Published in final edited form as:

Chem Mater. 2018 May 22; 30(10): 3330–3337. doi:10.1021/acs.chemmater.8b00679.

Data Mining for Parameters Affecting Polymorph Selection in Contorted Hexabenzocoronene Derivatives

Anna M. Hiszpanski^{†,‡}, Carmeline J. Dsilva[†], Ioannis G. Kevrekidis^{†,§}, and Yueh-Lin Loo^{*,†,||}

[†]Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, United States

[‡]Materials Science Division, Lawrence Livermore National Laboratory, Livermore, California 94550, United States

[§]Department of Applied Mathematics and Statistics, John Hopkins University, Baltimore, Maryland 21218, United States

^{||}Andlinger Center for Energy and the Environment, Princeton University, Princeton, New Jersey 08544, United States

Abstract

The macroscopic properties of molecular materials can be drastically influenced by their solid-state packing arrangements, of which there can be many (e.g., polymorphism). Strategies to controllably and predictively access select polymorphs are thus highly desired, but computationally predicting the conditions necessary to access a given polymorph is challenging with the current state of the art. Using derivatives of contorted hexabenzocoronene, cHBC, we employed data mining, rather than first-principles approaches, to find relationships between the crystallizing molecule, postdeposition solvent-vapor annealing conditions that induce polymorphic transformation, and the resulting polymorphs. This analysis yields a correlative function that can be used to successfully predict the appearance of either one of two polymorphs in thin films of cHBC derivatives. Within the postdeposition processing phase space of cHBC derivatives, we have demonstrated an approach to generate guidelines to select crystallization conditions to bias polymorph access. We believe this approach can be applied more broadly to accelerate the predictions of processing conditions to access desired molecular polymorphs, making progress toward one of the grand challenges identified by the Materials Genome Initiative.

Graphical Abstract

*Corresponding Author lloo@princeton.edu.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.chemma-ter.8b00679](https://doi.org/10.1021/acs.chemma-ter.8b00679). Experimental details; diffraction traces showing stability of 12F-cHBC film structure with time; table of solvents' Hansen solubility parameters, molar volume, and vapor pressure at 20 °C; table of solvents' classification by chemical synthons; table of molecule—solvent combination and resulting polymorph I fraction (PDF)

Table of Hansen solubility parameters of the solvents, their molar volume, and vapor pressure at 20 °C and classification by chemical synthons; table of molecule—solvent combination and resulting polymorph I fraction (XLSX)

The authors declare no competing financial interest.

INTRODUCTION

Polymorphism, or the ability of molecules to adopt more than one crystal structure, is a topic of significant interest for functional materials across a breadth of applications, including in pharmaceuticals, pigments, explosives, food additives, and organic electronics. Because the bulk properties of molecular compounds can vary significantly with different crystal structures,^{1,2} ultimately impacting their utility, strategies to more fully explore their polymorphic phase space to controllably and predictively access specific polymorphs are highly desired. Despite the intense interest in polymorphism, polymorphic discovery has largely been serendipitous, driven by trial-and-error experimentation due to challenges in accurately modeling the variety of weak and nondirectional intermolecular forces involved in molecular crystal growth. Further complicating computational polymorphic predictions is the observation that the energy differences between molecular polymorphs are often only a few kcal/mol, creating a rich energy landscape with many shallow local energy minima.^{3,4} Commonly, the crystal structure that is accessed is one that is kinetically favored given the specific processing conditions, as opposed to the thermodynamically favored polymorph.^{4,5}

Computationally capturing the structural rearrangement during polymorphic transformation remains challenging. An alternative and potentially simpler approach is to identify predictive correlations between crystallization conditions and resulting polymorphs using data mining approaches.^{6–10} Indeed, data mining to identify chemistry-processing-structure-function relationships is its own developing subdiscipline. Referred to as materials informatics or cheminformatics, this effort has been highlighted by the Materials Genome Initiative as a means to accelerate materials discovery and development.^{11–14} For example, data mining approaches have recently been used to identify correlations between molecules and their optoelectronic properties. These correlations have in turn been used to predict the optoelectronic properties, like the highest occupied and lowest unoccupied molecular orbital (HOMO and LUMO) energy levels, of new molecules without additional costly quantum chemical calculations and prior to synthesis.^{15,16}

Despite the demonstrated utility of data mining approaches, they have not been widely applied outside the pharmaceutical community to identify relationships between crystallization conditions and molecular polymorphs. While the pharmaceutical industry routinely uses automated robotic systems to explore varied crystallization conditions to yield

large single-crystal data sets from which correlations can be extracted,^{10,17–19} such automated single-crystal polymorph screening techniques are less explored in the broader materials community and less applicable to polycrystalline thin films. In polycrystalline thin-film formats, different molecular polymorphs have been accessed by changing a number of experimental parameters during film formation, including the deposition technique,²⁰ substrate temperature,^{21–24} choice of chemical modification of the substrate surface,^{23–26} film thickness,^{24,25,27–30} confinement effects,^{27,31,32} and the solvent choice,^{27,28} and the application of postdeposition processing techniques, like thermal annealing and solvent-vapor annealing.^{33–38} This rich experimental space is therefore rife with opportunities to apply data mining techniques for extracting more general and predictive relationships that are currently lacking between molecules of interest, crystallization conditions, and the resulting polymorph.

