

RESEARCH ARTICLE

# Organic materials database: An open-access online database for data mining

Stanislav S. Borysov<sup>1\*</sup>, R. Matthias Geilhufe<sup>1</sup>, Alexander V. Balatsky<sup>1,2</sup>

**1** Nordita, Center for Quantum Materials, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden, **2** ETH Institute for Theoretical Studies, ETH Zurich, 8092 Zurich, Switzerland

\* [borysov@kth.se](mailto:borysov@kth.se)



## Abstract

We present an organic materials database (OMDB) hosting thousands of Kohn-Sham electronic band structures, which is freely accessible online at <http://omdb.diracmaterials.org>. The OMDB focus lies on electronic structure, density of states and other properties for purely organic and organometallic compounds that are known to date. The electronic band structures are calculated using density functional theory for the crystal structures contained in the Crystallography Open Database. The OMDB web interface allows users to retrieve materials with specified target properties using non-trivial queries about their electronic structure. We illustrate the use of the OMDB and how it can become an organic part of search and prediction of novel functional materials via data mining techniques. As a specific example, we provide data mining results for metals and semiconductors, which are known to be rare in the class of organic materials.

## OPEN ACCESS

**Citation:** Borysov SS, Geilhufe RM, Balatsky AV (2017) Organic materials database: An open-access online database for data mining. PLoS ONE 12(2): e0171501. doi:10.1371/journal.pone.0171501

**Editor:** Oksana Ostroverkhova, Oregon State University, UNITED STATES

**Received:** October 25, 2016

**Accepted:** January 20, 2017

**Published:** February 9, 2017

**Copyright:** © 2017 Borysov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All electronic structure data are available from the <http://omdb.diracmaterials.org/>.

**Funding:** This work is supported by the Swedish Research Council, Grant No. 638-2013-9243, <http://www.vr.se/>; Knut and Alice Wallenberg Foundation, <https://www.wallenberg.com/kaw/>; European Research Council under the European Union's Seventh Framework Program (FP/2207-2013)/ERC Grant Agreement No. DM-321031, <https://erc.europa.eu/>; Villum Foundation through the Villum Center for Dirac Materials; Dr. Max

## Introduction

Computational materials science based on *ab initio* methods has a long history of more than half a century. Development of the density functional theory (DFT) framework in the 1960s by Hohenberg and Kohn [1] and Kohn and Sham [2] marked a clear breakthrough in providing an approach that is a standard tool in modern materials science [3]. In this connection, a variety of approaches to estimate the electron density have been considered and implemented [4–8]. By now, it has been established that the most prominent codes agree well in the calculation of physical quantities by showing errors comparable to the experiment [9]. Mostly, the calculations performed are focused on a particular material of interest and motivated, for example, by providing additional information to experiments (e.g. [10, 11]). This approach can be viewed as a “one-compound-at-a-time” analysis.

In the beginning of this century, the exponential growth of computational power and high demand for prediction of materials with target properties led to a new way of dealing with *ab initio* electronic methods referred to as materials informatics [12, 13]. This approach places the main effort on performing high-throughput computing and data mining [14–16] as well as the development of sufficient tools for that [17, 18]. One can call this approach an “aggregate informatics analysis”, where the properties of a single compound are captured approximately

Rössler, Walter Haefner Foundation; and ETH Zurich Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** Among the other research grants, the work is also supported by the Knut and Alice Wallenberg Foundation through the grant for the research project "Functional Dirac Materials," the Villum Foundation through the Villum Center for Dirac Materials, Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zurich Foundation. We confirm that this does not alter our adherence to PLOS ONE policies on sharing data and materials.

and main resource is placed on understanding global trends within the large datasets. Applications of this informatics-driven approach are wide-ranging and cover, for instance, the search for functional materials [19], topological insulators [20] or the prediction of stable crystal structures [21, 22]. Instead of recalculating material properties each time, results are made available in databases [23, 24].

Motivated by this new trend in materials informatics, we focus on organic and organometallic materials because of multiple reasons. Whereas inorganic materials are well-studied by the above described methods, organic crystals are investigated rarely. One of the main difficulties lies in the large-unit cells which can contain up to several hundred atoms. Even though reports on  $\mathcal{O}(N_{\text{atoms}})$  implementations are discussed in the literature [25, 26], usual DFT codes scale with  $\mathcal{O}(N_{\text{atoms}}^2 \log N_{\text{atoms}})$  up to  $\mathcal{O}(N_{\text{atoms}}^3)$  [27] leading to a high computational demand for large unit cells. New computational resources and modern code architectures have opened the path for such system sizes within the last decade [28, 29].

Organic crystals offer a high potential for technological applications [30, 31]. The main constituents of organic crystals are carbon, hydrogen, nitrogen, oxygen and, in rare cases, a low percentage of transition metal elements. This makes production of organics inexpensive and accessible in terms of raw materials. This potential for applications, utility and availability motivates the investigation of organic solar cells as realistic alternative to currently used cells based on inorganic semiconductors [32, 33]. Aside from application in organic solar cells, there are reports on *d*-wave superconductivity for the materials  $\kappa$ -(BEDT-TTF)<sub>2</sub>Cu(NCS)<sub>2</sub> [34] and  $\kappa$ -(BEDT-TTF)<sub>2</sub>Cu[N(CN)<sub>2</sub>]Br [35]. Due to the softness, some materials show interesting conduction phenomena under high pressure, like the material  $\alpha$ -(BEDT-TTF)<sub>2</sub>I<sub>3</sub>, where a tilted Dirac cone can be induced within the band structure close to the Fermi level [36]. The elastic properties of organic materials make them particularly suitable for various applications in flexible electronics [37, 38].

In this paper, we report on setting up a web database for organic crystals as a source for data mining projects promoting the *ab initio* investigation of organics and the prediction of organic functional materials. The database itself contains thousands (6461 at the time of writing) of calculated Kohn-Sham band structures. The implemented web interface allows for fast online search algorithms to identify materials with specified electronic properties.

The overall data flow chart for the organization of the database is shown in Fig 1. Details are discussed throughout the paper.

