

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 06-07-2022	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 1-Feb-2019 - 31-Jan-2022
---	--------------------------------	--

4. TITLE AND SUBTITLE Final Report: Physics-based approach to computational design of molecular solids	5a. CONTRACT NUMBER W911NF-19-1-0095
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102

6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES New York University 665 Broadway Suite 801 New York, NY 10012 -2331	8. PERFORMING ORGANIZATION REPORT NUMBER
--	--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 73498-CH.7

12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.
--

13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Mark Tuckerman
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	19b. TELEPHONE NUMBER 212-988-8471

RPPR Final Report

as of 12-Jul-2022

Agency Code: 21XD

Proposal Number: 73498CH

Agreement Number: W911NF-19-1-0095

INVESTIGATOR(S):

Name: Mark E Tuckerman
Email: mark.tuckerman@nyu.edu
Phone Number: 2129888471
Principal: Y

Organization: **New York University**

Address: 665 Broadway, New York, NY 100122331

Country: USA

DUNS Number: 041968306

EIN: 135562308

Report Date: 30-Apr-2022

Date Received: 06-Jul-2022

Final Report for Period Beginning 01-Feb-2019 and Ending 31-Jan-2022

Title: Physics-based approach to computational design of molecular solids

Begin Performance Period: 01-Feb-2019

End Performance Period: 31-Jan-2022

Report Term: 0-Other

Submitted By: Mark Tuckerman

Email: mark.tuckerman@nyu.edu

Phone: (212) 988-8471

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1

STEM Participants:

Major Goals: The overall goal of this project is to develop a comprehensive multi-layer physics-based computational method for modeling of molecular solids that will dramatically expand the current capabilities of in silico predictions. An exhaustive plan to reach this goal contains several elements, and we advanced work on some of them during the present funding period.

(a) First, have completed a project in predicting crystal structures of an energetic molecule, 4-amino-2,3,6-trinitrophenol from first-principles electronic structure calculations and molecular simulation. This physics-based approach consists of a series of steps. First, a tailor-made two-body potential energy surface was constructed with recently developed software, autoPES, using symmetry-adapted perturbation theory based on density-functional theory description of monomers [SAPT (DFT)]. The fitting procedure ensures asymptotic correctness of the PES by employing a rigorous asymptotic multipole expansion which seamlessly integrates with SAPT(DFT) interaction energies. Given the PES, crystal structure prediction is then performed by generating possible crystal structures with rigid molecules, minimizing these structures using the SAPT(DFT) force field, and running isothermal-isobaric molecular dynamics simulations with flexible molecules based on the tailor-made SAPT(DFT) intermolecular force field and a generic intramolecular one. This workflow constitutes a first-principles, bottom-up approach to CSP. In the course of this study, we demonstrated the importance of the intermolecular potential, which we found could be used with a generic bonded/intramolecular potential without significantly compromising accuracy.

(b) Kohn-Sham density functional theory (DFT) is a standard tool in most branches of chemistry, but accuracies for many molecules are limited to 2-3 kcal/mol with presently-available functionals. Wavefunction methods, such as coupled-cluster, routinely produce much higher accuracy, but computational costs limit their application to small molecules. One of our main goals is to leverage machine learning to calculate coupled-cluster energies from DFT densities with the aim of achieving quantum chemical accuracy (errors below 1 kcal/mol) on test data. In addition, density-based Delta-learning (learning only the correction to a standard DFT calculation) can significantly reduce the amount of training data required, particularly when molecular symmetries are included. We highlight the robustness of Delta-DFT by correcting "on the fly" DFT-based molecular dynamics (MD) simulations of resorcinol (C₆H₄(OH)₂) to obtain MD trajectories with coupled-cluster accuracy. We conclude that Delta-DFT facilitates running gas-phase MD simulations with quantum chemical accuracy, even for strained geometries and conformer changes where standard DFT fails. Extensions to bulk phases are currently under way. This work is now being extended to incorporate equivariance and, therefore, to leverage equivariant density-based learning models, which will ensure translational and rotational invariance. Ultimately, our goal is to employ this framework under periodic boundary conditions for studies of molecular crystals.

RPPR Final Report as of 12-Jul-2022

(c) Predicting molecular crystal structures starting only from a two-dimensional diagram of the constituent compound(s) remains a significant challenge. Most structure-prediction protocols require an interatomic interaction model that ultimately determines the quality of the results. We are currently developing a new topological approach to molecular crystal structure prediction. The idea is that in a stable structure, certain atoms occupy positions that correspond to minima of a set of geometric order parameters. By minimizing a distance and orientational metric that quantifies the proximity of atoms to these positions and orientations of molecules to certain crystallographic planes, stable structures and polymorphs for a given crystal can be predicted without reliance on an interaction model. This approach can accelerate the structure prediction step, however, predicting thermodynamics and kinetics of polymorphic transitions, a second critical step still requires force fields such as that discussed above.

(d) We are integrating SAPT(DFT) force fields generated by the AutoPES code developed in the Szalewicz group into a software pipeline for CSP being developed in the Tuckerman group. This pipeline consists of the following components: (i) Zero-temperature random structure searching using either energy optimization of topological algorithms; (ii) finite-temperature/finite-pressure molecular dynamics allowing the unit cells to be further optimized at experimental conditions and false minima on the potential energy surfaces to be eliminated; (iii) molecular dynamics based enhanced sampling based on driving a set of collective variables that allow the crystal free-energy landscape to be explored, thus testing the stability of candidate crystal structures and predict possible polymorphic transitions between them using collective variables derived using machine learning approaches. This protocol has been used in the Tuckerman group for the previous five years, and now other CSP groups are starting to recognize the importance of including finite-temperature molecular dynamics and enhanced sampling as part of a general CSP protocol. We have used this protocol to explore 1:1 co-crystals of resorcinol and urea, as will be discussed in the accomplishments section.