Here, we demonstrate the utility of data mining techniques in predicting which one of two molecular polymorphs in thin polycrystalline films of contorted hexabenzocoronene (cHBC) derivatives is accessed given the postdeposition solvent-vapor annealing conditions to which they are exposed. Due to their limited solubility, cHBC derivatives are thermally evaporated and form amorphous thin films. We previously showed that, by depositing an amorphous film of cHBC and subsequently inducing crystallization using postdeposition processing techniques like solvent-vapor annealing, we gain more control over the structural evolution of films and broader access to the crystal energy landscape and other polymorphs.^{34,39} Solvent-vapor annealing cHBC with hexanes induces crystallization of a monoclinic crystal structure with $P2_1/c$ space group symmetry that we refer to as polymorph I,³⁹ whereas solvent-vapor annealing with tetrahydrofuran (THF) induces crystallization of a second, likely triclinic, crystal structure that we refer to as polymorph II.³⁴ Recently, we demonstrated that this tunability in polymorphic selection can be extended to chemically modified derivatives of cHBC as well. Amorphous thin films of four fluorinated derivatives of cHBC decorated with 8, 12, 16, or 20 fluorine atoms at the molecule's periphery also adopt one of two crystal structures comparable to cHBC's polymorph I and polymorph II depending on the postdeposition processing conditions.³³

Given that (1) we can access two different polymorphs of cHBC and its derivatives by the choice of solvent used during solvent-vapor annealing and that (2) these two polymorphs appear to be common across a set of four fluorinated derivatives of cHBC, we have the opportunity to determine correlations between molecular structure and solvent choice that influence polymorphic selection. Such correlations can help guide the design of next-generation derivatives and the development of processing routes to access the prespecified crystal packing.

To generate a large data set that can be mined, we solvent-vapor anneal thin films of cHBC and its four fluorinated derivatives with twenty-seven different solvents. Solvent-vapor annealing is carried out until crystallization is complete; i.e., longer solvent vapor exposure does not alter the film structure. Annealing cHBC and its derivatives with some of the solvents yields polymorph I, while with others it yields polymorph II and yet with other solvents it yields a mixture of polymorphs I and II. As the resulting film structures are kinetically determined and do not represent thermodynamic equilibrium, Gibbs' phase rule

does not apply. Applying a data mining approach to our large experimental data training set, we developed a correlation between the fraction of polymorph I obtained upon solvent-vapor annealing with the molecular properties of the solvent and the cHBC derivative. Using a test data set separate from our training set, we further demonstrate how this correlative relationship can be used to predict the expected polymorph given the choices of solvent and cHBC derivative undergoing crystallization.

RESULTS AND DISCUSSION

Starting with amorphous thin films of cHBC and its four fluorinated derivatives, whose chemical structures are shown in Figure 1, we subjected the films to solvent-vapor annealing for 4 h to induce complete crystallization (see Supporting Information for details). The chemical structures of the twenty-seven solvents investigated, which were chosen to be as chemically diverse as possible, are also shown in Figure 1. As we previously reported, cHBC and 8F-, 12F-, 16F-, and 20F-cHBC have two distinct polymorphs.^{33,34} The presence of polymorph I is most readily identified by the position of the primary reflection, corresponding to the (100) plane, which occurs at approximately $q = 0.5 \text{ \AA}^{-1}$ in their X-ray diffraction traces. The precise placement of this reflection varies between 0.48 and $0.53 \text{ \AA}^{-1}$ depending on the cHBC derivative. The presence of polymorph II is most clearly identified by the position of its primary peak, which occurs at approximately $q = 0.7 \text{ \AA}^{-1}$ in the X-ray diffraction patterns of all the cHBC derivatives after they had been subjected to THF-vapor annealing. The precise placement of this peak varies between 0.65 and $0.70 \text{ \AA}^{-1}$ depending on the cHBC derivative.

Figure 2 shows optical micrographs and the corresponding one-dimensional X-ray diffraction traces (obtained by azimuthally integrating 2D-GIXD images) of cHBC thin films that were crystallized on exposure to toluene and benzene vapor. The fraction of polymorph I in each film can be approximated from the X-ray diffraction traces by dividing the integrated intensity of the primary reflection of polymorph I at $q = 0.5 \text{ \AA}^{-1}$ by the sum of the integrated intensities of the reflections at $q = 0.5 \text{ \AA}^{-1}$ and the primary reflection of polymorph II at $q = 0.7 \text{ \AA}^{-1}$. Because the integrated intensities of the reflections at $q = 0.5 \text{ \AA}^{-1}$ and $q = 0.7 \text{ \AA}^{-1}$ are comparable for completely crystalline thin films that adopt solely polymorph I and polymorph II, respectively, it is unnecessary to normalize the integrated intensities of these two reflections when estimating the fraction of the film adopting each polymorph. This exercise also assumes that the films are completely crystalline and that they adopt polymorphs I or II or a coexistence of both; extended postdeposition solvent-vapor annealing of the films confirms this assumption to be valid. Though we solvent-vapor annealed films for 4 h, crystallization is typically complete within 30 min, and longer solvent vapor exposure does not alter the film structure. Using this methodology, we estimated the fractions of polymorph I in the toluene- and benzene-vapor annealed cHBC films to be 0.93 and 0.10, respectively, with the remainder of the films adopting polymorph II.