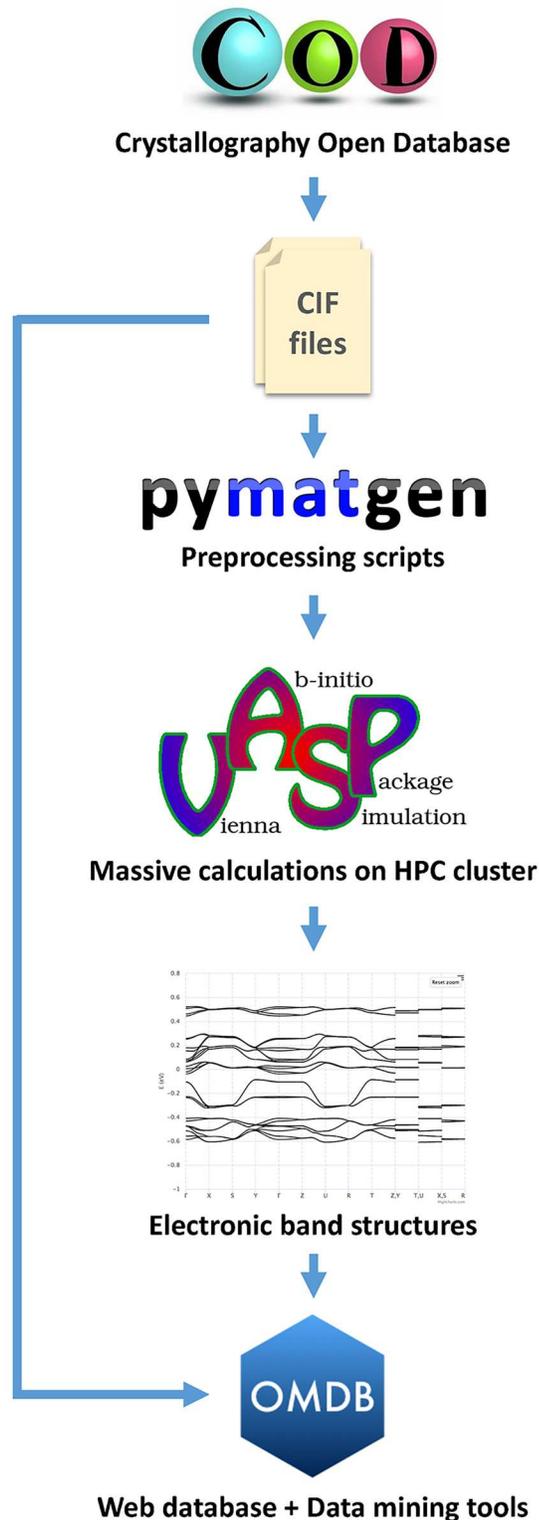
The paper is organized as follows. In Materials and Methods, we describe the crystallographic data and DFT calculation details along with the OMDB software implementation. In Results, the OMDB web interface and capabilities for data mining are introduced. Examples of the database usage for mining of novel functional materials such as organic metals and semiconductors are provided in Discussion. Finally, the scope and capabilities are summed up in Conclusions, where we also discuss the current status of the OMDB database and its potential future improvements.

## Materials and methods

### Crystallographic data

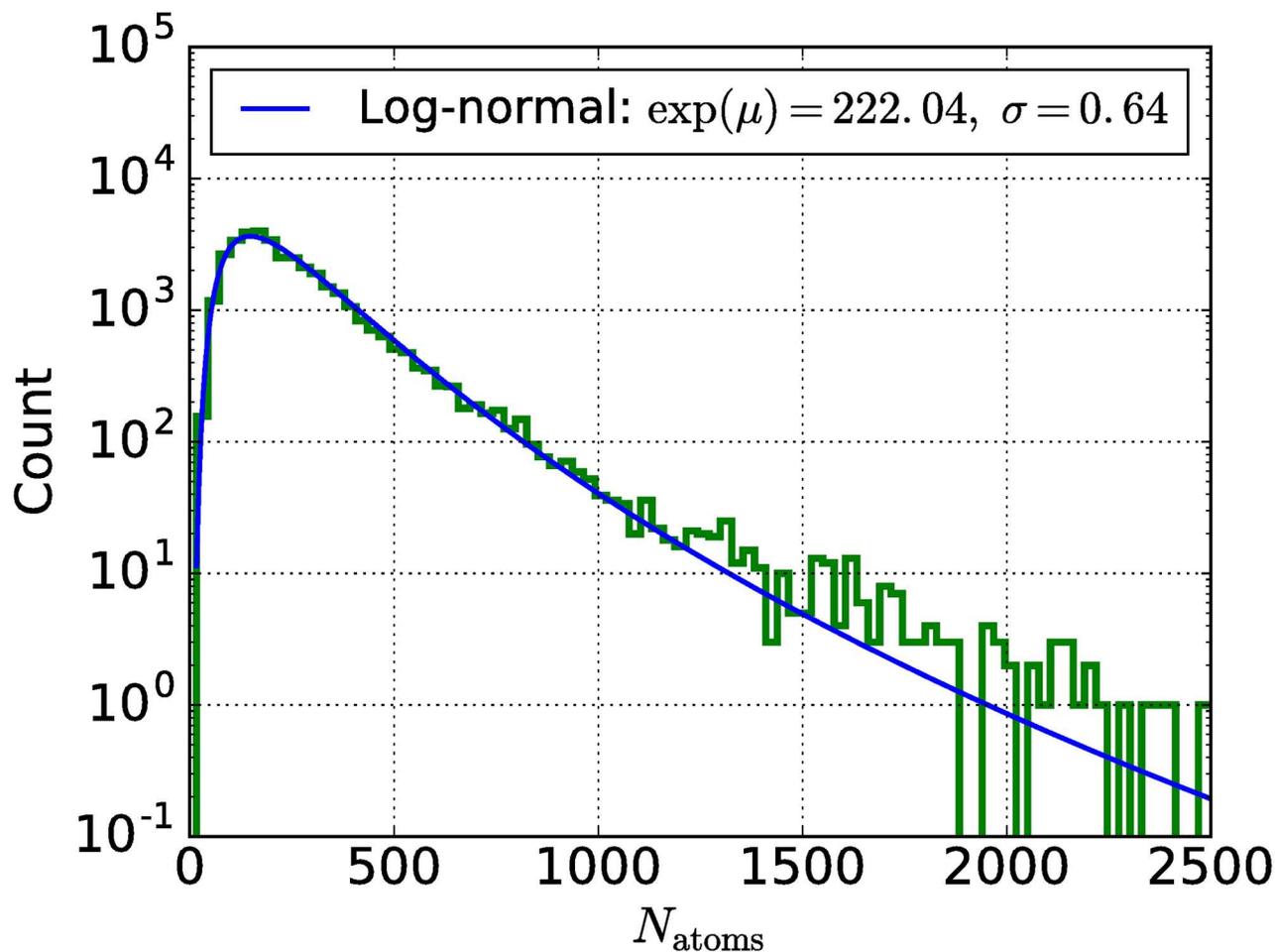
The structural information for organic compounds were taken from the Crystallography Open Database (COD) [39–41] which is available online at <http://crystallography.net>. The COD provides structural information in the Crystallographic Interchange File/Framework (CIF) files [42].