(e) When using theoretical crystal structure prediction calculations to predict structures of molecular crystals, these methods often fail to account for structures that might possess disordered motifs including stacking faults and polytypism. The powder X-ray diffraction (PXRD) pattern of the benzamide form II polymorph contains diffuse scattering consistent with a disordered structure. Previous attempts to elucidate its ordered structure by PXRD and CSP were incomplete because the average structure is actually a complex superposition of multiple polytypes, thus making structure determination particularly challenging by theoretical approaches. In order to address this problem, an approach to modeling such complex structures from PXRD data and CSP in combination with molecular dynamics simulations is proposed. Although the approach is applied, here, to benzamide II, it is applicable to other disordered crystal structures.

(f) The Szalewicz and Tuckerman groups are employing their joint protocol in a collaborative effort to predict crystal structures of targets in the 7th CCDC blind structure prediction competition. Our predictions have now all been submitted, and we are awaiting the final results. The conference on the competition will be held in September in Cambridge, UK.

Accomplishments: (a) The workflow described in the "Major Goals" section applied to the 4-amino-2,3,6-trinitrophenol led to the experimentally-observed structure being identified as one of the forms with the lowest lattice energy, demonstrating the success of a first-principles, bottom-up approach to CSP. The match to the experimental structure has an RMSD20 value of 0.35 angstroms, which lies well within the acceptable bounds prescribed by the CCDC blind structure prediction competition. The zero-temperature landscape, reduced landscape equilibrated at 100 K and 1 atm pressure using NPT molecular dynamics, and overlay of the predicted and experimental structure are all shown in Figure 1 of the accompanying figure file. We were able to argue that the accuracy of the intermolecular potential, here the SAPT(DFT)-based potential, is determinative of the crystal structure, while generic force fields can be used both for the structure generation step and for the intramolecular component. Using a generic representation of the intramolecular potential can be employed without significantly compromising that accuracy, even when some functional groups rotate easily within the crystalline environment. This force field approach simplifies the CSP workflow, without significantly compromising the accuracy of the prediction. A publication on this has appeared in *Crystal Growth & Design*.

(b) Machine learning is having a profound effect on our ability to solve complex quantum mechanical problems such as determining the electronic structure of a chemical system. Once trained on a handful of explicit, yet computationally costly, electronic structure calculation, machine learning (ML) models are able to predict the outcomes of these expensive calculations at the same accuracy as the level of theory used in training but several orders of magnitude faster using far fewer computational resources. Among the popular electronic structure methods, density functional theory remains one of the most widely used. Our previous work has focused on

RPPR Final Report as of 12-Jul-2022

learning the bijection between external potentials and ground-state densities established in the Hohenberg-Kohn (HK) theorem of density functional theory. This map, known as the ML-HK map, has allowed us to create simple, orbital-free machine-learned functionals capable of predicting the results of full Kohn-Sham density functional theory calculations, coupled-cluster calculations, excited-state energies from time-dependent density functional theory, and of generating highly accurate molecular dynamics trajectories. While this previous work constitutes a proof of concept of the utility of the ML-HK map, the ML model employed failed to incorporate an important property of the density, namely, its rotational equivariance. Building equivariance into an ML model would ensure that a system need not be constantly reoriented to a fixed grid, reduce the amount of training data needed, and simplify the use of ML-HK maps in generating molecular dynamics trajectories from trained ML models. The aim of this project has been to construct a deep message-passing neural network for learning the ML-HK map from a set of individual equivariant operations (see Figure 2). Beginning with an atom-centered basis-set representation of the density composed of radial Gaussians and spherical harmonics, a series of self- and pair-mixing layers and linear and nonlinear activations is combined to yield a robust approach for determining the basis-set expansion coefficients. The architecture is adapted from a model originally designed to predict molecular wave functions and is further imbued features important for use in molecular dynamics simulations. These include a protocol for learning energies and forces along with the density and a flexible basis that follows the system as it evolves in time. The new ML model will be tested on existing data sets on small molecules from our previous work, and a new set of “experiments” is proposed to sample two- and four-dimensional conformational free-energy landscapes flexible molecules for which crystal structures will be subsequently predicted.

(c) We demonstrated that the proximity of selected atoms and molecules to the zeros of selected order parameters. These zeroes occur on a set of crystallographic planes, and we have derived a set of equations for determining these. In addition, the orientation of molecules can be determined by diagonalizing moment of inertia tensor of a reference molecule. Rotations are then determined from the principal axes thus obtained, and these give yet another set of equations. The full set of equations we have obtained is enough to restrict an initial pool of structures to be quite small. These are then subject to a genetic algorithm that optimizes the structures by ensuring maximal adherence to the zeroes of the order parameters. Preliminary results on two polymorphs of aspirin and one of paracetamol all yield RMSD20 values of predicted structures within 0.5 Å of the experimental structures. A manuscript on this work is in preparation. Our predictions of these structures overlaid with the experimental structures are shown in Figure 3.