In studying numerous optical micrographs of films having a coexistence of the two polymorphs, we found that domains of polymorph II appear darker and grainier compared to those of polymorph I; these regions are therefore distinguishable in optical micrographs so

long as the domains are larger than ca. 10 μm . The optical micrographs of the toluene- and benzene-vapor annealed cHBC films show a mixture of polymorphs I and II; analysis of these images independently yielded fractions of polymorph I that are comparable to those extracted from the X-ray diffraction traces, i.e., 0.94 and 0.21 for toluene- and benzene-vapor annealed films, respectively. The larger discrepancy between the two estimates for the benzene-vapor annealed film compared to the toluene-vapor annealed film may stem from the fact that the domains in the benzene-vapor annealed film are larger, so we sample fewer domains within a given optical microscope image or within a given X-ray diffraction footprint, effectively yielding poorer statistics for polymorph I that is present in the film.

Expanding our examination to the full series of fluorinated cHBC derivatives, Figure 3 shows the fraction of films of cHBC and 8F-, 12F-, 16F-, and 20F-cHBC that adopt polymorph I upon solvent-vapor annealing with five different solvents: hexanes (squares), trichloroethylene (circles), dichloromethane (triangles pointing up), chloroform (triangles pointing down), and tetrahydrofuran (diamonds). We see that annealing with hexanes vapor yields polymorph I for all derivatives, whereas annealing with THF vapor yields exclusively polymorph II across the series of compounds. Annealing with the other three solvents yields a varying amount of the two polymorphs depending on the extent of fluorination of the cHBC derivatives. In total, we have 199 data points of cHBC derivatives films solvent-vapor annealed with solvents and yielding polymorph I, polymorph II, or a mixture of the two polymorphs (data includes repeats of some cHBC derivative and solvent combinations; data are compiled in Table S1). This compilation of data points to the importance of both the solvent choice and molecule type in dictating the fraction of polymorph I present.

Molecule—solvent combinations that yielded purely polymorph I or purely polymorph II did so consistently across repeated trials, but for molecule—solvent combinations that yielded a mixture of the two polymorphs, the precise fraction varied between trials. To quantify the experimental variability in the fraction of polymorph I in such mixed polymorph films, we repeated several solvent-molecule combinations from which we estimated the absolute spread in the fraction of polymorph I to be 0.3. For example, when we solvent-vapor annealed 16F-cHBC with dichloromethane, the fully crystallized film comprised 0.68 polymorph I. In a second but analogous experiment to assess reproducibility, we estimated 0.81 polymorph I in the fully crystallized 16F-cHBC film. We believe the spread stems from variability in sample preparation and crystallization conditions. While the error bar in predicting the fraction of polymorph I is large for films that adopt a mixture of the two polymorphs, it does not negate the impact of this exercise. Since coexistence of multiple polymorphs is practically undesirable, the value in this data mining approach is in its ability to articulate processing conditions for accessing either of the polymorphs *in its entirety*.

To identify by data mining the molecular and processing parameters that influence polymorph selection, we must first choose parameters that characterize and differentiate the cHBC derivatives and solvents. Given that cHBC and 8F-, 12F-, 16F-, and 20F-cHBC differ in their molecular structure by only the number of fluorines decorating the periphery of the molecule, we used the number of fluorines as the simplest and sole variable to distinguish between the cHBC derivatives. Characterizing the solvents is more challenging since they

exhibit greater chemical diversity, and so, a greater number of parameters have to be considered. We chose to consider the solvents' Hansen solubility parameters (which characterize each solvent's polarity, dispersion quality, and hydrogen-bonding potential),⁴⁰ molar volume, vapor pressure, and chemical structure motifs. Similarities in Hansen parameters are commonly used to predict the solubility of polymers and molecules in a given solvent.⁴⁰ During solvent-vapor annealing, the solvent vapor plasticizes cHBC films, enabling the molecules to rearrange and crystallize.⁴¹ Thus, the size of the solvent molecule and the concentration of solvent in the vapor phase (quantified by vapor pressure at room temperature) may affect the kinetics of crystallization and are both important to consider. Table S2 contains the Hansen solubility parameters, molar volumes, and vapor pressures (at 20 °C) of the twenty-seven solvents we explored. The Hansen solubility parameters and molar volume were obtained from Hansen's handbook,⁴⁰ and the vapor pressures were extracted from material safety data sheets (MSDS). We also chose to categorize the solvents by their chemical synthons, noting if their chemical functionalities include aromatics, halogens, alkenes, alkynes, ethers, alcohols, thiols, or ketones, to determine if there is any correlation between chemical solvent families and the favored polymorph. Table S3 contains these classifications with 1 indicating the presence of the specific synthon in the solvent and 0 indicating its absence. Performing principal component analysis⁴² on the solvents' Hansen solubility parameters and chemical synthons, we found no significant correlations between these variables, indicating they are appropriate independent input parameters for data mining.