Although there are about 300,000 materials in the COD containing carbon, we decided to focus first on the 50,211 previously synthesized materials described in four dedicated experimental organic chemistry journals: "Organometallics" [43], "Organic Letters" [44], "Journal of



**Fig 1. The OMDB data flow chart.** Crystallographic data contained in the COD database in the CIF format is converted to DFT input by applying the Pymatgen package. DFT electronic structure calculations are performed using the VASP package. The DFT output (band structures and density of states) along with the basic crystallographic data from the CIF files are stored in the OMDB database, which also provides data mining tools to retrieve materials with specified by users electronic structure properties.

doi:10.1371/journal.pone.0171501.g001



**Fig 2. Histogram of number of atoms per unit cell for 37,941 organic compounds from four experimental organic chemistry journals contained within the COD database.** Blue solid line denotes log-normal fit with a median value  $\exp(\mu)$  of 222.04 atoms and a standard deviation  $\sigma$  of 0.64  $\ln(\text{atoms})$ .

doi:10.1371/journal.pone.0171501.g002

Organic Chemistry” [45] and “Organic & Biomolecular Chemistry” [46]. However, it was not possible to do DFT calculations for all of them. Incomplete structures or structures with fractional occupation of ionic sites were excluded (12,270 structures or 24% of the initial data). For the remaining 37,941 materials, the main limitation lies on the polynomial complexity of DFT algorithms with respect to a number of atoms in the unit cell. Organic crystals have on average larger unit cells comparing to inorganic crystal structures. For illustration purpose, a histogram of  $N_{\text{atoms}}$  per unit cell for the considered materials is shown in Fig 2. The shape remarkably follows a log-normal distribution with median value of 222 atoms per unit cell.

To further elaborate on this point, we split the 37,941 input materials into four classes depending on  $N_{\text{atoms}}$  per unit cell. Rough estimation of computational resources provided in Table 1 shows that it would require more than 70 million core hours of calculations on a typical modern CPU to cover this subset of materials. Given medium-scale HPC computing resources available, we were able to calculate materials with up to 120 of atoms in the unit cell, which have led to 6461 database entries at the time of writing the paper. DFT calculations for the materials with larger unit cells and other carbon-based structures from the rest of the journals are in progress.

**Table 1. Summary of the input crystallographic data and rough estimation of the computational resources required to complete DFT calculations for the 37,941 organic compounds from four experimental organic chemistry journals contained within the COD database.** Core hours (c×h) are estimated based on the actual computational time of self-consistency calculations followed by density of states and band structure calculations on a single-core Intel Xeon 2.2 GHz assuming  $\mathcal{O}(N_{\text{atoms}}^2 \log N_{\text{atoms}})$  complexity of the DFT algorithm.

Class	$N_{\text{atoms}}$	$N$ mat. (%)	c×h per mat. mean	c×h total
Small	≤120	6,363 (16.8%)	95	610K
Medium	121–250	16,052 (42.3%)	430	7,000K
Large	251–500	11,598 (30.6%)	1,735	20,000K
Extra large	≥501	3,928 (10.3%)	11,070	43,000K
<b>Total</b>		<b>37,941</b>		<b>70,610K</b>

doi:10.1371/journal.pone.0171501.t001

## Electronic structure calculations

CIF files from the COD database were transformed into input files for the Vienna Ab initio Simulation Package (VASP) [8, 27, 47] by applying the Pymatgen package [41]. For the DFT-based calculations, the projector augmented wave method [48–51] was applied as implemented in VASP and Quantum ESPRESSO [4]. The exchange-correlation functional was approximated by the generalized gradient approximation (GGA) according to Perdew, Burke and Ernzerhof [52]. Within VASP, the precision flag was set to “NORMAL” and therefore the energy cut-off is given by the maximum of the specified maxima within the POTCAR files. For example for carbon, this value is given by 400 eV. To properly describe the influence of transition metal elements, the calculations were performed spin-polarized. The provided structural information were kept and no further relaxation was considered. For the integration in  $\vec{k}$ -space, a  $6 \times 6 \times 6$   $\Gamma$ -centered Monkhorst-Pack grid [53] was chosen for the self-consistent cycle. The  $\vec{k}$ -path for the band structure calculations was automatically generated by the Pymatgen package.

## Database implementation and version control system

The OMDB database is implemented as an open-access database available online at <http://omdb.diracmaterials.org>. Both the website's back and front ends were implemented using the PHP language. Additionally, for the front end, the JavaScript language was used. The core of the OMDB is a MySQL database, where all information about materials from the CIF files (e.g. chemical formula, crystal lattice parameters and symmetry group) as well as the output of the DFT calculations (electronic band structures and density of states) are stored. Extended outputs of DFT calculations (e.g., charge distribution, magnetization, orbital projected density of states) for each material are stored in the server's file system. Every material has a unique OMDB identifier assigned. Furthermore, the COD identifier is kept (if available) to maintain consistency with the COD database.