(d) Predictions of cocrystals of urea and resorcinol make use of the Shannon information entropy, constructed from an intermolecular orientational spatial distribution function, to drive a search for crystal structures via enhanced molecular dynamics was shown to be an efficient way to map out a landscape of putative polymorphs. We showed that driven adiabatic free energy dynamics, a particular enhanced-sampling approach, combined with these entropy variables, can transform the stable phase into alternate polymorphs. Density functional theory calculations confirm that a structure obtained from the enhanced molecular dynamics is stable at pressures above 1 GPa. Figure 4 shows the energy-density plots from the finite-temperature crystal structure prediction protocol at different pressures. Figure 5 shows the density functional theory energy curves, which includes the enthalpy differences between the P21/c and P212121 and the P-1 and P212121 structures as an inset. An important conclusion from this work is that from a computational perspective, cocrystals are inherently more complex than monocystals owing to the richness of the effective $Z' = 2$ crystal structure landscape. We demonstrated the value of enhanced-sampling techniques in rapidly identifying polymorphs not easily found in a standard CSP search and generating transitions between different structures. The ability to generate structures readily with higher Z' values makes this method complementary to a standard CSP search with the minimal number of molecules in the asymmetric unit cell. Results were published in *Crystal Growth & Design*.

(e) In benzamide II polytypes, centric H-bonded amide dimers are found to be the favored H-bonding motif. The stacking faults are also associated with conformational changes of phenyl rings, in agreement with observed diffraction intensities. A characteristic average structure containing the pertinent one-dimensional motifs is derived. Figure 6 shows PXRD patterns associated with various disorder models derived from our approach together with the experimental pattern. In the end, taking a linear combination of different disorder models yields a model exhibiting excellent agreement with the experimental pattern. Results were published in *Crystal Growth & Design*.

(f) The Szalewicz and Tuckerman groups have employed their joint protocol in a collaborative effort to predict crystal structures of targets in the 7th CCDC blind structure prediction competition. This collaboration allowed us to predict structures for a majority of the issued targets (phase 1) and to predict the rankings of lists supplied in phase 2. As the results have not yet been released, we will defer a more detailed report on the results until the interim report due at the end of the summer.

RPPR Final Report as of 12-Jul-2022

Training Opportunities: The project has provided the opportunity for two half-time postdocs and a graduate student to receive training in a number of areas of theoretical and computational physics and chemistry and large-scale software development. These include crystal structure prediction, theoretical statistical mechanics, molecular dynamics based enhanced sampling calculations, free energy reconstruction techniques, machine learning algorithms, and large-scale software development in the form of a novel parallel pipeline for free-energy based crystal structure prediction. All of these are skills that can also be used outside of the realm of theoretical chemistry. For the last four years (including this year), Tuckerman has been a participant in NYU's GSTEM program, which aims to provide summer research opportunities for female high school students in the greater New York area. He will continue to engage with this program in the last cycle, allowing GSTEM students to become involved in projects involving crystal structure prediction, machine learning, and code development. In 2020, Tuckerman co-organized a symposium at the ACS National meeting entitled, "Molecular Crystal Polymorphism: How, When & Why Molecules Pack in the Solid State", which provided opportunities for PIs, postdocs, and students to present and learn some of the latest approaches in the area of predicting crystal polymorphism. In early 2021, Tuckerman gave a pedagogical lecture entitled, "Processing and upscaling enhanced sampling simulations with machine learning" at the Virtual Winter School on Computational Chemistry, run out of the Australian National University in Canberra. This lecture provided students with an introduction to machine learning techniques and showed their applicability in statistical mechanics and structure prediction tasks. Tuckerman also participates in the New York City based BioBus program, giving outreach lectures to underserved high-school students, the last such lectures having been given in October of 2020 and in May of 2021. These lectures generally focus on computational chemistry ("Chemistry in a virtual laboratory") but use examples of crystal structure prediction, in part, to illustrate the concepts. In 2022, Tuckerman will deliver two pedagogical lectures, including a Webinar on materials design and a pedagogical lecture at the Erice Free Energy School on rare-event sampling and its synergistic combination with machine learning for graduate students.

RPPR Final Report

as of 12-Jul-2022

Results Dissemination: Papers published:

1. "Quantum chemical accuracy from density functional approximations via machine learning". M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K. -R. Mueller, K. Burke. *Nature Comm.* 11, 5223 (2020).
2. "Generating Cocrystal Polymorphs with Information Entropy Driven by Molecular Dynamics Based Enhanced Sampling". H. Song, L. Vogt-Maranto, R. Wiscons, A. J. Matzger, M. E. Tuckerman *J. Phys. Chem. Lett.* 11, 9751 (2020).
3. "Crystal Structure Prediction as a Tool for Identifying Components of Disordered Structures from Powder Diffraction: A Case Study of Benzamide II". E. J. Chan, A. G. Shtukenberg, M. E. Tuckerman, B. Kahr *Cryst. Growth & Design* 21, 5544 (2021).
4. "Crystal structure predictions for 4-amino-2,3,6-trinitrophenol using a tailor-made first-principles-based force field". M. P. Metz, M. Shahbaz, H. Song, L. Vogt-Maranto, M. E. Tuckerman, K. Szalewicz *Cryst. Growth & Design* 22, 1182 (2022).
5. "Topological crystal structure prediction". N. Galanakis, M. E. Tuckerman (in preparation).
6. "Equivariant neural networks for density-enhanced molecular dynamics". M. Bogojeski, L. Vogt-Maranto, K. -R. Müller, K. Burke, M. E. Tuckerman (in preparation).
7. "Report on the seventh blind test of organic molecular crystal structure prediction". (in preparation).