Our training data set does not exhibit any direct correlations with any one solvent characteristic for the tendency of cHBC and its fluorinated derivatives to preferentially adopt polymorph I. We thus performed multiple linear regression analysis on the data set to ascertain a relationship between the fraction of the thin film that adopted polymorph I (the dependent variable that varies between 0 and 1) and the solvent parameters and number of fluorines on cHBC derivatives (the independent variables). As described before, our goal is to identify the conditions under which accessing either of the two polymorphs across the entire film is most likely. Thus, we specifically chose to use a logistic regression model,^{42,43} which is commonly used when the dependent variable is categorical (i.e., pass or fail, win or lose, 0 or 1, or as desired in our case, polymorph I or polymorph II). The logistic function that is the basis of this model is an S-shape curve (sigmoid curve) with the following form:^{42,43}

$$[\text{fraction polymorph I}] = 1/(1 + e^{(-y)}) \quad (1)$$

where e is Euler's number and y is a linear relationship comprising all the independent variables under study.

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n \quad (2)$$

Here, x_i represents the independent variables and β_j represents their weighting coefficients. We refer to each term in this equation as a parameter, and any function that describes our system can contain up to 15 parameters given that we have 13 independent variables to describe the solvents (i.e., three Hansen solubility parameters, molar volume, vapor pressure, and eight different synthons), one to describe differences between the cHBC derivative (i.e., the number of fluorines), and one constant for linear regression. We used the Akaike information criterion, AIC,⁴⁴ to quantify the quality of fit to our data. AIC imposes a penalty for functions that require more parameters.⁴⁴ In this framework, a lower AIC indicates a function that more accurately describes our data set.⁴⁴ Of the 199 experimental data points, we also randomly withheld 25 data points from this multiple linear regression analysis to later validate our model.

Given that we did not know which parameters or how many parameters were required to describe the correlation between the independent and dependent variables in our data set, we first tried fitting the training data set (199 minus 25 data points) to all possible functions having up to 15 parameters, requiring that the number of fluorines always be included as a parameter in the function, as this is the only variable that differentiates the cHBC derivatives. Figure 4a contains the resulting AICs of these functions, with each data point representing an attempted function. The vertical spread in AIC reveals the spread in the quality of fit for all possible functions having the specified number of parameters. Looking specifically at the functions that yield the lowest AIC (identified with the red circle), Figure 4a indicates that the quality of fit levels off with functions having eight or more parameters. Thus, having more than eight parameters in any possible function provides little gain in improving the quality of fit to our data.

Of the functions that yield the lowest AIC as a function of the number of parameters, we investigated which of the 14 variables are involved in these functions to identify which physical or chemical properties of the solvents may be most relevant for polymorphic selection. Figure 4b provides a color plot with the variables on the y-axis and the number of parameters included in the functions on the x-axis. White indicates that the parameter is not included in the function in question; red indicates that the parameter is positively correlated with the appearance of polymorph I (i.e., β is positive), and blue indicates that the parameter is negatively correlated with the appearance of the polymorph I (i.e., β is negative) and instead positively correlated with the appearance of polymorph II. From Figure 4b, we see that the number of fluorines is positively correlated with the appearance of polymorph I in all the functions examined. Other parameters that consistently appear across the different “best case” functions are the Hansen solubility parameter of the solvent describing hydrogen bonding, which is negatively correlated with the appearance of polymorph I, and solvents having aromatic, ketone, and alcohol groups, all of which positively correlate with the appearance of the polymorph I.

Given that the quality of fit between the function and our data set levels off after the inclusion of 8 parameters, we focused on functions having 8 parameters only. Figure 5a shows the distribution of AIC for all the possible functions having 8 parameters. Of these functions, we selected 20 functions having the lowest AIC scores (red box in Figure 5a) to further determine which parameters appear consistently across these 20 best 8-parameter

functions to gain insight into the solvent properties that are most relevant for polymorphic selection. Figure 5b reveals which parameters are active in each function, represented by the same color scheme used in Figure 4b. Looking across the 20 best functions, we observe that several parameters are active in all these functions. Specifically, the solvent's hydrogen-bonding Hansen solubility parameter is consistently negatively correlated with the appearance of polymorph I, and the presence of ketone and alcohol synthons in the solvent is consistently positively correlated with the appearance of polymorph I. Fluorination of cHBC is also positively correlated with the appearance of polymorph I. These correlations are thus consistent with previous observations in Figure 4b, which considered the best functions independent of the numbers of parameters. Of the 20 8-parameter functions examined, the 8-parameter function with the lowest AIC score is

$$y = 0.574(\text{Hansen polarity}) - 1.32(\text{Hansen H-bonding}) \quad (3) \\ + 0.103(\text{solvent molar volume}) + 4.44(\text{aromatic}) \\ + 26.5(\text{alcohol}) + 8.41(\text{ketone}) + 0.179 \\ (\text{no. F on cHBC}) - 8.87$$