We use Git [54] to keep track of the development of the database as it represents a widely used version control software. Hence, it is possible to recover a complete history of all modifications of any database entry. Related changes history for each entry is shown on the material's information page.

## Results

### Database user interface

The user interface and functionality of the OMDB website have been developed in the style of the functionality of the COD database. It allows users to browse through all database entries or

particular previously data-mined groups of materials, for example metals or semiconductors. The website also provides a basic search mechanism, where the user can specify full or partial chemical formula, chemical name or symmetry group of interest to retrieve a list of relevant materials. The OMDB also provides a more advanced electronic band structure search, which is described in the following subsection.

The information page for a selected material (Fig 3) shows basic information about its crystal structure followed by the link to the COD entry if available. It also provides interactive electronic band structure and density of states plots implemented by using the Highcharts JavaScript library [55]. It allows users to zoom in to a specific energy or  $\vec{k}$ -path range. Furthermore, it is possible to download the plots in one of the popular graphic file formats (PNG, JPEG, PDF or SVG).

## Electronic band structure search

In addition to the basic material retrieval system, the OMDB provides an interface for an advanced band structure search, which can be divided into the two following categories:

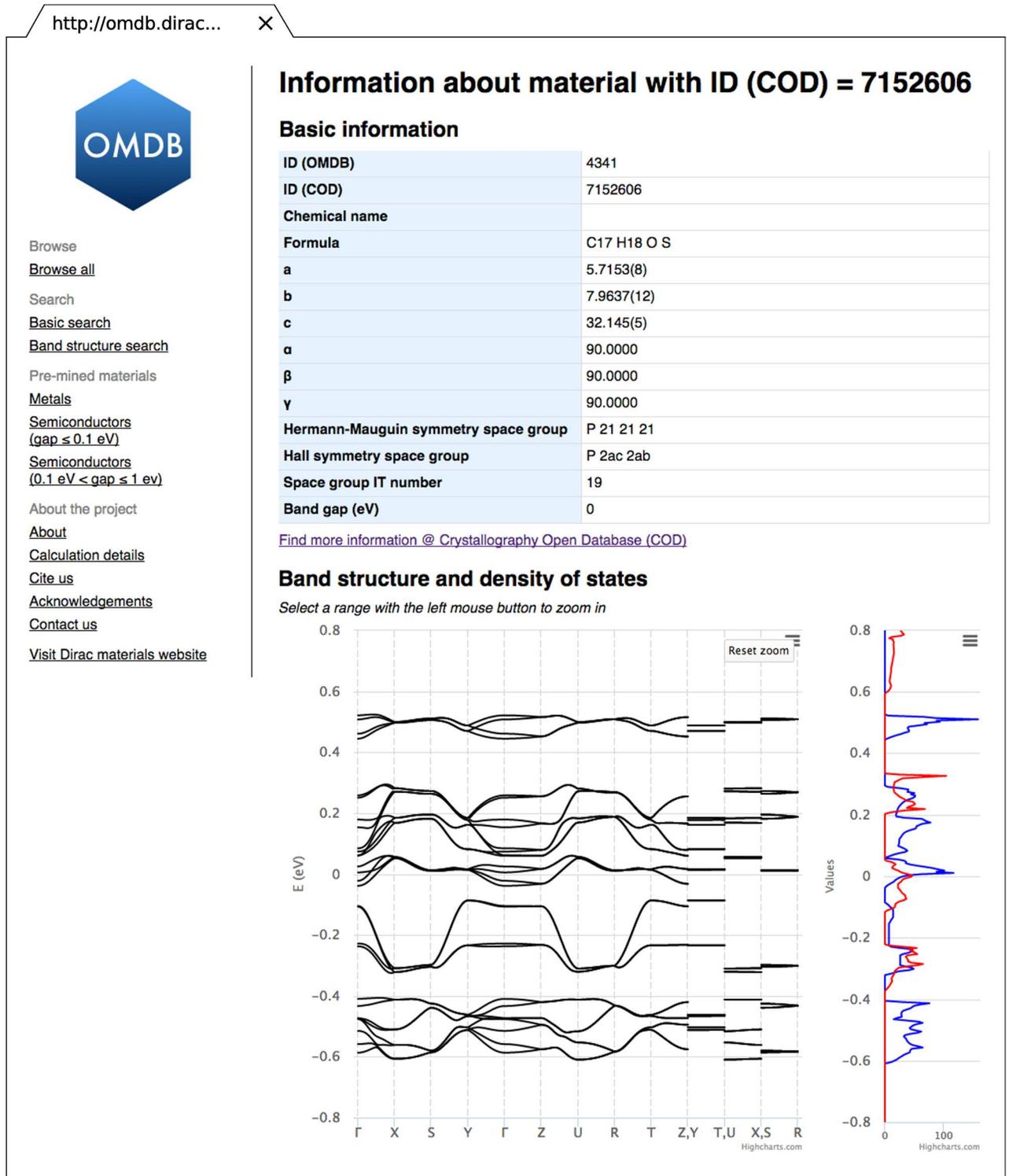
1. *“Hard” criteria search.* The database users can provide a rigorous definition of the particular properties the band structure needs to satisfy, for example, presence or absence of a spectral gap of a particular size in a specified energy range.
2. *“Soft” criteria search.* The database users can search for a graphical pattern by making use of a similarity measure, for example, root mean square error (RMSE) or more advanced probability measures [56, 57]. For example, a pattern can specify two crossing straight lines for the search of Dirac materials [58] like graphene or two touching parabolic bands for the search of other semi-metals.

The crucial difference between these two retrieval techniques is that the former completely discards search results which does not satisfy specified search criteria while the latter can only range materials according to some similarity measure, i.e., a single real number. In the latter case, discarding of search results can be based on an essentially subjective threshold value of the similarity measure.

