2 level on the optimized crystal structure. The MP2 calculation is used to obtain more accurate and reliable prediction results. Besides, to treat the macromolecular system, such as the p -Tol₂S₂ (has 30 atoms) crystal in the present paper, the embedded-fragment method^{29–31} is used to obtain the total energy by taking an appropriate combination of the energies of monomers and dimers, which are embedded in the electrostatic field of the rest in the crystalline environment. Via the Gibbs free energy calculation, we determine that phase α is the most stable structure from those predicted and experimental structures.

Methods

Crystal structure predictions. The credibility of computing methods for crystal structure prediction has been improved over the past years. Here we use the MOLPAK and USPEX tools to determine the crystal forms from the molecular structure. MOLPAK can create unit cells that have the smallest volumes per molecule by allowing surrounding molecules to collapse along the three axes of a Cartesian coordinate system until a predetermined repulsion threshold with the central molecule is reached^{32,33}. While the USPEX program, adapting the evolutionary algorithm that can find the stable structures for systems with various rigid molecular shapes^{36,34}. The initial structures are generally generated by selected space groups with the initial known structures under atmospheric pressure. Structure relaxation is implemented from low to high precision using the VASP program. All the structures are examined and their initial sorting was performed based on lattice energy or enthalpy after full relaxation. However selected potential structures were resorted according to root-mean-square deviations (RMSDs), given in Tables S1 and S2.

Structure optimization and energy calculation. After generation of the initial crystal structures using MOLPAK and USPEX, for further calculation, 20 structures from MOLPAK and 18 structures from USPEX were selected based on lower lattice energies and enthalpies, respectively. To minimize the enthalpy, we adopt quasi-Newton algorithm³⁵ for the crystal structure optimization. Because of the large molecular size and the quantity of molecules in the supercell, the embedded-fragment method^{30,31,36} with ω B97XD/6-31G* was used for the calculation of enthalpy and Gibbs free energy of the p -Tol₂S₂ crystal. The Hessian matrix approximation was updated using the BFGS procedure³⁷ and the convergence criterion for the maximum gradient was set to 0.001 Hartree/Bohr. The embedded fragment method, as the name implies, is a process that breaks a large system into small fragments. The embedded fragment method divides the total energy per unit cell of the crystal into an appropriate combination of the energies of monomers and dimers, and therefore can treat the macromolecules effectively. The individual molecule is designated as a segment, and the interaction energies between two segments, in close contact with each other, are calculated by quantum mechanics (QM). The interaction energy between the two distal segments is calculated according to Coulomb interactions. The DFT calculations, accounting for the environmental influences, are embedded in an electrostatic field represented by the point charges of the remaining system.

The internal energy of a unit cell for the molecular crystal system is calculated by:

$$E_e = \sum_i E_{i(0)} + \sum_{\substack{i,j, i < j \\ R_{ij} \leq \lambda}} (E_{i(0)j(0)} - E_{i(0)} - E_{j(0)}) + \frac{1}{2} \sum_{N=-S}^S (1 - \delta_{n0}) \sum_{\substack{i,j \\ R_{ij} \leq \lambda}} (E_{i(0)j(n)} - E_{i(0)} - E_{j(n)}) + E_{LR}, \quad (1)$$

where the three-integer index of unit cell is represented by a variable 'n'. The quantum mechanical (QM) energy of the *i*th molecules in the *n*th unit cell is represented by $E_{i(n)}$. The QM energy of dimers is represented by $E_{i(0)j(n)}$, where *i*th and *j*th represent number of molecules with respect to 0th central unit cell and *n*th unit cells^{29,38–41}. The crystal system is represented by a $3 \times 3 \times 3$ supercell. The 1st part of Eq. (1) calculates all the single molecular energies in the 0th unit cell, which is in the center. The 2nd part represents the two-body QM interaction that has a shorter distance than λ (where λ is a given cutoff distance which is set to 4 Å in this work). The 3rd part of Eq. (1) provides the interactions between one molecule in the central unit cell and the other in *n*th unit cell whose distance is shorter than λ . The QM is used to calculate short range interactions (the first three parts in Eq. (1)). It is calculated in electrostatic field of the rest, where ω B97XD/6-31G* level is used to fit electrostatic potential variations. The background charges are represented by the $11 \times 11 \times 11$ supercell. The Coulomb's charge-charge interaction is employed to approximately treat long-range interactions between two molecules of dimers whose distance is larger than λ . The long-range electrostatic interactions are represented by E_{LR} in a $41 \times 41 \times 41$ supercell.

