Temperature-Accelerated Method for Exploring Polymorphism in Molecular Crystals Based on Free Energy

Tang-Qing Yu[1,*] and Mark E. Tuckerman[2,†]

[1]Department of Chemistry, New York University, New York, New York 10003, USA

[2]Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, New York 10003, USA

(Received 23 March 2011; revised manuscript received 4 May 2011; published 29 June 2011)

The ability of certain organic molecules to form multiple crystal structures, known as polymorphism, has important ramifications for pharmaceuticals and high energy materials. Here, we introduce an efficient molecular dynamics method for rapidly identifying and thermodynamically ranking polymorphs. The new method employs high temperature and adiabatic decoupling to the simulation cell parameters in order to sample the Gibbs free energy of the polymorphs. Polymorphism in solid benzene is revisited, and a resolution to a long-standing controversy concerning the benzene II structure is proposed.

DOI: 10.1103/PhysRevLett.107.015701 PACS numbers: 64.70.kt, 05.70.Ce, 31.15.xv, 61.66.Hq

Structural diversity abounds in nature and in chemistry, an area where structural diversity has profound implications is molecular crystals. Small organic molecules can crystallize into a variety of different forms, giving rise to the phenomenon of *polymorphism*. While polymorphism is important in numerous contexts, there are few in which the stakes are as high as they are in pharmaceutical applications [1]. In the anti-AIDS drug ritonavir, for example, an unexpected insoluble crystalline form of the compound appeared in the manufacturing process after the drug's launch in 1996 requiring a massive and costly recall in 1998. From this and many other examples [1], it is clear that *a priori* prediction and thermodynamic ranking of the different crystalline polymorphs of a given compound are important problems in which suitable computational techniques can play an important role.

Considerable effort in predicting crystal structures has been invested over several decades, and numerous theoretical methods have been developed [2]. In order to evaluate the performance of different approaches, blind tests of crystal structure prediction have been held every few years by the Cambridge Structural Database [3–6]. The rate of success in predicting crystal structures is increasing, yet it is still low, and theoretical prediction of crystal polymorphism remains an unsolved problem [2]. A reason for this is that most methods base their rankings on lattice energies of different crystal forms, so that thermodynamic information is lost. For free-energy-based methods, the theoretical challenge of exploring crystalline polymorphism stems from the need to sample a complex and rough energy landscape in order to obtain free energy differences between different structures. In fact, polymorphism prediction has been compared to the conformational exploration of proteins [7]. While the two challenges are very different, they share important features, and, consequently, methods developed for biophysical structure prediction could potentially be adapted for crystal polymorphism exploration [7–11].

In this Letter, we introduce a new technique for rapid exploration and ranking of crystalline polymorphs. The new approach employs our adiabatic free energy dynamics (AFED) [12] scheme for mapping out multidimensional free energy surfaces in complex systems [12]. AFED is a molecular dynamics approach in which an adiabatic decoupling is imposed between a set of collective variables (CVs) of interest and the remainder of the system by assigning the former a high mass. The CVs are also maintained at a temperature T_s much higher than the physical temperature T in order to ensure facile barrier crossing. Under these conditions, it can be shown that if a multidimensional histogram in the collective variables is accumulated over the course of an AFED calculation, then the free energy surface at temperature T is given by the logarithm of this histogram multiplied by $-k_B T_s$. For crystal polymorphism exploration, we construct an isothermal-isobaric (NPT) AFED-based approach, termed *crystal-AFED*, in which the three vectors of the simulation cell are chosen as CVs.