While the “soft” search technique is only implemented within the offline database version at the moment, the “hard” one is fully functional with acceptable for online usage search execution time. Currently, it provides search possibility for gap presence/absence of particular size in the energy range specified by the user (Fig 4). The other possibilities for this type of search, for instance, the number of electronic bands crossing a particular energy level (which might be important for the discovery of new superconductors), number of electrons or magnetization in the particular energy range, and number of states at the Fermi level, will be implemented in the nearest future. The database users are always encouraged to suggest new search functionality missing within the present version of the database.

## Discussion

As an application of the search tools developed for band structure data mining, we searched for all materials with either zero or small band gap  $\Delta \leq 1$  eV around the Fermi energy. No distinction between direct and indirect band gaps has been made, i.e.,  $\Delta$  was defined as a distance between the minimum energy of the lowest conduction band and the maximum energy of the highest valence band independently of the momentum vector  $\vec{k}$ . Such materials, metals and semiconductors respectively, are of high practical interest for the organic electronics industry. However, these properties are rarely observed in organic crystals, which are mostly wide-gap



**Fig 3. Web interface of the OMDB database.** A web page with basic crystallographic information about a material, interactive band structure and density of states plots.

doi:10.1371/journal.pone.0171501.g003

http://omdb.dirac... X

## Band structure search

**Restrictions**

Enabled

*Material should*

have a gap of size from  to  eV

have no gap (material is a conductor)

*Where?*

anywhere

around energy level  eV

within the energy range from  to  eV

*Reference energy*

$E_{\text{Fermi}}$  from the DFT calculations

Highest valence band energy

Lowest conduction band energy

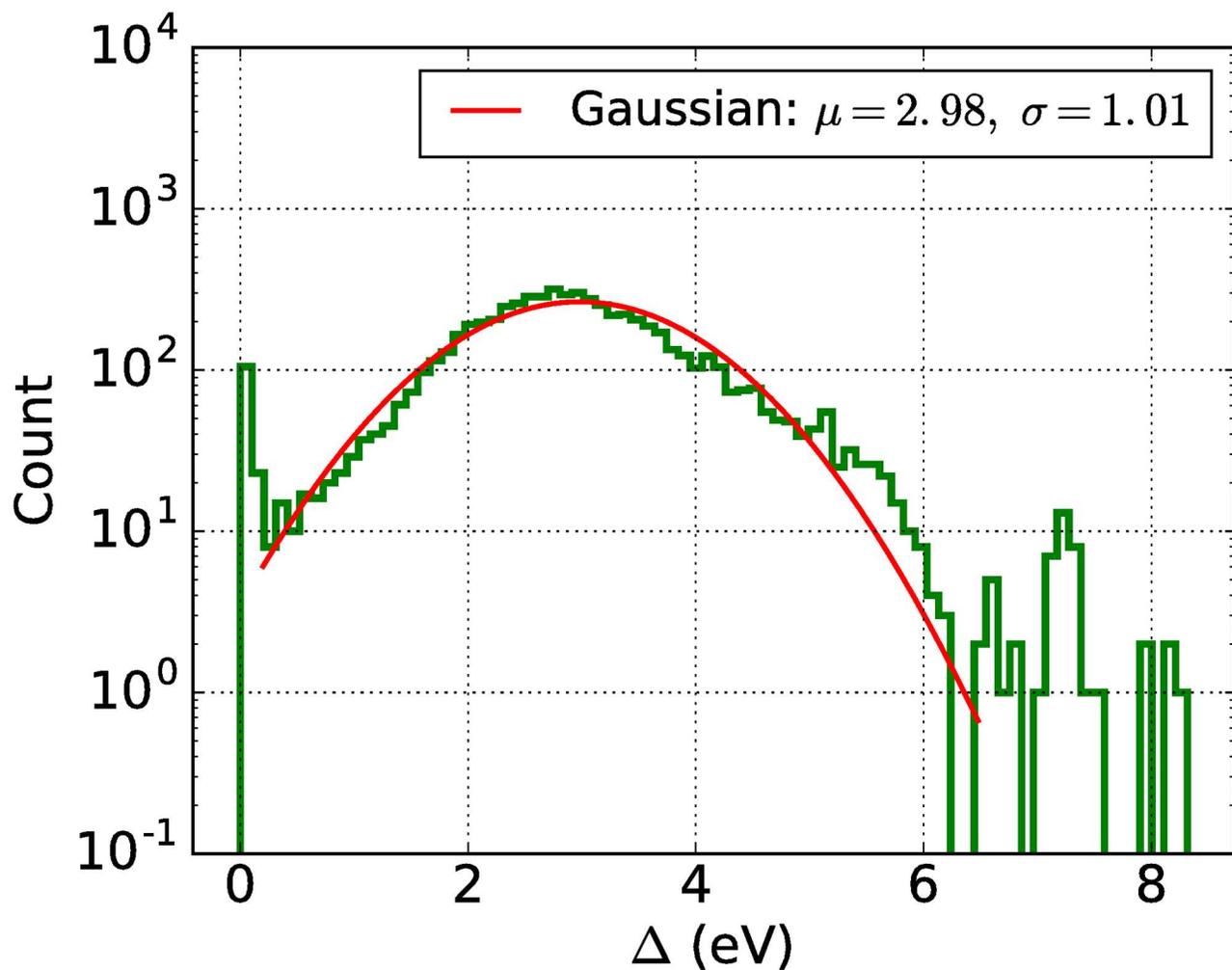
Search

**Fig 4. Web interface of the advanced electronic band structure search.** Users can search for gap presence/absence of particular size in the energy range.

doi:10.1371/journal.pone.0171501.g004

insulators [59]. It can also be verified from the histogram of the band gaps of all materials within the OMDB depicted in Fig 5. Remarkably, its bulk shape is close to the (truncated) Gaussian distribution with a mean value of 2.98 eV and a standard deviation of 1.01 eV. Nevertheless, there are a few outliers with a band gap close to zero. In total, by using the implemented OMDB band structure search, 93 suspect materials to be organic metals ( $\Delta = 0$  eV), 11 narrow band gap semiconductors ( $0 < \Delta \leq 0.1$  eV) and 151 semiconductors ( $0.1 < \Delta \leq 1$  eV) were identified. The semiconductors as well as metals are tabulated on the OMDB website.