Invited Conference and Workshop Lectures:

1. "Topology, molecular simulation and machine learning as routes to exploring structure and phase behavior in atomic and molecular crystals." Fall 2022 ACS Meeting, Chicago, IL, Aug., 2022.
2. "Topology, molecular simulation and machine learning as routes to exploring structure and phase behavior in atomic and molecular crystals." American Conference on Theoretical Chemistry, Lake Tahoe, CA, July, 2022.
3. "Topology, molecular simulation and machine learning as routes to exploring structure and phase behavior in atomic and molecular crystals." MOLSIM2022 meeting in honor of the 80th birthday of Michael Klein, Erice, Italy, June, 2022.
4. "Topology, molecular simulations, and machine learning as routes to structure and phase behavior in atomic and molecular crystals." ACS's Mid-Atlantic Regional Meeting symposium: Application of Crystal Engineering, Technical College of New Jersey, Ewing, NJ, June, 2022.
5. "Machine learning the theorems of density functional theory." McGill Molecular Sciences Mini-Meeting, McGill University, Montréal, Québec, Canada, June, 2022.
6. "Using machine learning to process high-dimensional free energy surfaces into low-dimensional paths." CECAM workshop: Recent Advances in Machine Learning Accelerated Molecular Dynamics, Pisa, Italy, March, 2022.
7. "Molecular simulation and machine learning as routes to exploring structure and phase behavior in atomic and molecular crystals." ACS Pacificchem Conference, Honolulu, HI, Dec., 2021.
8. "Topological crystal structure prediction." ACS Pacificchem Conference, Honolulu, HI, Dec., 2021.
9. "Using machine learning to process high-dimensional free energy surfaces into low-dimensional paths." RARE Meeting, Coorg India, Dec., 2021.
10. "Machine learning the theorems of density functional theory." Telluride Scientific Research Center (TSRC) workshop on Nonequilibrium Phenomena, Nonadiabatic Dynamics, and Spectroscopy, Telluride, CO, July, 2021.
11. "Topology, molecular simulation and machine learning as routes to exploring structure and phase behavior in atomic and molecular crystals." Physical Chemistry Division of the ACS National Meeting, March, 2021.

RPPR Final Report as of 12-Jul-2022

12. "Processing and upscaling enhanced sampling simulations with machine learning." Scaling Cascades in Complex Systems conference, Freie Universität, Berlin, Feb., 2021.

(11 invited lectures in 2020 were postponed due to COVID).

13. "Machine learning of force fields for organic molecular crystals." Flatiron workshop: Machine Learning for Organic Force Fields, New York, NY, Dec., 2019.

14. "Combining statistical mechanics and machine learning in the prediction of structure and phase behavior in atomic and molecular crystals." CECAM workshop on Molecular Materials: Simulation at the Turn of the Decade – Celebrating 50 years of CECAM, Lausanne, Switzerland, Sept., 2019.

15. "Combining enhanced sampling with machine learning in the generation of high-dimensional free energy landscapes of complex molecular systems." Division of Computer in Chemistry, ACS National Meeting, San Diego, CA, Aug., 2019.

16. "Topology, molecular simulation, and machine learning as routes to exploring structure and phase behavior in atomic and molecular crystals." Current Trends in Theoretical Chemistry conference at the Universidad San Francisco, Quito, Ecuador, July, 2019.

17. "Machine learning of electron densities, energies, and free energies: Bypassing the Schrödinger equation and costly phase-space integrals." ACS Spring National Meeting, Orlando, FL, April, 2019.

Invited Departmental Colloquia:

1. "Topology, molecular simulation, and machine learning as routes to prediction of structure and phase behavior in molecular and atomic crystals." Department of Chemistry, City University of New York, New York, NY, March, 2022.

2. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals." Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, Canada, Nov., 2021.

3. "Path integral based enhanced sampling on machine-learned potential energy surface for correct calculation of quantum free energy landscapes". VISTA seminar, Department of Chemistry, University of Buffalo, April, 2021.

4. "Topological, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals". Department of Chemistry, University of Oregon, April, 2021.

5. "Topological, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals". Department of Chemistry, Fritz-Haber Institute, Jan., 2021.

6. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals". Department of Chemistry, University of Cambridge, Nov., 2020.

7. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals". Department of Chemistry, Princeton University, Princeton, NJ, Sept., 2020.

8. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals". Department of Chemistry, University of Toronto, Toronto, Canada, Aug., 2020.

9. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals." ICTP, Trieste, Italy, June, 2020.

10. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals." Department of Chemistry, Fordham University, Bronx, NY, Feb., 2020.

RPPR Final Report as of 12-Jul-2022

11. "A combined statistical mechanics/machine-learning approach to the prediction of structure and phase behavior in atomic and molecular crystals." Invited computational science seminar, University of Pennsylvania, Philadelphia, PA, Nov., 2019.
12. "Topology, molecular simulation, and machine learning as routes to the prediction of structure and phase behavior in atomic and molecular crystals." SISSA, Trieste, Italy, July, 2019.
13. "A combined statistical mechanics/machine-learning approach to the prediction of structure and phase behavior in atomic and molecular crystals." Department of Chemistry, University of Stony Brook, Stony Brook, NY, April, 2019.
14. "A combined statistical mechanics/machine-learning approach to the prediction of structure and phase behavior in atomic and molecular crystals." IIT Kanpur, India, March, 2019.