The enthalpy H_e for each unit cell can be calculated by considering the effect of external pressure as follow:

$$H_e = E_e + PV, \quad (2)$$

where the external pressure is represented by P and the unit cell volume is represented by V .

$$U_v = \frac{1}{K} \sum_n \sum_{\mathbf{k}} \omega_{n\mathbf{k}} \left(\frac{1}{2} + \frac{1}{e^{\beta\omega_{n\mathbf{k}}} - 1} \right), \quad (3)$$

$$S_v = \frac{1}{\beta TK} \sum_n \sum_{\mathbf{k}} \left\{ \frac{\beta\omega_{n\mathbf{k}}}{e^{\beta\omega_{n\mathbf{k}}} - 1} - \ln(1 - e^{-\beta\omega_{n\mathbf{k}}}) \right\}, \quad (4)$$

$$G_e = H_e + U_v - TS_v. \quad (5)$$

The harmonic approximation is used to calculate zero-point vibrational energy U_v and the entropy S_v , which are shown in Eqs. (3) and (4), respectively, where the frequency of phonon with lattice vector \mathbf{k} is represented by $\omega_{n\mathbf{k}}$. $\beta = 1/k_0 T$ and k_0 is the Boltzmann constant. The capital K is the product of all \mathbf{k} , which are evenly spaced grid points in the reciprocal unit cell. The \mathbf{k} -grid of $21 \times 21 \times 21$ has been used in this study, where $K = 9261$. The calculation of Gibbs free energy with effects of temperature and pressure in a unit cell is given by Eq. (5).

As MP2/6-31G* is more accurate than DFT/ ω B97XD/6-31G* but its computational time and cost are very high. Therefore, to be more efficient and effective, we optimized crystal structure as well as calculated enthalpy (H_e) over single point energy, zero-point vibrational energy (U_v), and entropy (S_v) at DFT/ ω B97XD/6-31G* level. Moreover, to obtain better accuracy, MP2/6-31G* level was used to recalculate enthalpy (H_e) only, over single point energy based on crystal structure calculated by DFT/ ω B97XD/6-31G*.

Raman spectra simulation. For a periodic molecular system, the dynamical force constant matrix is calculated as follow:

$$D(r_p, r_q, k) = \frac{1}{\sqrt{m_p m_q}} \sum_{n=-R}^R H(r_{p,0}, r_{q,n}) e^{-ikR(n)}, \quad (6)$$

where the masses of atoms *p* and *q* are represented by m_p and m_q , respectively.

The \mathbf{k} is a Brillouin zone point which is known. The 2nd order derivative of the total energy in a unit cell which correspond to atom *p* and *q* in the 0th and the *n*th cell, is represented by $H(r_{p,0}, r_{q,n})$ in Eq. (6) at geometry equilibrium. The number of nearby unit cells for QM treatment was truncated at $R = 1$, and the number of \mathbf{k} -points was set to 21 in each of the three dimensions. The vibrational frequencies and the corresponding normal modes of the periodic molecular system can be obtained by solving the dynamical matrix $D(r_p, r_q, k)$. The zone-center ($\mathbf{k} = 0$) vibrations presenting nonzero intensities results in the activation of IR- and Raman spectra. The force-constant matrix $D(\mathbf{0})$ was used to calculate the vibrational frequency. Therefore, we can obtain the IR intensity (I_{n0}) and Raman intensity (R_{n0}) for the fundamental transition of the mode in the *n*th phonon branch with wave vector $\mathbf{k} = 0$, demonstrated by the following derivatives:

$$I_{n0} \propto \sum_i^{l,m,n} \left(\frac{\partial \mu_i}{\partial Q_{n0}} \right)^2, \quad (7)$$