Modern DFT approaches usually fail in accurate band gap estimations as long as no explicit correlation corrections are applied [60]. As mentioned in Materials and Methods, the GGA approximation of the exchange-correlation functional is used, which is known to systematically underestimate band gaps by about 30–100% [61–63] (see also related discussion on the Materials Project website [64]). Going beyond GGA to improve the accuracy of the DFT band gaps [61, 65–67] and adding experimental data when available is one of the future directions planned for the OMDB. So far, a warning concerning the accuracy of GGA band gaps is shown together with the electronic structures on the website. Nevertheless, the GGA band gap errors can be regarded as statistically systematic in some sense. Particularly, the large number of calculated materials opens up the possibility for a general discussion of trends and features within the electronic structures. It is important to stress that the main goal of the presented



**Fig 5. Histogram of the band gap  $\Delta$  (without distinction between direct and indirect gaps) for 6461 organic materials within the OMDB database.** Red solid line denotes Gaussian fit with a mean value  $\mu$  of 2.98 eV and a standard deviation  $\sigma$  of 1.01 eV.

doi:10.1371/journal.pone.0171501.g005

database (and most of the other databases containing output from high-throughput DFT calculations) is to provide users with general guidance in the search space.

The application of pattern search algorithms will be available soon within the online version of the website. So far, the offline version has been successfully applied for the search of 3D organic Dirac-point [68] and Dirac-line [69] materials together with an investigation of their topological protection properties for particular crystal symmetry groups.

## Conclusions

We presented the new electronic structure database on organometallics and pure organic materials. The Organic Materials Database (OMDB) currently contains 6461 entries and is accessible via a web-interface at <http://omdb.diracmaterials.org>. At the current stage, the OMDB database builds the connection between already available structural information, taken from the Crystallography Open Database (COD), with the *ab initio* electronic structure calculations based on the density functional theory (DFT). The implemented structure of the database also allows for an extension beyond the materials contained in the COD database. The

presented analysis for the 37,941 materials described in four experimental organic chemistry journals have shown that the number of atoms in their unit cells follows log-normal distribution with the median value of 222 atoms. This relatively large number represents a challenge for high-throughput DFT calculations for organic crystals as the algorithm scales polynomially with the number of atoms. We have roughly estimated that more than 70 millions of core hours of calculations on a typical modern CPU are required to cover this relatively small subset of organic materials. Given medium-scale HPC computational resources, we were able to calculate materials with up to 120 of atoms in the unit cell so far. We plan to extend our calculations to the crystal structures with larger unit cells and materials from other chemical journals in the nearest future.

Although the performed DFT calculations are not fine-tuned to each separate material, the large amount of provided Kohn-Sham band structures and densities of states allows for a general discussion of trends and features within the electronic structures. The core feature of the OMDB is to provide advanced tools aimed for efficient data mining studies of materials with specified electronic target properties. As an example, we discussed the distribution of the band gaps for the calculated materials. Surprisingly, it shows a simple (truncated) Gaussian shape with a mean value of 2.98 eV and a standard deviation of 1.01 eV. Hence, identifying organic metals or semiconductors is a non-trivial task. The probability of randomly finding a metal using high-throughput DFT calculations is given by 1.4% and of finding a semiconductor with a gap less than 1 eV is less than 2.5%. Although DFT band gaps are usually underestimated, the presented procedure helps to shrink the search space and provide guidance for further theoretical and experimental work. In exchange with the research community, we actively plan to extend the existing OMDB search tools to include broader options related to properties of electronic band structures and density of states.

## Acknowledgments

The work is supported by the Swedish Research Council Grant No. 638-2013-9243, the Knut and Alice Wallenberg Foundation, the Villum foundation, the European Research Council under the European Union's Seventh Framework Program (FP/2207-2013)/ERC Grant Agreement No. DM-321031, Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zurich Foundation. The authors also acknowledge computational resources from the Max Planck Institute of Microstructure Physics in Halle (Germany) and the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre at Linköping University.