Honors and Awards: Institute Lectureship, Indian Institute of Technology Kanpur, March, 2019.

Elected to Membership in Sigma Xi Society, August, 2020.

Elected Fellow of the American Association for the Advancement of Science, January, 2022.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)

Participant: Eric J. Chan

Person Months Worked: 7.00

Funding Support:

Project Contribution:

National Academy Member: N

Participant Type: Graduate Student (research assistant)

Participant: Hongxing Song

Person Months Worked: 2.00

Funding Support:

Project Contribution:

National Academy Member: N

Participant Type: PD/PI

Participant: Mark E. Tuckerman

Person Months Worked: 1.00

Funding Support:

Project Contribution:

National Academy Member: N

ARTICLES:

RPPR Final Report as of 12-Jul-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: The Journal of Physical Chemistry Letters

Publication Identifier Type: DOI

Publication Identifier: 10.1021/acs.jpcllett.0c02647

Volume: 11

Issue: 22

First Page #: 9751

Date Submitted: 7/6/22 12:00AM

Date Published: 11/1/20 4:00AM

Publication Location: New York, NY

Article Title: Generating Cocrystal Polymorphs with Information Entropy Driven by Molecular Dynamics-Based Enhanced Sampling

Authors: Hongxing Song, Leslie Vogt-Maranto, Ren Wiscons, Adam J. Matzger, Mark E. Tuckerman

Keywords: ChemistryScience & Technology - Other TopicsMaterials SciencePhysics

Abstract: Predicting structures of organic molecular cocrystals is a challenging task when considering the immense number of possible intermolecular orientations. Use of the Shannon information entropy, constructed from an intermolecular orientational spatial distribution function, to drive a search for crystal structures via enhanced molecular dynamics can be an efficient way to map out a landscape of putative polymorphs. Here, the Shannon entropy is used to generate a set of collective variables for differentiating polymorphs of a 1:1 cocrystal of resorcinol and urea. We show that driven adiabatic free energy dynamics, a particular enhanced-sampling approach, combined with these entropy variables, can transform the stable phase into alternate polymorphs. Density functional theory calculations confirm that a structure obtained from the enhanced molecular dynamics is stable at pressures above 1 GPa. We thus show that enhanced sampling should be considered an integral component of crystal structure

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published

Journal: Crystal Growth & Design

Publication Identifier Type: DOI

Publication Identifier: 10.1021/acs.cgd.1c00261

Volume: 21

Issue: 10

First Page #: 5544

Date Submitted: 7/6/22 12:00AM

Date Published: 8/1/21 4:00AM

Publication Location: New York, NY

Article Title: Crystal Structure Prediction as a Tool for Identifying Components of Disordered Structures from Powder Diffraction: A Case Study of Benzamide II

Authors: Eric J. Chan, Alexander G. Shtukenberg, Mark E. Tuckerman, Bart Kahr

Keywords: Chemistry Crystallography Materials Science

Abstract: Theoretical crystal structure prediction (CSP) calculations of molecular crystals often generate structures representative of the components associated with disorder, stacking faults, and polytypism. The powder X-ray diffraction (PXRD) pattern of benzamide II contains diffuse scattering consistent with a disordered structure. Previous attempts to elucidate its ordered structure by PXRD and CSP were incomplete because the average structure is actually a complex superposition of multiple polytypes, thus making structure determination particularly challenging. To address this problem, an approach to modeling such complex structures from PXRD data and CSP in combination with molecular dynamics simulations is proposed. Using this approach, the potential building blocks for polytypes possessing a twodimensional stacking fault structure are identified. For benzamide II polytypes, centric H-bonded amide dimers are found to be the preferred H-bonding motif. The stacking faults are also associated

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report
as of 12-Jul-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Crystal Growth & Design
Publication Identifier Type: DOI **Publication Identifier:** 10.1021/acs.cgd.1c01117
Volume: 22 **Issue:** 2 **First Page #:** 1182
Date Submitted: 7/6/22 12:00AM **Date Published:** 1/1/22 5:00AM
Publication Location: New York, NY

Article Title: Crystal Structure Predictions for 4-Amino-2,3,6-trinitrophenol Using a Tailor-Made First-Principles-Based Force Field

Authors: Michael P. Metz, Muhammad Shahbaz, Hongxing Song, Leslie Vogt-Maranto, Mark E. Tuckerman, Krzy

Keywords: Chemistry Science & Technology Materials Science Physics

Abstract: Predictions of crystal structures from first-principles electronic structure calculations and molecular simulations have been performed for an energetic molecule, 4-amino-2,3,6-trinitrophenol. This physics-based approach consists of a series of steps. First, a tailor-made two-body potential energy surface (PES) was constructed with recently developed software, autoPES, using symmetry-adapted perturbation theory based on a density-functional theory description of monomers [SAPT(DFT)]. The fitting procedure ensures asymptotic correctness of the PES by employing a rigorous asymptotic multipole expansion, which seamlessly integrates with SAPT(DFT) interaction energies. Next, crystal structure prediction (CSP) was performed by generating possible crystal structures with rigid molecules, minimizing these structures using the SAPT(DFT) force field, and running isothermal-isobaric molecular dynamics (MD) simulations with flexible molecules based on the tailor-made SAPT (DFT) intermolecular force