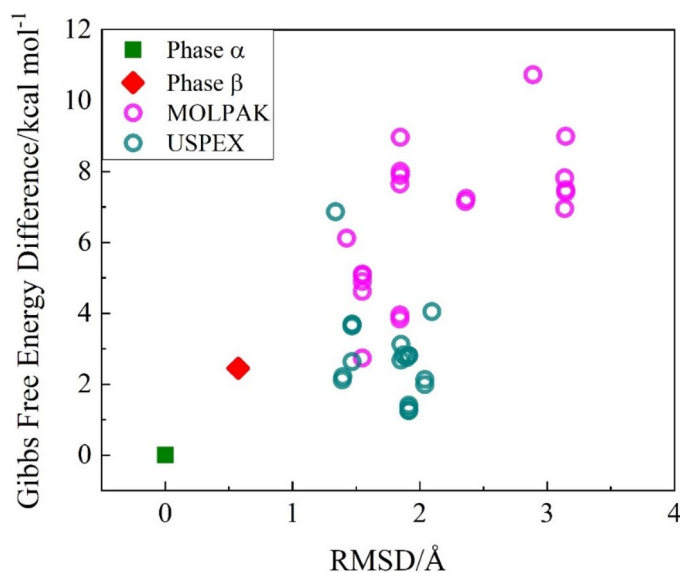


Figure 2. The Gibbs free energy differences per unit cell between the predicted crystal structures and experimental phase α as a function of the unit cell RMSD (with respect to phase α) under 298 K and standard atmospheric pressure, where the purple and green circles signify the 20 MOLPAK and 18 USPEX predicted structures, respectively. The green square and red diamond represent Gibbs free energies of the phases α and β , respectively, which were obtained from experiment.

$$R_{n0} \propto \frac{3}{2} \left(\sum_i^{l,m,n} \frac{\partial \alpha_{ij}}{\partial Q_{n0}} \right)^2 + \frac{21}{2} \sum_i^{l,m,n} \sum_j^{l,m,n} \left(\frac{\partial \alpha_{ij}}{\partial Q_{n0}} \right)^2. \quad (8)$$

The corresponding normal mode is represented by Q_{n0} . The dipole moment derivative $\partial \mu_i / \partial Q_{n0}$ and the polarizability derivative $\partial \alpha_{ij} / \partial Q_{n0}$ in the central unit cell can be derived with the embedded-fragment quantum mechanical method.

Results and discussion

The experimental data of the p -Tol₂S₂ (CH₃-C₆H₄-S)₂ in the Cambridge Structural Database^{42,43} is used to predict the crystal structure. Phase α has a space group of P2₁, with a monoclinic unit cell and a volume of 628.741 Å³. The cell lattice parameters are $a = 7.5927$ Å, $b = 5.71318$ Å, $c = 14.722$ Å, $\alpha = 90^\circ$, $\beta = 94.7615^\circ$, $\gamma = 90^\circ$. However, phase β has a space group of P1 with a unit cell volume of 545.963 Å³. The lattice parameters of unit cell are $a = 7.3057$ Å, $b = 5.5093$ Å, $c = 14.038$ Å, $\alpha = 95.14^\circ$, $\beta = 97.23^\circ$, $\gamma = 85.36^\circ$. Figure 1 represents the molecular structures of both (α and β) phases.

Crystal structure predictions. The crystal structure prediction was performed by the MOLPAK and USPEX programs. From MOLPAK, we obtained 31 standard clusters of crystal structures with common space groups of (P2₁2₁2₁, P $\bar{1}$, P2₁/c, Cc, C2/c, P2₁, Pna2₁ and Pbca). Each cluster contains 300+ crystal structures. Therefore, MOLPAK generated total crystal structures more than 9000. From these 9000+ crystal structures, only 20 crystal structures were selected according to lower lattice energies, and these selected structures were further optimized. All the selected 20 structures from MOLPAK have same lattice energy of -16.91 kcal/mol, given in Table S1. We obtained same lattice energy for 20 selected structures because of two main reasons: first, these structures are similar to each other with minor difference in their lattice parameters, and second, the MOLPAK software does not calculate lattice energy very accurately. Therefore, to obtain accurate energy landscape, we optimized selected structures and calculated their Gibbs free energy at the QM level.

On the other hand, USPEX generated 20 generations with each generation generated about 20 crystal structures. Therefore, the total crystal structures generated by USPEX were about 400. From USPEX predictions, only 18 crystal structures were selected based on enthalpy sorting. After selection of potential candidates we sorted them again, according to RMSDs, given in Table S2. The lattice parameters of phase α , phase β and the 38 predicted new crystal structures by MOLPAK and USPEX are given in Tables S1 and S2 as Supporting Information.