## Author contributions

**Conceptualization:** AVB SSB RMG.

**Data curation:** SSB RMG.

**Formal analysis:** SSB RMG.

**Funding acquisition:** AVB.

**Investigation:** AVB SSB RMG.

**Methodology:** AVB SSB RMG.

**Project administration:** AVB.

**Resources:** AVB RMG.

**Software:** SSB RMG.

**Supervision:** AVB.

**Validation:** SSB RMG.

**Visualization:** SSB.

**Writing – original draft:** SSB RMG AVB.

**Writing – review & editing:** SSB RMG AVB.

## References

1. Hohenberg P, Kohn W. Inhomogeneous Electron Gas. *Phys Rev.* 1964; 136:B864–B871. doi: [10.1103/PhysRev.136.B864](https://doi.org/10.1103/PhysRev.136.B864)
2. Kohn W, Sham LJ. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev.* 1965; 140:A1133–A1138. doi: [10.1103/PhysRev.140.A1133](https://doi.org/10.1103/PhysRev.140.A1133)
3. Jones RO. Density functional theory: Its origins, rise to prominence, and future. *Rev Mod Phys.* 2015; 87:897–923. doi: [10.1103/RevModPhys.87.897](https://doi.org/10.1103/RevModPhys.87.897)
4. Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter.* 2009; 21(39):395502 (19pp). doi: [10.1088/0953-8984/21/39/395502](https://doi.org/10.1088/0953-8984/21/39/395502) PMID: [21832390](https://pubmed.ncbi.nlm.nih.gov/21832390/)
5. Ebert H, Ködderitzsch D, Minár J. Calculating condensed matter properties using the KKR-Green's function method—recent developments and applications. 2011; 74(9):096501.
6. Lüders M, Ernst A, Temmerman WM, Szotek Z, Durham PJ. Ab initio angle-resolved photoemission in multiple-scattering formulation. 2001; 13(38):8587.
7. Vitos L. Total-energy method based on the exact muffin-tin orbitals theory. *Physical Review B.* 2001; 64(1):014107. doi: [10.1103/PhysRevB.64.014107](https://doi.org/10.1103/PhysRevB.64.014107)
8. Kresse G, Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B.* 1996; 54(16):11169. doi: [10.1103/PhysRevB.54.11169](https://doi.org/10.1103/PhysRevB.54.11169)
9. Lejaeghere K, Bihlmayer G, Björkman T, Blaha P, Blügel S, Blum V, et al. Reproducibility in density functional theory calculations of solids. *Science.* 2016; 351(6280). doi: [10.1126/science.aad3000](https://doi.org/10.1126/science.aad3000) PMID: [27013736](https://pubmed.ncbi.nlm.nih.gov/27013736/)
10. Nayak SK, Langhammer HT, Adeagbo WA, Hergert W, Müller T, Böttcher R. Chromium point defects in hexagonal BaTiO<sub>3</sub>: A comparative study of first-principles calculations and experiments. *Physical Review B.* 2015; 91(15):155105. doi: [10.1103/PhysRevB.91.155105](https://doi.org/10.1103/PhysRevB.91.155105)
11. Tikkanen J, Geilhufe M, Frontzek M, Hergert W, Ernst A, Paturi P, et al. The low-temperature magnetostructure and magnetic field response of Pr<sub>0.9</sub>Ca<sub>0.1</sub>MnO<sub>3</sub>: the roles of Pr spins and magnetic phase separation. *Journal of Physics: Condensed Matter.* 2016; 28(3):036001. PMID: [26732100](https://pubmed.ncbi.nlm.nih.gov/26732100/)
12. Rajan K. Materials informatics. *Materials Today.* 2005; 8(10):38–45. [http://dx.doi.org/10.1016/S1369-7021\(05\)71123-8](http://dx.doi.org/10.1016/S1369-7021(05)71123-8).
13. Rodgers JR, Cebon D. Materials Informatics. *MRS Bulletin.* 2006; 31:975–980. doi: [10.1557/mrs2006.223](https://doi.org/10.1557/mrs2006.223)
14. Morgan D, Ceder G, Curtarolo S. High-throughput and data mining with ab initio methods. *Measurement Science and Technology.* 2005; 16(1):296. doi: [10.1088/0957-0233/16/1/039](https://doi.org/10.1088/0957-0233/16/1/039)
15. Jain A, Hautier G, Moore CJ, Ong SP, Fischer CC, Mueller T, et al. A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science.* 2011; 50(8):2295–2310. <http://dx.doi.org/10.1016/j.commatsci.2011.02.023>.
16. Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, Levy O. The high-throughput highway to computational materials design. *Nature materials.* 2013; 12(3):191–201. doi: [10.1038/nmat3568](https://doi.org/10.1038/nmat3568) PMID: [23422720](https://pubmed.ncbi.nlm.nih.gov/23422720/)
17. Curtarolo S, Setyawan W, Hart GLW, Jahnatek M, Chepulskii RV, Taylor RH, et al. AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science.* 2012; 58:218–226. <http://dx.doi.org/10.1016/j.commatsci.2012.02.005>.
18. Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, et al. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science.* 2012; 58:227–235. <http://dx.doi.org/10.1016/j.commatsci.2012.02.002>.