The equations of motion underlying crystal-AFED are those of Martyna, Tobias, and Klein [13,14], which generate an anisotropic NPT ensemble. Suppose the N atoms of a system at temperature T have masses $m_1, \ldots, m_N$, positions $\mathbf{r}_1, \ldots, \mathbf{r}_N$, and momenta $\mathbf{p}_1, \ldots, \mathbf{p}_N$ and interact via a potential $U(\mathbf{r}_1, \ldots, \mathbf{r}_N)$. If the periodic simulation cell of the system is described by vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ ($V = \mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$), the Martyna-Tobias-Klein equations read

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{\mathbf{p}_\mathbf{g}}{W}\mathbf{r}_i, \qquad \dot{\mathbf{h}} = \frac{\mathbf{p}_\mathbf{g}\mathbf{h}}{W}, \tag{1a}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \frac{\mathbf{p}_\mathbf{g}}{W}\mathbf{p}_i - \frac{1}{N_f}\frac{\mathrm{Tr}[\mathbf{p}_g]}{W}\mathbf{p}_i + \mathcal{H}(T), \tag{1b}$$

$$\dot{\mathbf{p}}_\mathbf{g} = V[\mathbf{P}^{(\mathrm{int})} - P\mathbf{I}] + \frac{1}{N_f}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}\mathbf{I} + \mathcal{H}(T), \tag{1c}$$

where $\mathcal{H}(T)$ indicates a heat bath coupling at T, $\mathbf{h}$ is a 3×3 matrix whose columns are the three cell vectors,

$V = \det(\mathbf{h})$, $\mathbf{F}_i = -\partial U/\partial \mathbf{r}_i + \mathbf{f}_i^{(\mathrm{constr})}$ is the force on particle i, including any constraint forces, $\mathbf{I}$ is the 3×3 identity matrix, and N_f is the number of degrees of freedom in the system. The variable $\mathbf{p_g}$ is a 3×3 matrix that acts as a "barostat," ensuring that the internal pressure tensor of the system $\mathbf{P}^{(\mathrm{int})}$, whose elements are $P_{\alpha\beta}^{(\mathrm{int})} = \det(\mathbf{h})^{-1} \sum_{i=1}^{N} [p_{i,\alpha} p_{i,\beta}/m_i + r_{i,\alpha} F_{i,\beta}]$, averages to the external value $P\mathbf{I}$. $W = (N_f/3 + 1)kT\tau^2$ is a masslike parameter where τ determines the time scale of the barostat motion. Note that the two heat baths in Eqs. (1) separately maintain the particles and barostat at temperature T.

Using $\mathbf{h}$ as CVs in crystal-AFED to generate the Gibbs free energy $G(\mathbf{h}, T)$ requires assigning a temperature $T_{\mathbf{h}} \gg T$ to $\mathbf{h}$. This is tantamount to replacing $\mathcal{H}(T)$ in Eq. (1c) with $\mathcal{H}(T_{\mathbf{h}})$. We must also introduce a large value for the mass parameter W in order to ensure an adiabatic decoupling between the cell and particle dynamics. For the large temperature separation, we employ the robust generalized Gaussian moment thermostating method [15] as the heat bath coupling. In order to enhance fluctuations in the cell matrix, we couple diagonal and off-diagonal elements of $\mathbf{p_g}$ to separate thermostats as described in Ref. [16]. In addition, each atomic degree of freedom is coupled to its own thermostat. Over a crystal-AFED calculation, we accumulate a probability distribution $P_{\mathrm{adb}}(\mathbf{h}, T, T_{\mathbf{h}})$ of the cell matrix and recover the Gibbs free energy surface $G(\mathbf{h}, T)$ at temperature T from [12] $G(\mathbf{h}, T) = -k_B T_{\mathbf{h}} \ln P_{\mathrm{adb}}(\mathbf{h}, T, T_{\mathbf{h}})$. Although this formula is, in principle, straightforward to apply, analysis of the high dimensional distribution P_{adb} requires a sophisticated approach such as a clustering algorithm [17,18]. Finally, ideal structures can be generated from crystal-AFED by selecting candidate structures and optimizing them to remove thermal noise at the external pressure P.