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

DISSERTATIONS:

Publication Type: Thesis or Dissertation

Institution: New York University

Date Received: 06-Jul-2022

Completion Date: 1/31/22 4:32PM

Title: Polymorph prediction for molecular crystals by combining random search and molecular dynamics

Authors: Hongxing Song

Acknowledged Federal Support: N

Partners

I certify that the information in the report is complete and accurate:

Signature: Mark E. Tuckerman

Signature Date: 7/6/22 12:07PM

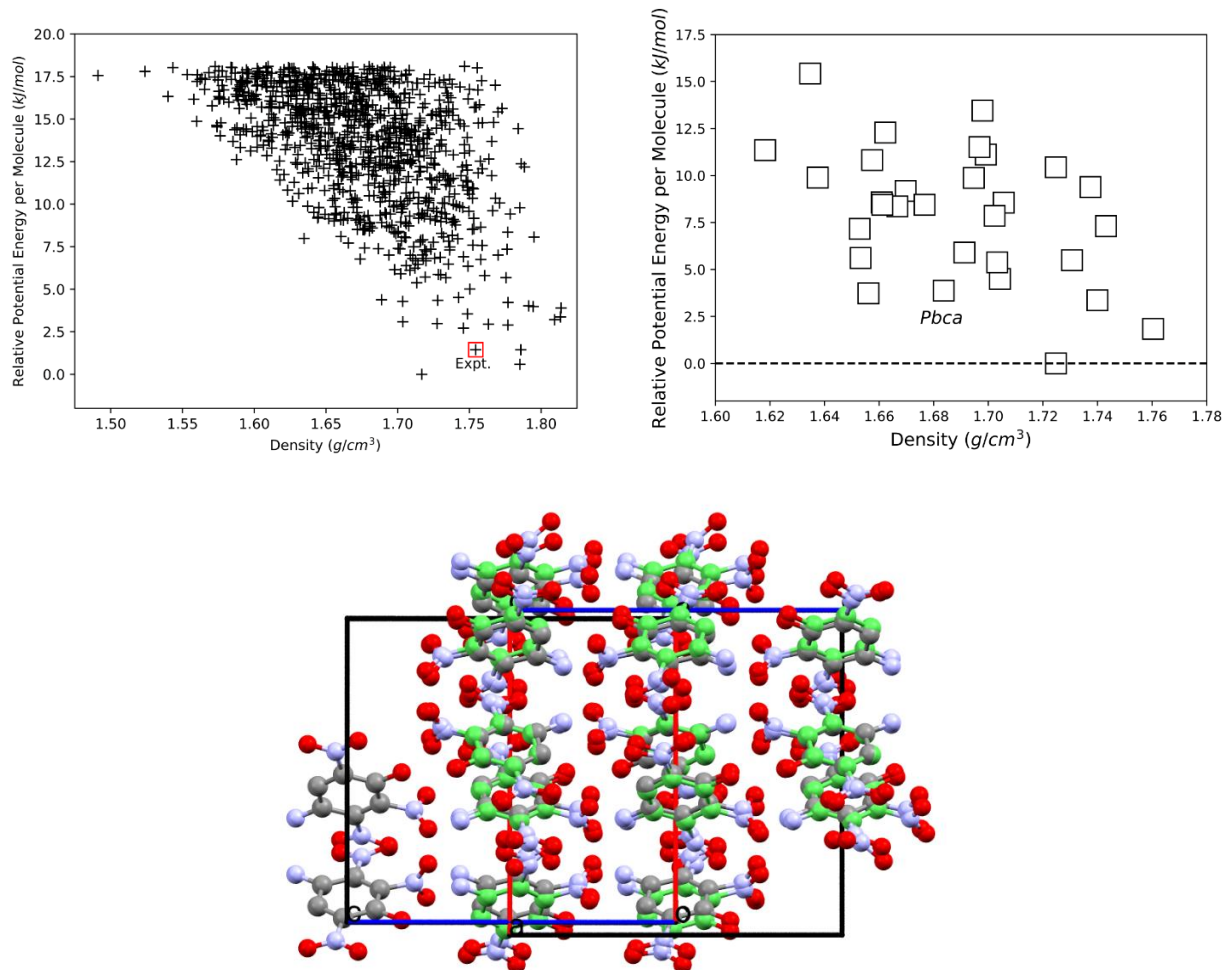


Figure 1: (Top left) Zero-temperature CSP landscape for 4-amino-2,3,6-trinitrophenol crystal. The experimental structure is indicated. (Top right) Reduced CSP landscape after equilibration via NPT molecular dynamics at 100 K and 1 atm pressure. The experimental structure (*Pbca*) is indicated. (Bottom) Overlay of the predicted (green spheres) and experimental (colored by element) structures.