To validate the accuracy of the predictions of our model, we entered into eqs 1 and 3 the molecule and solvent parameters for each of the 25 data points that were previously withheld from the learning data set, obtaining for each film an expected fraction of polymorph I. Figure 6 shows the absolute difference between the expected and experimentally observed fraction of polymorph I for each of the 25 data points. The test cases yield a median absolute error of 0.023 and a root-mean-square error of 0.34, which is within the absolute spread of experiments. Indeed, 16 of the 25 samples were predicted within 0.1 absolute fraction of the experimentally derived values of polymorph I.

The relationships we obtained are correlative and not exhaustive because they are limited to the independent variables we considered. Thus, eq 3 is not representative of any fundamental relationship. However, the repeated appearance of specific molecular and solvent parameters (e.g., number of fluorines on cHBC, Hansen hydrogen-bonding, and presence of ketone and alcohol groups in the solvent) across the best models gives us confidence that these parameters are influencing polymorph selection.

That solvents with ketone and alcohol synthons, which have high tendencies to H-bond, are positively correlated with the appearance of polymorph I whereas the Hansen H-bonding parameter is negatively correlated with the appearance of polymorph I is seemingly contradictory. We speculate that this trend arises due to the fact that cHBC and its fluorinated derivatives have poor solubilities in these two classes of solvents. When precipitating the cHBC derivatives during synthesis or when trying to grow single crystals, we have often used alcohols and ketone-bearing acetone as the nonsolvent. We surmise that these nonsolvents induce aggregation of cHBC and its derivatives in their polymorph I form. Solvents with high H-bonding tendency and in which cHBC and its fluorinated derivatives have good solubility (e.g., THF) yield polymorph II.

These trends point to the complexity of molecule—molecule and molecule—solvent interactions at play and the nontriviality of gaining physical insights with which solid-state

assembly takes place from data mining. While data mining by itself does not provide these physical insights, it makes targeted studies to explore specific questions such as these more tractable by reducing the experimental phase space and parameters of interest. For example, we found here that fluorination of cHBC is positively correlated with the appearance of polymorph I. Through an entirely different set of experiments specifically focused on understanding the effects of fluorination on polymorph selection in cHBC derivatives, we tried to interconvert fluorinated cHBC films having polymorph I to polymorph II and films having polymorph II to polymorph I. We found that fluorinated cHBC films irreversibly converted from polymorph II to polymorph I.³³ Furthermore, the energy barrier to convert fluorinated cHBC films from polymorph II to polymorph I decreases with increasing fluorination,³³ and the empirical correlation determined in our present study reinforces that the fluorination of cHBC derivatives favors polymorph I.

We previously speculated that fluorination of cHBC favors polymorph I because it disrupts the balance of intermolecular C...C and C...H interactions upon fluorination.³³ Large polycyclic aromatics, like cHBC, tend to adopt one of two different molecular packing motifs, brickwork or herringbone, depending on the balance of C-C (i.e., $\pi\cdots\pi$) and C...H (i.e., $\pi\cdots\text{H}-\text{C}$) interactions.⁴⁵⁻⁴⁸ The brickwork packing motif tends to be favored when C-C intermolecular interactions dominate over C...H interactions, and the herringbone packing motif tends to be favored when C...H intermolecular interactions dominate over C...C interactions.⁴⁵⁻⁴⁸ Polymorph I has a brickwork packing motif, and we believe polymorph II has a herringbone packing motif.³³ It follows that the brickwork packing motif (polymorph I) would be increasingly favored over the herringbone packing motif (polymorph II) as the amount of intermolecular C...H interactions that favor the herringbone packing decreases with increasing fluorination of cHBC.