The Gibbs free energy. The stability of 40 crystal structures are compared in Fig. 2, including phase α , phase β and 38 predicted new structures by MOLPAK and USPEX. Figure 2 shows the difference of Gibbs free energy per unit cell among 38 predicted structures, phase α and phase β of the p -Tol₂S₂ as a function of unit cell RMSD. In Fig. 2, the purple and green circles represent the 38 predicted structures by MOLPAK and USPEX, whereas the green square and red diamond denote the experimental Gibbs free energies of phases α and β ,

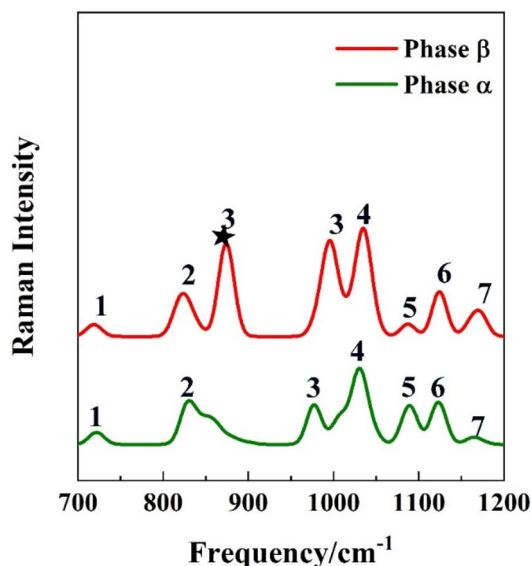


Figure 3. Calculated Raman spectra of *p*-Tol₂S₂ phase α (green) and phase β (red) under standard atmospheric pressure.

respectively. The enthalpy over single point energy of 40 crystal structures was calculated at the MP2/6-31G* level based on the DFT/ ω B97XD/6-31G* optimized crystal structures.

The values of RMSD vary from 1 to 3.5 Å, which suggest that the crystal structure prediction method is reliable. Moreover, the lowest value of RMSD is smaller than 2 Å demonstrating the accuracy of prediction. The correlation between RMSD and Gibbs free energy (as shown in Fig. 2) reveals that the free energy of phase α , locating at the bottom of the lattice energy landscape, is the most stable structure among all predicted and experimental structures. Furthermore, the selected 20 structures from MOLPAK were possessing same lattice energies, given in Table S1. However, in Fig. 2 Gibbs free energy varies from about 1.5 to 8 kcal/mol, which is evidence of accuracy of Gibbs free energy calculation.

To verify the rationality of predicted structures, we selected eight candidates from 38 predicted structures by USPEX and MOLPAK programs to visualize overlay only. Figures S1 and S2 in the Supporting Information present the similarity of crystal structures and the molecular conformations via the calculations of the structural RMSDs, phase α and the predicted structures based on the RDKit tool⁴⁴. We selected a subset of 8 structures from 38 structures based on difference in RMSDs. The green molecules represent the structures of *p*-Tol₂S₂ phase α determined by X-ray diffraction and the red molecules denote the predicted structures based on USPEX and MOLPAK programs. The molecular overlay shows the reliability of the predicted structures.

Vibrational spectra. The crystal structure of a molecule can be identified by vibrational spectroscopy^{45,46}. The calculated full range of Raman spectra are provided in Figs. S3 and S4 of the Supporting Information. For the purpose of highlighting the differences of phase α and phase β , we select parts of the calculated Raman and IR spectra of *p*-Tol₂S₂ crystals. The Raman and IR spectra are shown in Figs. 3 and 4. In Fig. 3, the Raman spectra are calculated from 700 to 1200 cm⁻¹, where phase α (green curve) and phase β (red curve) are predicted to display seven and eight discernible intense Raman bands, respectively. The band 3*, locating at ~874 cm⁻¹, in the red curve is the characteristic peak for phase β . Similarly, there are five and four IR bands for phase α (green curve) and phase β (red curve), respectively. The band 2*, locating at ~3220 cm⁻¹, in the green curve is the characteristic peak for phase α . Therefore, the band 3* in the Raman spectra and the band 2* in the IR spectra, are the characteristic bands to distinguish phases α and β of *p*-Tol₂S₂. Figures 3 and 4 show that although the structures of phases α and β are very similar (as shown in Fig. 1), they can be identified by the characteristic bands in the vibrational spectra. The vibrational spectra of different pressure about phase α and phase β are demonstrated in Figs. S5–S8 in the Supporting Information. We can observe the vibrational spectra as a function of different pressures.