In order to test the performance of crystal-AFED, we chose solid benzene, which has attracted considerable experimental and theoretical attention [19] due to its richness in polymorphs. Since the experimental work of Thiéry and Léger [20], several theoretical studies on benzene polymorphism followed [9,21–26]. While benzene I and benzene III are consistently identified both theoretically and experimentally, the benzene II structure has remained controversial, largely due to sample imperfection. A monoclinic unit cell was originally proposed (denoted II88) based on limited powder diffraction patterns [20]. Raman scattering [21] supports the existence of benzene II, yet the identification of the structure has proved elusive. Three hypothetical structures (II96, II98, and II01) for benzene II were proposed (see Table I), based on lattice energy minimization after molecular packing analysis for II96 and II98 and a molecular dynamics search for II01. Since both II98 and II01 have low lattice energies and powder patterns similar to experiment [24,25], it is unclear which structure best represents the true benzene II phase, and additional experiments are needed [19,23]. Benzene polymorphism was not explored thoroughly until a recent metadynamics study [9], and in that work, no clarification is given concerning benzene II. Rather, II98 is taken as the true benzene II structure, and no information about the thermodynamic stability of different benzene polymorphs is given. Using crystal-AFED, we will propose a resolution of the benzene II controversy.

Crystal-AFED calculations of solid benzene at two physical temperatures, $T = 300$ K and $T = 100$ K, and a pressure of 2 GPa were performed by using the PINY_MD package [27] with the Gromos 96 force field [28] used in Ref. [9]. Technical details are given in Ref. [16]. The parameter W is calculated by using $\tau = 8.5$ ps for $T = 100$ K and $\tau = 10$ ps for $T = 300$ K. In choosing the value of $T_{\mathbf{h}}$, we observed that phase transitions are rare for values less than 28 000 K. In the range $T_{\mathbf{h}} \sim 30\,000$–$32\,000$ K, we generate transitions between all of the polymorphs (higher temperatures could also be used). Such a large $T_{\mathbf{h}}$ value suggests high barriers between different free energy minima.

A small trajectory ensemble of a system of 192 benzene molecules was generated by using crystal-AFED with a 1 fs time step. The initial conditions for each trajectory were randomly generated following Ref. [9]. At the chosen values of $T_{\mathbf{h}}$, the system can cross energy barriers freely, leading to smooth solid-to-solid phase transitions. (See [9] for a putative phase diagram.) Example crystal-AFED trajectories are shown in Fig. 1, where (c) shows the transition from benzene I to benzene III (Fig. 2 shows unit cell structures) from a trajectory at 300 K, while (a) and (b) show the evolution of the cell lengths, angles, and molar volume from trajectories at 100 and 300 K. These figures indicate that the cell lengths and angles (derived from $\mathbf{h}$) can distinguish the different polymorphs and,

TABLE I. Proposed phase II structure for benzene.

	SG[a]	a	b	c	β	V_m	Z	P	T(K)
II88 [20]	⋯	4.40	4.87	9.80	102.1	102.66	2	3.1	293
II96 [23]	$P2_1/c$	5.43	6.68	6.65	121.6	102.63	2	3.1	293
II98 [24]	$P4_32_12$	5.29	5.29	14.29	90	100.0	4	3.0	⋯
II01 [25][b]	⋯	5.43	5.43	7.34	107.1	⋯	⋯	3.1	293

[a]SG = space group, cell lengths in angstroms, cell angles in degrees, molar volume V_m in Å^3, and pressure P in GPa.
[b]II01 is essentially phase III with one or more line defects [denoted III(d)]. As the form is not unique, the given cell parameters are approximated from phase III(d) [25].