CONCLUSIONS

While macroscopic properties of molecular materials are known to depend on their crystal structures, controllably and predictively accessing a given crystal structure remains challenging. Starting with amorphous films of cHBC and four of its fluorinated derivatives, solvent-vapor annealing provides access to polymorph I, polymorph II, or a mixed fraction of the two depending on solvent choice. Applying a data mining approach to our data, we found a correlative relationship between the crystallizing molecule, the postdeposition crystallizing conditions, and the resulting polymorph that has allowed us to predict the fraction of polymorph I that is present in test samples. Such experimentally derived correlative relationships can be extremely powerful for guiding future design of molecules and development of processing conditions to access desired polymorphs and cannot be derived using current computational methods. Our correlative relationship that predicts polymorph access may be directly applicable to other polycyclic aromatic hydrocarbons and their fluorinated derivatives that are similar to cHBC. More broadly, the methods of polymorph screening and data mining that we employed are applicable to other molecular materials beyond organic semiconductors, like pharmaceuticals, pigments, explosives, and food additives, for which predicting and controlling molecular polymorphism is critical but as yet nearly nonexistent.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Dr. Arthur Woll (Cornell High Energy Synchrotron Source) for his assistance with GIXD experiments and Dr. Matthew Bruzek and Prof. John Anthony (University of Kentucky, Lexington) for providing the fluorinated pentacene quinone precursor used in the synthesis of fluorinated cHBC derivatives (NSF DMR-1035217). This work was supported by the NSF MRSEC program through the Princeton Center for Complex Materials (DMR-1420451) and the DMREF Program (DMR-DMR-1627453). GIXD experiments were conducted at CHESS, which is supported by NSF and NIH/NIGMS under award DMR-1332208. A.M.H. acknowledges support through the National Defense Science and Engineering Graduate (NDSEG) Fellowship (Air Force Office of Scientific Research 32 CFR 168a). Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344. LLNL-JRNL-739547.

REFERENCES

- (1). Bernstein J Polymorphism in Molecular Crystals; Oxford University Press: New York, 2002.
- (2). Chung H; Diao Y Polymorphism as an emerging design strategy for high performance organic electronics. *J. Mater. Chem. C* 2016, 4 (18), 3915–3933.
- (3). Nyman J; Day GM Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* 2015, 17 (28), 5154–5165.
- (4). Price SL Predicting crystal structures of organic compounds. *Chem. Soc. Rev.* 2014, 43 (7), 2098–2111. [PubMed: 24263977]
- (5). Hiszpanski AM; Khlyabich PP; Loo Y-L Tuning kinetic competitions to traverse the rich structural space of organic semiconductor thin films. *MRS Commun.* 2015, 5 (3), 407–421.
- (6). Hofmann DWM Data Mining in Organic Crystallography In *Data Mining in Crystallography*; Hofmann DWM, Kuleshova LN, Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2010; pp 89–134.
- (7). Yu LX; Lionberger RA; Raw AS; D’Costa R; Wu H; Hussain AS Applications of process analytical technology to crystallization processes. *Adv. Drug Delivery Rev.* 2004, 56 (3), 349–369.
- (8). Sheikhzadeh M; Murad S; Rohani S Response surface analysis of solution-mediated polymorphic transformation of buspirone hydrochloride. *J. Pharm. Biomed. Anal.* 2007, 45 (2), 227–236. [PubMed: 17658715]
- (9). Fujiwara M; Nagy ZK; Chew JW; Braatz R D. First-principles and direct design approaches for the control of pharmaceutical crystallization. *J. Process Control* 2005, 15 (5), 493–504.
- (10). Aaltonen J; Alleso M; Mirza S; Koradia V; Gordon KC; Rantanen J Solid form screening — A review. *Eur. J. Pharm. Biopharm.* 2009, 71 (1), 23–37. [PubMed: 18715549]
- (11). Jain A; Shyue Ping O; Hautier G; Chen W; Richards WD; Dacek S; Cholia S; Gunter D; Skinner D; Ceder G; Persson KA Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* 2013, 1, 011002.
- (12). Takahashi K; Tanaka Y Materials informatics: a journey towards material design and synthesis. *Dalton Trans.* 2016, 45 (26), 10497–10499. [PubMed: 27292550]
- (13). Agrawal A; Choudhary A Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. *APL Mater.* 2016, 4 (5), 053208.
- (14). Kalidindi SR; Graef MD Materials Data Science: Current Status and Future Outlook. *Annu. Rev. Mater. Res.* 2015, 45 (1), 171–193.
- (15). Pereira F; Xiao K; Latino DARS; Wu C; Zhang Q; Aires-de-Sousa, J. Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals. *J. Chem. Inf. Model.* 2017, 57 (1), 11–21. [PubMed: 28033004]