Conclusions

It is highly desirable to control polymorph for all the manufacturing industries which manufacture and/or utilize crystalline products. This will enable us to enhance functionality of the products to next higher level. A high precision method to screen different polymorphs for designing of crystals and to find out the most stable structure that confers an optimal product would, therefore, be extremely valuable. In this paper, we carried out a theoretical approach to screen different polymorphs of *p*-Tol₂S₂, a lubricating and energy storage material, based on the DFT and MP2 methods with the embedded fragment approach. With the theoretical calculation, we pick out the most stable structure phase α from the predicted and experimental structures. Considering the small RMSDs, there will be hundreds of predicted structures between them and the conformation of existing phases

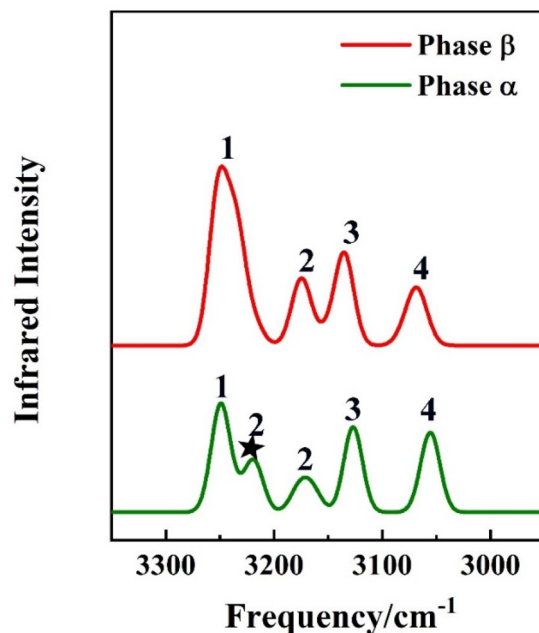


Figure 4. Calculated IR spectra of *p*-Tol₂S₂ phase α (green) and phase β (red) under standard atmospheric pressure.

α and β , the proposed method has successfully performed the predictions for sufficiently similar conformations along with the identification of the most stable structure. The predicted vibrational spectra provide an effective way to identify polymorphs with the characteristic peaks of IR and Raman spectra.

Received: 1 January 2020; Accepted: 16 March 2021

Published online: 29 March 2021

References

1. Qu, J. *et al.* Synergistic effects between phosphonium-alkylphosphate ionic liquids and zinc dialkyldithiophosphate (ZDDP) as lubricant additives. *Adv. Mater.* **27**, 4767–4774 (2015).
2. Wang, B. J. *et al.* An intercalated graphene/(molybdenum disulfide) hybrid fiber for capacitive energy storage. *J. Mater. Chem. A* **5**, 925–930 (2017).
3. Arisawa, M. & Yamaguchi, M. Rhodium-catalyzed disulfide exchange reaction. *J. Am. Chem. Soc.* **125**, 6624–6625 (2003).
4. Sobczak, S. & Katrusiak, A. Colossal strain release by conformational energy up-conversion in a compressed molecular crystal. *J. Phys. Chem. C* **121**, 2539–2545 (2017).
5. Taniguchi, T., Fujii, T., Idota, A. & Ishibashi, H. Reductive addition of the benzenethiyl radical to alkynes by amine-mediated single electron transfer reaction to diphenyl disulfide. *Org. Lett.* **11**, 3298–3301 (2009).
6. Wu, M., Tse, J. S., Wang, S. Y., Wang, C. Z. & Jiang, J. Z. Origin of pressure-induced crystallization of Ce₇₅Al₂₅ metallic glass. *Nat. Commun.* **6**, 6493 (2015).
7. Jenei, Z. *et al.* Structural phase transition in vanadium at high pressure and high temperature: Influence of nonhydrostatic conditions. *Phys. Rev. B* **83**, 054101 (2011).
8. Hui, X., Duclos, S. J., Ruoff, A. L. & Vohra, Y. K. New high-pressure phase transition in zirconium metal. *Phys. Rev. Lett.* **64**, 204–207 (1990).
9. Catalan, G. & Scott, J. F. Physics and applications of Bismuth ferrite. *Adv. Mater.* **21**, 2463–2485 (2009).
10. Anzellini, S., Dewaele, A., Mezouar, M., Loubeyre, P. & Morard, G. Melting of iron at earth's inner core boundary based on fast X-ray diffraction. *Science* **340**, 464–466 (2013).
11. Kuriakose, M. *et al.* In situ imaging of the dynamics of photo-induced structural phase transition at high pressures by picosecond acoustic interferometry. *New J. Phys.* **19**, 053026 (2017).
12. Li, L., Liu, T., Zhao, L. & Yuan, W. K. CO₂-induced phase transition of isotactic poly-1-butene with form III upon heating. *Macromolecules* **44**, 4836–4844 (2011).
13. Neumann, M. A., de Streek, J. V., Fabbiani, F. P. A., Hidber, P. & Grassmann, O. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nat. Commun.* **6**, 7793 (2015).
14. Boldyreva, E. V., Ivashkevskaya, S. N., Sowa, H., Ahsbabs, H. & Weber, H. P. Effect of high pressure on crystalline glycine: A new high-pressure polymorph. *Dokl. Phys. Chem.* **396**, 111–114 (2004).
15. Boldyreva, E. V. High-pressure-induced structural changes in molecular crystals preserving the space group symmetry: Anisotropic distortion/isosymmetric polymorphism. *Cryst. Eng.* **6**, 235–254 (2003).
16. Fabbiani, F. P. A. & Pulham, C. R. High-pressure studies of pharmaceutical compounds and energetic materials. *Chem. Soc. Rev.* **35**, 932–942 (2006).
17. Ostrowska, K., Kropidłowska, M. & Katrusiak, A. High-pressure crystallization and structural transformations in compressed R,S-ibuprofen. *Cryst. Growth Des.* **15**, 1512–1517 (2015).
18. Terayama, K., Yamashita, T., Oguchi, T. & Tsuda, K. Fine-grained optimization method for crystal structure prediction. *NPJ Comput. Mater.* **4**, 32 (2018).