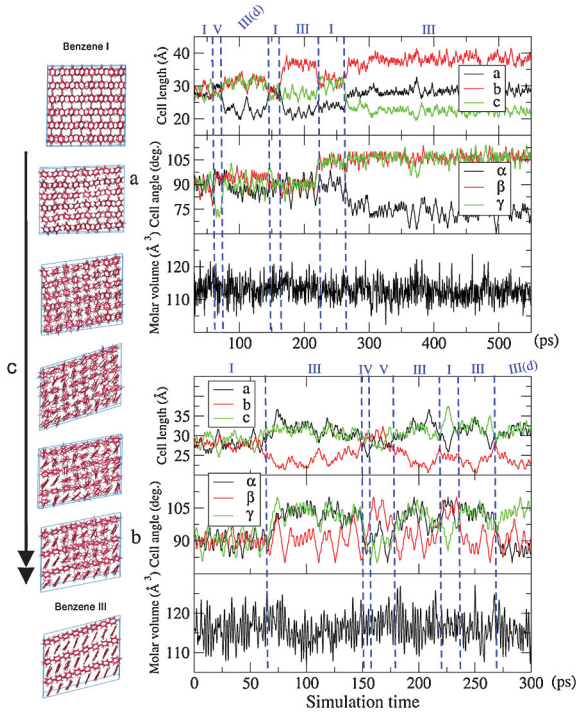

FIG. 1 (color). (a) Crystal-AFED trajectories (with $T_{\mathbf{h}} = 31\,000$ K) for benzene at 100 K; (b) crystal-AFED trajectories for benzene at 300 K; (c) a smooth phase transition from a crystal-AFED trajectory of benzene at 300 K.

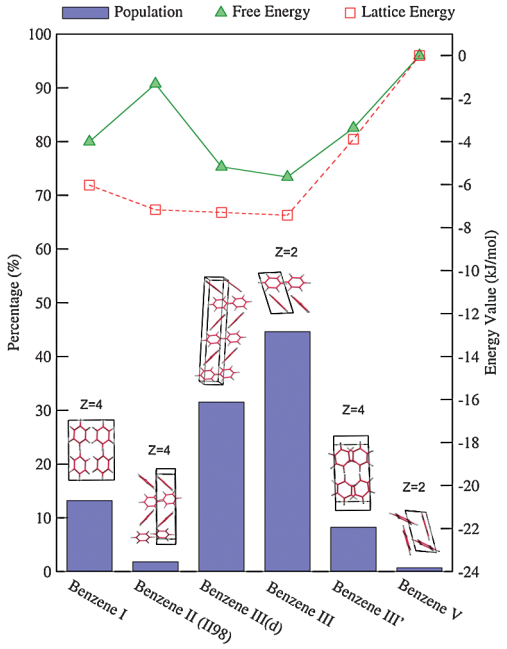

FIG. 2 (color). Probability distribution, unit cell structures, and corresponding free energies and lattice energies of the stable polymorphs of benzene at 100 K and 2 GPa obtained via crystal-AFED. The unit cell for the mixed stacking structure [III(d)] is one representative structure among all those generated in the simulation. The lattice energy is the sum of the intermolecular energy and the PV contribution at 0 K and 2 GPa, which, for III (d), is averaged over several different mixed forms. Both the lattice energy and the free energy of benzene V are set to zero as reference values.

hence, that $\mathbf{h}$ is a good set of CVs for enhanced sampling despite the large values of $T_{\mathbf{h}}$ needed. Using just six independent trajectories at 100 K, we found all of the crystal structures in 470 ps, while at 300 K, with 12 trajectories, we found them within 700 ps.

In order to determine the crystallographic space group, a configuration is selected as a crystal candidate if the configuration is stable in the simulation for $\gtrsim 10$ ps. The unit cell and space group of the optimized structure of the candidate are then determined by using the program PLATON [29]. The predicted crystal structures are summarized in Table II, and a detailed comparison with metadynamics [9] and experiment [20] is given in Ref. [16].

Within the Gromos force field, the predicted benzene I structure deviates slightly from the experimental structure and is actually the same as benzene I′ in the metadynamics study of Ref. [9]. However, we refer to this phase simply as benzene I rather than renaming it benzene I′ [9] because it is the only stable phase corresponding to the experimental I structure (see [16]).