- (16). Montavon G; Rupp M; Gobre V; Vazquez-Mayagoitia A; Hansen K; Tkatchenko A; Muller K-R; Anatole von Lilienfeld O Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* 2013, 15 (9), 095003.
- (17). Morissette SL; Almarsson Ö; Peterson ML; Remenar JF; Read MJ; Lemmo AV; Ellis S; Cima MJ; Gardner CR High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids. *Adv. Drug Delivery Rev.* 2004, 56 (3), 275–300.
- (18). Morissette SL; Soukasene S; Levinson D; Cima MJ; Almarsson Ö Elucidation of crystal form diversity of the HIV protease inhibitor ritonavir by high-throughput crystallization. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100 (5), 2180–2184. [PubMed: 12604798]
- (19). Lee EH A practical guide to pharmaceutical polymorph screening & selection. *Asian J. Pharm. Sci.* 2014, 9 (4), 163–175.
- (20). Wedl B; Resel R; Leising G; Kunert B; Salzmänn I; Oehzelt M; Koch N; Vollmer A; Duhm S; Werzer O; Gbabode G; Sferrazza M; Geerts Y Crystallisation kinetics in thin films of dihexyl-terthiophene: the appearance of polymorphic phases. *RSC Adv.* 2012, 2 (10), 4404–4414.
- (21). Stevens LA; Goetz KP; Fonari A; Shu Y; Williamson RM; Bredas J-L; Coropceanu V; Jurchescu OD; Collis GE Temperature-Mediated Polymorphism in Molecular Crystals: The Impact on Crystal Packing and Charge Transport. *Chem. Mater.* 2015, 27 (1), 112–118.
- (22). Mattheus CC; Dros AB; Baas J; Meetsma A; de Boer JL; Palstra TTM Polymorphism in pentacene. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* 2001, 57, 939–941.
- (23). Mattheus CC; Dros AB; Baas J; Oostergetel GT; Meetsma A; de Boer JL; Palstra TTM Identification of polymorphs of pentacene. *Synth. Met.* 2003, 138 (3), 475–481.
- (24). Schweicher G; Paquay N; Amato C; Resel R; Koini M; Talvy S; Lemaury V; Cornil J; Geerts Y; Gbabode G Toward Single Crystal Thin Films of Terthiophene by Directional Crystallization Using a Thermal Gradient. *Cryst. Growth Des.* 2011, 11 (8), 3663–3672.
- (25). Pfattner R; Mas-Torrent M; Bilotti I; Brillante A; Milita S; Liscio F; Biscarini F; Marszałek T; Ulanski J; Nosal A; Gazicki-Lipman M; Leufgen M; Schmidt G; Molenkamp LW; Laukhin V; Veciana J; Rovira C High-Performance Single Crystal Organic Field-Effect Transistors Based on Two Dithiophene-Tetrathiafulvalene (DT-TTF) Polymorphs. *Adv. Mater.* 2010, 22 (37), 4198–4203. [PubMed: 20564710]
- (26). Schiefer S; Huth M; Dobrinevski A; Nickel B Determination of the Crystal Structure of Substrate-Induced Pentacene Polymorphs in Fiber Structured Thin Films. *J. Am. Chem. Soc.* 2007, 129 (34), 10316–10317. [PubMed: 17672461]
- (27). Giri G; Li R; Smilgies D-M; Li EQ; Diao Y; Lenn KM; Chiu M; Lin DW; Allen R; Reinspach J; Mannsfeld SCB; Thoroddsen ST; Clancy P; Bao Z; Amassian A One-dimensional self-confinement promotes polymorph selection in large-area organic semiconductor thin films. *Nat. Commun.* 2014, 5, 3573. [PubMed: 24736391]
- (28). Chen J; Shao M; Xiao K; Rondinone AJ; Loo Y-L; Kent PRC; Sumpter BG; Li D; Keum JK; Diemer PJ; Anthony JE; Jurchescu OD; Huang J Solvent-type-dependent polymorphism and charge transport in a long fused-ring organic semiconductor. *Nanoscale* 2014, 6 (1), 449–456. [PubMed: 24217182]
- (29). Yuan Y; Giri G; Ayzner AL; Zoombelt AP; Mannsfeld SCB; Chen J; Nordlund D; Toney MF; Huang J; Bao Z Ultrahigh mobility transparent organic thin film transistors grown by an offcentre spin-coating method. *Nat. Commun.* 2014, 5, 3005. [PubMed: 24398476]
- (30). Jones AOF; Geerts YH; Karpinska J; Kennedy AR; Resel R; Rothel C; Ruzié C; Werzer O; Sferrazza M Substrate-Induced Phase of a [1]Benzothieno[3,2-b]benzothiophene Derivative and Phase Evolution by Aging and Solvent Vapor Annealing. *ACS Appl. Mater. Interfaces* 2015, 7 (3), 1868–1873. [PubMed: 25569633]
- (31). Diao Y; Tee BCK; Giri G; Xu J; Kim DH; Becerril HA; Stoltenberg RM; Lee TH; Xue G; Mannsfeld SCB; Bao Z Solution coating of large-area organic semiconductor thin films with aligned single-crystalline domains. *Nat. Mater.* 2013, 12 (7), 665–671. [PubMed: 23727951]
- (32). Diao Y; Lenn KM; Lee W-Y; Blood-Forsythe MA; Xu J; Mao Y; Kim Y; Reinspach JA; Park S; Aspuru-Guzik A; Xue G; Clancy P; Bao Z; Mannsfeld SCB Understanding Polymorphism in Organic Semiconductor Thin Films through Nanoconfinement. *J. Am. Chem. Soc.* 2014, 136 (49), 17046–17057. [PubMed: 25333565]