19. Ryan, K., Lengyel, J. & Shatruk, M. Crystal structure prediction via deep learning. *J. Am. Chem. Soc.* **140**, 10158–10168 (2018).
20. Hoser, A. A., Sovago, I., Lanzac, A. & Madsen, A. O. A crystal structure prediction enigma solved: the gallic acid monohydrate system—Surprises at 10 K. *Chem. Commun.* **53**, 925–928 (2017).
21. Hulme, A. T., Price, S. L. & Tocher, D. A. A new polymorph of 5-fluorouracil found following computational crystal structure predictions. *J. Am. Chem. Soc.* **127**, 1116–1117 (2005).
22. Bhardwaj, R. M. *et al.* Exploring the experimental and computed crystal energy landscape of olanzapine. *Cryst. Growth Des.* **13**, 1602–1617 (2013).
23. Baías, M. *et al.* De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography. *J. Am. Chem. Soc.* **135**, 17501–17507 (2013).
24. Anghel, A. T., Day, G. M. & Price, S. L. A study of the known and hypothetical crystal structures of pyridine: Why are there four molecules in the asymmetric unit cell?. *CrystEngComm* **4**, 348–355 (2002).
25. Holden, J. R., Zuyue, D. & Ammon, H. L. Prediction of possible crystal structures for C-, H-, N-, O-, and F-containing organic compounds. *J. Comput. Chem.* **14**, 422–437 (1993).
26. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX—Evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713–720 (2006).
27. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
28. Oganov, A. R. & Glass, C. W. Evolutionary crystal structure prediction as a tool in materials design. *J. Phys. Condens. Matter* **20**, 064210 (2008).
29. Li, J. J., Sode, O., Voth, G. A. & Hirata, S. A solid-solid phase transition in carbon dioxide at high pressures and intermediate temperatures. *Nat. Commun.* **4**, 141–155 (2013).
30. Liu, J. F., Qi, L. W., Zhang, J. Z. H. & He, X. Fragment quantum mechanical method for large-sized ion-water clusters. *J. Chem. Theory Comput.* **13**, 2021–2034 (2017).
31. He, X., Zhu, T., Wang, X. W., Liu, J. F. & Zhang, J. Z. H. Fragment quantum mechanical calculation of proteins and its applications. *Accounts Chem. Res.* **47**, 2748–2757 (2014).
32. Ammon, H. L. *et al.* Structure of 1,3,5,7-tetranitro-3,7-diazabicyclo 3.3.0 octane. Structure solution from molecular packing analysis. *Acta Crystallogr. Sect. B Struct. Commun.* **52**, 352–356 (1996).
33. McArdle, P., Gilligan, K., Cunningham, D., Dark, R. & Mahon, M. A method for the prediction of the crystal structure of ionic organic compounds—The crystal structures of o-toluidinium chloride and bromide and polymorphism of bicifadine hydrochloride. *CrystEngComm* **6**, 303–309 (2004).
34. Kuklin, M. S. & Karttunen, A. J. Crystal structure prediction of magnetic transition-metal oxides by using evolutionary algorithm and hybrid DFT methods. *J. Phys. Chem. C* **122**, 24949–24957 (2018).
35. Dennis, J. E. Jr. & Moré, J. J. Quasi-Newton methods, motivation and theory. *SIAM Rev.* **19**, 46–89 (1977).
36. Beran, G. J. O. Modeling polymorphic molecular crystals with electronic structure theory. *Chem. Rev.* **116**, 5567–5613 (2016).
37. Head, J. D. & Zerner, M. C. A Broyden-Fletcher-Goldfarb-Shanno optimization procedure for molecular geometries. *Chem. Phys. Lett.* **122**, 264–270 (1985).
38. Liu, J. F. & He, X. Accurate prediction of energetic properties of ionic liquid clusters using a fragment-based quantum mechanical method. *Phys. Chem. Chem. Phys.* **19**, 20657–20666 (2017).
39. Liu, J. F., He, X., Zhang, J. Z. H. & Qi, L. W. Hydrogen-bond structure dynamics in bulk water: Insights from ab initio simulations with coupled cluster theory. *Chem. Sci.* **9**, 2065–2073 (2018).
40. Liu, J. F., He, X. & Zhang, J. Z. H. Structure of liquid water—A dynamical mixture of tetrahedral and “ring-and-chain” like structures. *Phys. Chem. Chem. Phys.* **19**, 11931–11936 (2017).
41. He, X., Sode, O., Xantheas, S. S. & Hirata, S. Second-order many-body perturbation study of ice Ih. *J. Chem. Phys.* **137**, 204505 (2012).
42. Bruno, I. J. *et al.* New software for searching the Cambridge structural database and visualizing crystal structures. *Acta Crystallogr. Sect. B Struct. Commun.* **58**, 389–397 (2002).
43. Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. Sect. B-Struct. Commun.* **58**, 380–388 (2002).
44. Landrum, G., RDKit: Open-source cheminformatics. Release 2014.03.1. *Zenodo*. <https://doi.org/10.5281/zenodo.10398>.
45. Khalil, M., Demirdoven, N. & Tokmakoff, A. Obtaining absorptive line shapes in two-dimensional infrared vibrational correlation spectra. *Phys. Rev. Lett.* **90**, 047401 (2003).
46. Hirata, S., Gilliard, K., He, X., Li, J. J. & Sode, O. Ab initio molecular crystal structures, spectra, and phase diagrams. *Accounts Chem. Res.* **47**, 2721–2730 (2014).