In order to analyze the distribution of crystalline polymorphs and estimate Gibbs free energy differences, we generated a larger ensemble of trajectories at 100 K, totaling 5.43 ns of simulation time from which 54 300 configurations were collected. These were filtered for disordered or amorphous structures using the orientational distribution function $P(\cos\theta)$, where $\cos\theta = \mathbf{u}_i \cdot \mathbf{u}_j$ and $\mathbf{u}_i$ and $\mathbf{u}_j$ are unit vectors perpendicular to the planes of molecules i and j, respectively. $P(\cos\theta)$ will exhibit distinct peaks if the structure is ordered and will be essentially uniform in a disordered state. As a measure of disorder, we employ the information theory entropy corresponding to $P(\cos\theta)$, $S = -\int_{-1}^{1} P(\cos\theta)\ln P(\cos\theta)d(\cos\theta)$, as a filter of disordered states. Larger values of S correspond to a more uniform $P(\cos\theta)$ and more disordered states. We found that an optimal value for S for such a filtering in our study is 4.7.

TABLE II. The benzene polymorphs found in crystal-AFED. A short anisotropic NPT simulation is performed on a crystal candidate to obtain an averaged cell matrix. By using this averaged cell, the structure is quenched to 0 K via simulated annealing to obtain a perfect crystal structure at finite temperature. SG = space group; cell lengths are in angstroms, angles are in degrees, and molar volume (V_m) is in Å^3. The cell angles α and γ are both 90°.

	I	II (II98)	III	III′	IV	V
a	9.44	5.68	5.57	10.70	9.41	5.59
b	7.14	5.68	5.63	5.54	5.86	3.92
c	7.26	14.88	7.68	6.93	6.44	9.82
β	90.0	90.0	110.6	108.4	90.0	93.3
V_m	122.0	120.0	112.5	97.0	88.7	107.5
Z	4	4	2	4	4	2
SG	$Cmca$	$P4_12_12$	$P2_1/c$	$C2/c$	$Pbam$	$P2_1/c$

After screening in this manner, we clustered the remaining 40 001 configurations by using a Gaussian mixture model [17,18] based on the values of the cell lengths and angles (see the clustering diagram in Ref. [16]). From the clusters, we calculated the populations of each polymorph. Because several clusters share the same underlying structure, they must be merged in order to evaluate the final distribution. The population distribution of benzene polymorphs is shown in Fig. 2 with free energies and lattice energies indicated.

From Fig. 2, we see that the free energy and lattice energy measures are comparable and result in a consistent stability ordering except for phase II98. II98 is often found in a mixed stacking structure (see [16]), which is essentially a phase III with line defects [denoted III(d)]. The free energy analysis shows that mixed stacking structures have a considerable thermodynamic stability, while the stability of pure II98 is low despite its low lattice energy. This result addresses a long-standing controversy: Crystal-AFED predicts that mixed stacking structures are thermodynamically more stable at the simulated conditions. Hence, what is observed experimentally for phase II is actually such a mixed structure [II01 or III(d)] rather than II98 [25]. Indeed, phase III(d) has a powder diffraction pattern that gives the closest match to experiment [25].

Benzene IV is found to be stable (or metastable) only at pressures above 5 GPa, and, in our simulations, IV was observed to change quickly to I under an ordinary anisotropic NPT simulation. Crystal-AFED trajectories visit the IV structure but remain there for times sufficiently short as to give them negligible contribution to the distribution ($\Delta G > 0$). A typical pathway observed in crystal-AFED at 2 GPa is from benzene I to V with IV appearing as an intermediate state.

The Gromos force field appears to give a reasonably good prediction of benzene polymorphism at pressures in the range 0–4 GPa [9]. Our overall conclusion, therefore, is that at 2 GPa benzene III is the most stable form at 100 K, while mixed stacking structures have a comparable stability with III. The latter could be due to the fact that 2 GPa is near the phase transition region under this force field. Phase I is the third most stable structure at 2 GPa, which is consistent with the putative phase diagram in Ref. [9].