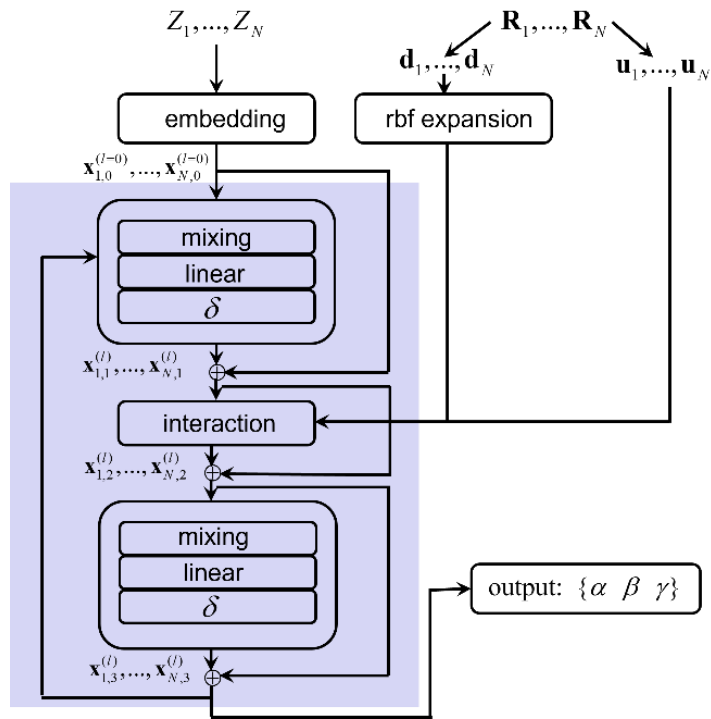


Figure 2: Message-passing equivariant neural network architecture for learning the ML-HK map. The density is decomposed into a set of atom-centered radial basis functions and spherical harmonics, which are mixed with each other in the “mixing” layers. Activation layers ensure nonlinearity, and ultimately, a set of expansion coefficients (β, γ) and parameters of the radial functions (α) are output.

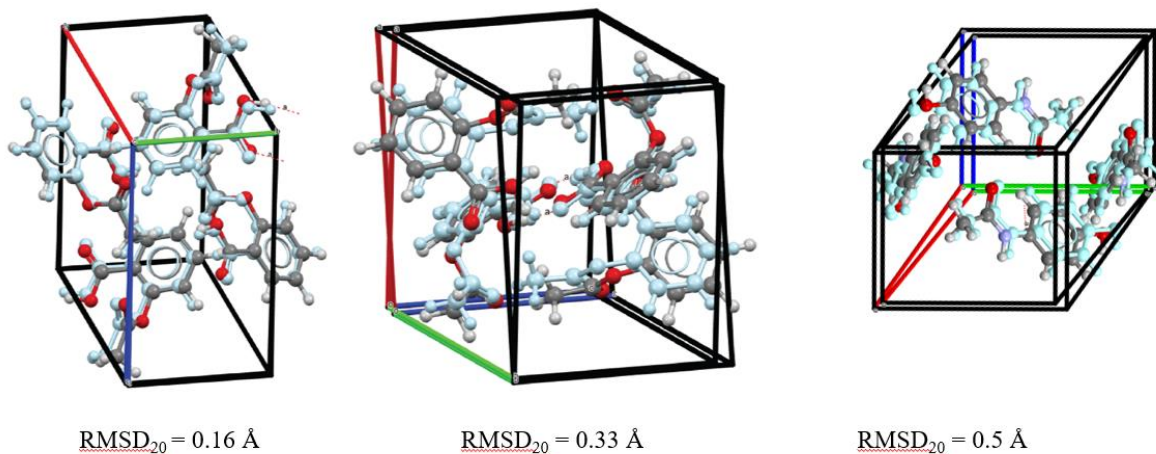


Figure 3: Topological crystal structure predictions for two polymorphs of aspirin (left and middle structures) and paracetamol (right). Colored structures are predicted, light blue are experimental structures. RMSD_{20} values are shown below each structure.

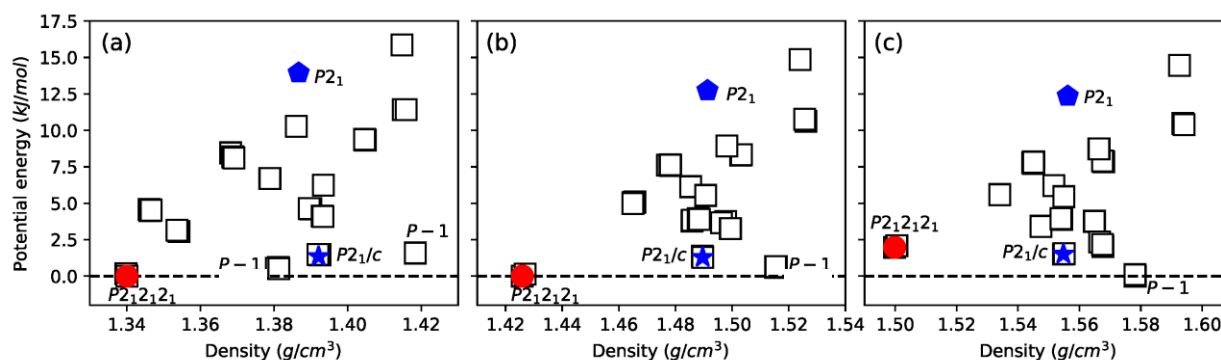


Figure 4: Potential energy vs. density plots for the top 30 structures (black open squares) from standard CSP. The potential energy (given per f.u.) and the density are both averaged from flexible cell NPT MD at $T=300$ K but different pressures: (a) $P=1$ bar, (b) $P=1$ GPa, and (c) $P=2$ GPa. The experimental structure (red solid circle) and the metastable polymorphs predicted in the d-AFD trajectories (blue solid star and pentagon) are also shown, including their space groups.