Acknowledgements

The authors are grateful for the financial support provided by the National Natural Science Foundation of China (Nos. 21901157, 21673074, 21761132022, and 21703289), the National Key R&D Program of China (No. 2016YFA0501700), the Intergovernmental International Scientific and Technological Cooperation of Shanghai (No. 17520710200), the Science and Technology Major Project of Anhui Province (No. 18030901093), the Foundation of Anhui Laboratory of Molecule-Based Materials (No. FZJ19014), the Shanghai Municipal Natural Science Foundation (No. 18ZR1412600), the Young Top-Notch Talent Support Program of Shanghai, and the NYU-ECNU Center for Computational Chemistry at NYU Shanghai. We also thank the Supercomputer Center of East China Normal University (ECNU Multifunctional Platform for Innovation 001) for providing us computational time.

Author contributions

J.L. (Jinjin Li), X.H. (Xiao He) and J.L. (Jinyun Liu) proposed the idea and supervised the project. X.H. (Xuan Hao), J.L. (Jinfeng Liu) and I.A. performed the calculation and wrote the paper. H. L., Y.H. and W.H. completed the data analysis. I.A. revised the English grammar. All authors contributed to revising of the paper.

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/s41598-021-86519-1>.

Correspondence and requests for materials should be addressed to J.L., X.H. or J.L.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021