- (33). Hiszpanski AM; Woll AR; Kim B; Nuckolls C; Loo Y-L Altering the Polymorphic Accessibility of Polycyclic Aromatic Hydrocarbons with Fluorination. *Chem. Mater.* 2017, 29 (10), 4311–4316.
- (34). Hiszpanski AM; Baur RM; Kim B; Tremblay NJ; Nuckolls C; Woll AR; Loo Y-L Tuning Polymorphism and Orientation in Organic Semiconductor Thin Films via Post-deposition Processing. *J. Am. Chem. Soc.* 2014, 136 (44), 15749–15756. [PubMed: 25317987]
- (35). Purdum GE; Yao N; Woll A; Gessner T; Weitz RT; Loo Y-L Understanding Polymorph Transformations in Core-Chlorinated Naphthalene Diimides and their Impact on Thin-Film Transistor Performance. *Adv. Funct. Mater.* 2016, 26 (14), 2357–2364.
- (36). Gundlach DJ; Jackson TN; Schlom DG; Nelson SF Solvent-induced phase transition in thermally evaporated pentacene films. *Appl. Phys. Lett.* 1999, 74 (22), 3302–3304.
- (37). Amassian A; Pozdin VA; Li R; Smilgies D-M; Malliaras GG Solvent vapor annealing of an insoluble molecular semiconductor. *J. Mater. Chem.* 2010, 20 (13), 2623–2629.
- (38). Campione M; Tavazzi S; Moret M; Porzio W Crystal-to-crystal phase transition in alpha-quaterthiophene: An optical and structural study. *J. Appl. Phys.* 2007, 101 (8), 083512.
- (39). Hiszpanski AM; Lee SS; Wang H; Woll AR; Nuckolls C; Loo Y-L Post-deposition Processing Methods To Induce Preferential Orientation in Contorted Hexabenzocoronene Thin Films. *ACS Nano* 2013, 7 (1), 294–300. [PubMed: 23228001]
- (40). Hansen CM Hansen Solubility Parameters: A Users Handbook, 2nd ed.; CRC Press: Boca Raton, FL, 2007.
- (41). Dickey KC; Anthony JE; Loo YL Improving Organic Thin-Film Transistor Performance through Solvent-Vapor Annealing of Solution-Processable Triethylsilylethynyl Anthradithiophene. *Adv. Mater.* 2006, 18 (13), 1721–1726.
- (42). Hastie T; Tibshirani R; Friedman J The Elements of Statistical Learning: Data Mining, Interference, and Prediction, 2nd ed.; Springer: New York, 2009.
- (43). Han J; Kamber M; Pei J Data Mining: Concepts and Techniques; Morgan Kaufmann Publishers Inc: San Francisco, 2011.
- (44). Osborne JW Best Practices in Quantitative Methods, 1st ed.; SAGE Publications, Inc: Thousand Oaks, CA, 2007.
- (45). Desiraju GR; Gavezzotti A Crystal structures of polynuclear aromatic hydrocarbons. Classification, rationalization and prediction from molecular structure. *Acta Crystallogr., Sect. B: Struct. Sci.* 1989, 45 (5), 473–482.
- (46). Loots L; Barbour LJ A simple and robust method for the identification of [small pi]-[small pi] packing motifs of aromatic compounds. *CrystEngComm* 2012, 14 (1), 300–304.
- (47). Spackman MA; McKinnon JJ Fingerprinting intermolecular interactions in molecular crystals. *CrystEngComm* 2002, 4 (66), 378–392.
- (48). Schatschneider B; Phelps J; Jezowski S A new parameter for classification of polycyclic aromatic hydrocarbon crystalline motifs: a Hirshfeld surface investigation. *CrystEngComm* 2011, 13 (24), 7216–7223.

Figure 1.
Chemical structures of the five cHBC derivatives and twenty-seven solvents under study.

Figure 2.

(a, b) Polarized optical micrographs and (c, d) twodimensional grazing-incidence X-ray diffraction (2D-GIXD) images of thin films of contorted hexabenzocoronene (cHBC) crystallized by solvent-vapor annealing with (left column: a, c) toluene vapor and (right column: b, d) benzene vapor. Azimuthally integrating the 2D-GIXD images produces the traces shown in (e). Polymorphs I and II have their primary diffraction peaks at $q = 0.5 \text{ \AA}^{-1}$ and $q = 0.7 \text{ \AA}^{-1}$, respectively.

Figure 3. Quantification of the fraction of cHBC and fluorinated cHBC derivative films that adopt polymorph I after exposure to different solvent vapors. Films are fully crystalline (i.e., film is entirely polymorph I or II or some mixture), and no further changes are observed with longer solvent-vapor annealing.

Figure 4.

(a) The AIC, which measures the quality of a function in describing the data and imposes a penalty for having more parameters in the function, of all possible combinations of parameters, plotted against the number of parameters in each function. A lower AIC indicates a better fit. The leveling off in the AIC beyond 8 parameters indicates that using additional parameters provides minimal benefit in yielding a better function. (b) Visual guide to the variables included in each of the “best case” functions shown in (a).

Figure 5.

(a) Distribution of AIC for all functions having 8 parameters; the 20 best functions having the lowest AIC score are boxed in red. (b) Visual guide to the variables included in each of the 20 best functions having eight parameters.

Figure 6. Absolute error between the predicted (using eq 3) and experimentally obtained fraction of polymorph I in 25 cases examined.