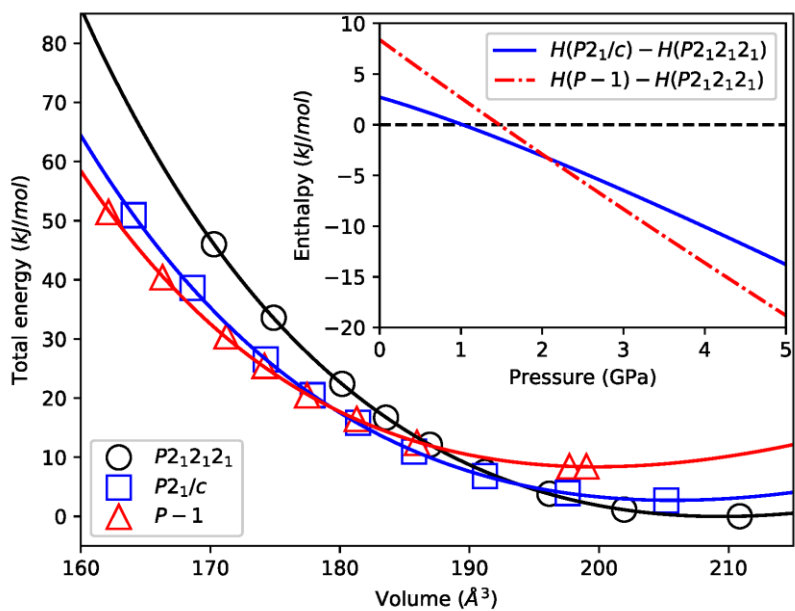


Figure 5: Relative total energy calculations by DFT-D vs. cell volume for three low energy structures. The E-V curves are fitted to the Murnaghan equation of states model, which is then used to calculate the enthalpy (per f.u.) as a function of pressure (see inset). The $P2_1/c$ structure is predicted to become more stable than the experimentally observed $P2_12_12_1$ polymorph (zero reference of energy and enthalpy) at pressures between 1.02 GPa and 2.07 GPa.

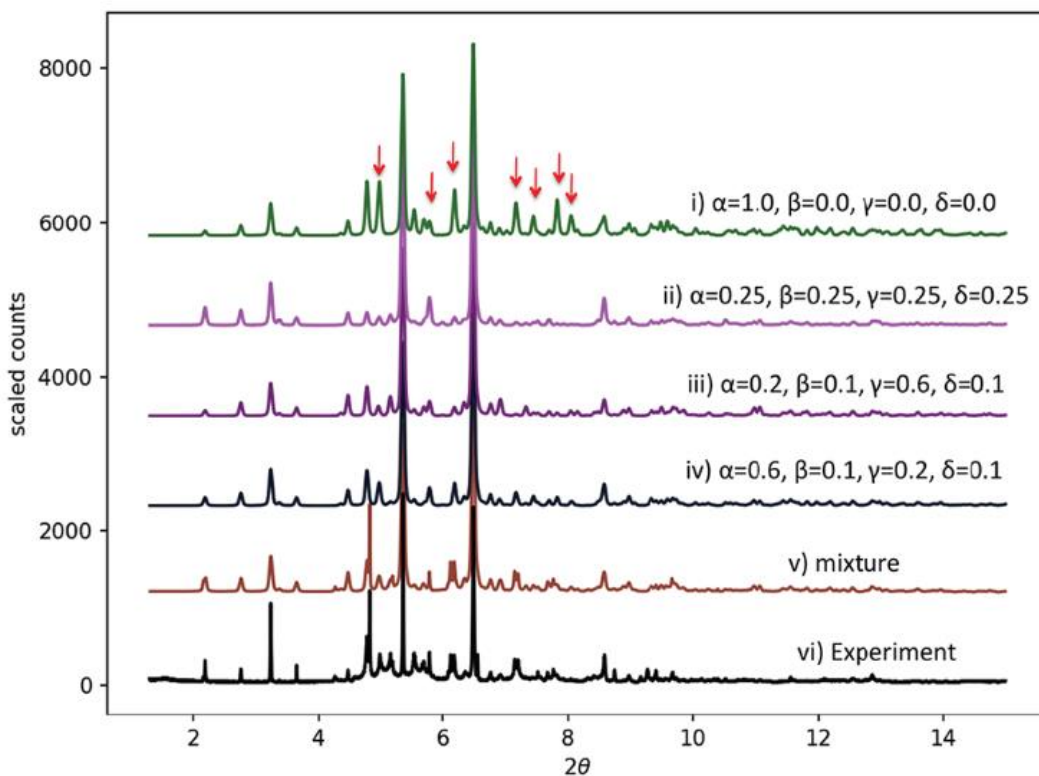


Figure 6: PXRD patterns (i – iv) calculated from the Bragg structure for benzamide II. The site occupation labels $\alpha, \beta, \gamma, \delta$ represent respective polytype sequences ABABABAB, BABABABA, DCDCDCDC, and CDCDCDCD, depending on whether the phenyl ring faces the oxygen or nitrogen atom. A and B designate these two orientations in one of the benzamide II molecules, C and D designate these two orientations in the other. Structure v is a linear combination of the structure i – iv. The red arrows indicate peaks that would require considerable attenuation if the mixture of the crystal phases were considered as not being fully ordered. When a disordered Bragg model is used, a favorable attenuation of such peaks emerges more naturally. The respective Pearson correlation coefficients for i – v are 0.81, 0.77, 0.78, 0.79, and 0.83.