We have introduced the crystal-AFED approach for exploring crystalline polymorphism under conditions of constant external temperature and pressure in full atomic detail. Crystal-AFED offers several advantages over metadynamics. Being based on a rigorous anisotropic NPT approach with well defined external temperature and pressure, crystal-AFED exhibits remarkable stability. It is straightforward to implement in existing anisotropic NPT simulation codes and contains just two adjustable parameters. Finally, as we have shown for solid benzene, from the multidimensional distribution, we are able to estimate the relative free energies of the different crystalline polymorphs, which were crucial for proposing a resolution to the benzene II controversy.

We thank Dr. Paolo Raiteri for useful exchanges and discussions. This work was supported by the National Science Foundation (CHE-1012545 and DMR-0820341).

*tqy200@nyu.edu
†mark.tuckerman@nyu.edu

[1] J. Bernstein, *Polymorphism in Molecular Crystals* (Oxford University, Oxford, 2002).
[2] S. M. Woodley and R. Catlow, Nature Mater. **7**, 937 (2008).
[3] J. P. M. Lommerse *et al.*, Acta Crystallogr. Sect. B **56**, 697 (2000).
[4] W. D. Sam Motherwell *et al.*, Acta Crystallogr. Sect. B **58**, 647 (2002).
[5] G. M. Day *et al.*, Acta Crystallogr. Sect. B **61**, 511 (2005).
[6] G. M. Day *et al.*, Acta Crystallogr. Sect. B **65**, 107 (2009).
[7] J. D. Dunitz and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **101**, 14 309 (2004).
[8] R. Martoňák, A. Laio, and M. Parrinello, Phys. Rev. Lett. **90**, 075503 (2003).
[9] P. Raiteri, R. Martoňák, and M. Parrinello, Angew. Chem., Int. Ed. **44**, 3769 (2005).
[10] T. Zykova-Timan, P. Raiteri, and M. Parrinello, J. Phys. Chem. B **112**, 13 231 (2008).
[11] P. G. Karamertzanis *et al.*, J. Phys. Chem. B **112**, 4298 (2008).
[12] L. Rosso *et al.*, J. Chem. Phys. **116**, 4389 (2002).
[13] G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).
[14] T. Q. Yu *et al.*, Chem. Phys. **370**, 294 (2010).
[15] Y. Liu and M. E. Tuckerman, J. Chem. Phys. **112**, 1685 (2000).
[16] See supplemental material at http://link.aps.org/supplemental/10.1103/PhysRevLett.107.015701 for technical details of the crystal-AFED calculations, a summary of crystal structures obtained and comparison with the metadynamics study of Ref. [9], a clustering diagram, and an illustration of mixed stacking structures.
[17] G. J. McLachlan and D. Peel, *Finite Mixture Models* (Wiley, Hoboken, NJ, 2000).
[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soc. Ser. B **39**, 1 (1977).
[19] A. V. Dzyabchenko, Russ. J. Phys. Chem. A **82**, 1663 (2008).
[20] M. M. Thiéry and J. M. Léger, J. Chem. Phys. **89**, 4255 (1988).
[21] F. Cansell, D. Fabre, and J.-P. Petitet, J. Chem. Phys. **99**, 7300 (1993).
[22] T. Shoda *et al.*, J. Mol. Struct., Theochem **313**, 321 (1994); **333**, 267 (1995).
[23] M.-M. Thiéry and C. Rérat, J. Chem. Phys. **104**, 9079 (1996).
[24] B. P. van Eijck *et al.*, Acta Crystallogr. **54**, 291 (1998).
[25] Y. Yonetani and K. Yokoi, Mol. Phys. **99**, 1743 (2001).
[26] L. Ciabini *et al.*, Phys. Rev. B **72**, 094108 (2005).
[27] M. E. Tuckerman *et al.*, Comput. Phys. Commun. **128**, 333 (2000).
[28] D. van der Spoel *et al.*, J. Biomol. NMR **8**, 229 (1996).
[29] A. L. Spek, J. Appl. Crystallogr. **36**, 7 (2003).