19. Ortiz C, Eriksson O, Klintonberg M. Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Computational Materials Science*. 2009; 44(4):1042–1049. doi: [10.1016/j.commatsci.2008.07.016](https://doi.org/10.1016/j.commatsci.2008.07.016)
20. Klintonberg M, Haraldsen J, Balatsky A. Computational search for strong topological insulators: an exercise in data mining and electronic structure. *Applied Physics Research*. 2014; 6(4):31. doi: [10.5539/apr.v6n4p31](https://doi.org/10.5539/apr.v6n4p31)
21. Sarmiento-Pérez R, Cerqueira TFT, Körbel S, Botti S, Marques MAL. Prediction of Stable Nitride Perovskites. *Chemistry of Materials*. 2015; 27(17):5957–5963. doi: [10.1021/acs.chemmater.5b02026](https://doi.org/10.1021/acs.chemmater.5b02026)
22. Curtarolo S, Morgan D, Persson K, Rodgers J, Ceder G. Predicting Crystal Structures with Data Mining of Quantum Calculations. *Phys Rev Lett*. 2003; 91:135503. doi: [10.1103/PhysRevLett.91.135503](https://doi.org/10.1103/PhysRevLett.91.135503) PMID: [14525315](https://pubmed.ncbi.nlm.nih.gov/14525315/)
23. Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*. 2013; 1(1):011002. doi: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323)
24. Rasmussen FA, Thygesen KS. Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides. *The Journal of Physical Chemistry C*. 2015; 119(23):13169–13183. doi: [10.1021/acs.jpcc.5b02950](https://doi.org/10.1021/acs.jpcc.5b02950)
25. Guerra CF, Snijders J, Te Velde G, Baerends E. Towards an order-N DFT method. *Theoretical Chemistry Accounts*. 1998; 99(6):391–403. doi: [10.1007/s002140050353](https://doi.org/10.1007/s002140050353)
26. Zeller R. Towards a linear-scaling algorithm for electronic structure calculations with the tight-binding Korringa-Kohn-Rostoker Green function method. *Journal of Physics: Condensed Matter*. 2008; 20(29):294215.
27. Kresse G, Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*. 1996; 6(1):15–50. doi: [10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0)
28. Hafner J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of computational chemistry*. 2008; 29(13):2044–2078. doi: [10.1002/jcc.21057](https://doi.org/10.1002/jcc.21057) PMID: [18623101](https://pubmed.ncbi.nlm.nih.gov/18623101/)
29. Thiess A, Zeller R, Bolten M, Dederichs PH, Blügel S. Massively parallel density functional calculations for thousands of atoms: KKRnano. *Phys Rev B*. 2012; 85:235103. doi: [10.1103/PhysRevB.85.235103](https://doi.org/10.1103/PhysRevB.85.235103)
30. MacDiarmid AG. “Synthetic Metals”: A Novel Role for Organic Polymers (Nobel Lecture). *Angewandte Chemie International Edition*. 2001; 40(14):2581–2590. doi: [10.1002/1521-3773\(20010716\)40:14%3C2581::AID-ANIE2581%3E3.0.CO;2-2](https://doi.org/10.1002/1521-3773(20010716)40:14%3C2581::AID-ANIE2581%3E3.0.CO;2-2) PMID: [11458347](https://pubmed.ncbi.nlm.nih.gov/11458347/)
31. Gershenson ME, Podzorov V, Morpurgo AF. *Colloquium*: Electronic transport in single-crystal organic transistors. *Rev Mod Phys*. 2006; 78:973–989. doi: [10.1103/RevModPhys.78.973](https://doi.org/10.1103/RevModPhys.78.973)
32. Hoppe H, Sariciftci NS. Organic solar cells: An overview. *Journal of Materials Research*. 2011; 19:1924–1945. doi: [10.1557/JMR.2004.0252](https://doi.org/10.1557/JMR.2004.0252)
33. Brabec CJ, Dyakonov V, Parisi J, Sariciftci NS. Organic photovoltaics: concepts and realization. vol. 60. Springer Science & Business Media; 2013.
34. Arai T, Ichimura K, Nomura K, Takasaki S, Yamada J, Nakatsuji S, et al. Tunneling spectroscopy on the organic superconductor  $\kappa$ -(BEDT-TTF)<sub>2</sub>Cu(NCS)<sub>2</sub> using STM. *Phys Rev B*. 2001; 63:104518. doi: [10.1103/PhysRevB.63.104518](https://doi.org/10.1103/PhysRevB.63.104518)
35. Ichimura K, Takami M, Nomura K. Direct Observation of d-Wave Superconducting Gap in  $\kappa$ -(BEDT-TTF)<sub>2</sub>Cu[N(CN)<sub>2</sub>]Br with Scanning Tunneling Microscopy. *Journal of the Physical Society of Japan*. 2008; 77(11):114707. doi: [10.1143/JPSJ.77.114707](https://doi.org/10.1143/JPSJ.77.114707)
36. Tajima N, Tamura M, Nishio Y, Kajita K, Iye Y. Transport Property of an Organic Conductor  $\alpha$ -(BEDT-TTF)<sub>2</sub>I<sub>3</sub> under High Pressure-Discovery of a Novel Type of Conductor. *Journal of the Physical Society of Japan*. 2000; 69(2):543–551. doi: [10.1143/JPSJ.69.543](https://doi.org/10.1143/JPSJ.69.543)
37. Nathan A, Ahnood A, Cole MT, Lee S, Suzuki Y, Hiralal P, et al. Flexible Electronics: The Next Ubiquitous Platform. *Proceedings of the IEEE*. 2012; 100(Special Centennial Issue):1486–1517. doi: [10.1109/JPROC.2012.2190168](https://doi.org/10.1109/JPROC.2012.2190168)
38. Kim DH, Ghaffari R, Lu N, Rogers JA. Flexible and Stretchable Electronics for Biointegrated Devices. *Annual Review of Biomedical Engineering*. 2012; 14(1):113–128. doi: [10.1146/annurev-bioeng-071811-150018](https://doi.org/10.1146/annurev-bioeng-071811-150018) PMID: [22524391](https://pubmed.ncbi.nlm.nih.gov/22524391/)
39. Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research*. 2012; 40(D1):D420–D427. doi: [10.1093/nar/gkr900](https://doi.org/10.1093/nar/gkr900)

40. Gražulis S, Chateigner D, Downs RT, Yokochi AFT, Quirós M, Lutterotti L, et al. Crystallography Open Database—an open-access collection of crystal structures. *Journal of Applied Crystallography*. 2009; 42(4):726–729. doi: [10.1107/S0021889809016690](https://doi.org/10.1107/S0021889809016690)
41. Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*. 2013; 68:314–319. <http://dx.doi.org/10.1016/j.commatsci.2012.10.028>.
42. Hall SR, Allen FH, Brown ID. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A*. 1991; 47(6):655–685.
43. “Organometallics”;. <http://pubs.acs.org/journal/orgnd7>.
44. “Organic Letters”;. <http://pubs.acs.org/journal/orlef7>.
45. “Journal of Organic Chemistry”;. <http://pubs.acs.org/journal/jocean>.
46. “Organic & Biomolecular Chemistry”;. <http://pubs.rsc.org/en/journals/journalissues/ob>.
47. Kresse G, Hafner J. *Ab initio* molecular dynamics for liquid metals. *Phys Rev B*. 1993; 47:558–561. doi: [10.1103/PhysRevB.47.558](https://doi.org/10.1103/PhysRevB.47.558)
48. Blöchl PE. Projector augmented-wave method. *Physical Review B*. 1994; 50(24):17953. doi: [10.1103/PhysRevB.50.17953](https://doi.org/10.1103/PhysRevB.50.17953)
49. Kresse G, Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*. 1999; 59(3):1758. doi: [10.1103/PhysRevB.59.1758](https://doi.org/10.1103/PhysRevB.59.1758)
50. Vanderbilt D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical Review B*. 1990; 41(11):7892. doi: [10.1103/PhysRevB.41.7892](https://doi.org/10.1103/PhysRevB.41.7892)
51. Kresse G, Hafner J. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. *Journal of Physics: Condensed Matter*. 1994; 6(40):8245.
52. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Physical review letters*. 1996; 77(18):3865. doi: [10.1103/PhysRevLett.77.3865](https://doi.org/10.1103/PhysRevLett.77.3865) PMID: [10062328](https://pubmed.ncbi.nlm.nih.gov/10062328/)
53. Monkhorst HJ, Pack JD. Special points for Brillouin-zone integrations. *Physical Review B*. 1976; 13(12):5188. doi: [10.1103/PhysRevB.13.5188](https://doi.org/10.1103/PhysRevB.13.5188)
54. Git;. <http://git-scm.com>.
55. Highsoft AS;. <http://highcharts.com>.
56. Keogh EJ, Smyth P. A Probabilistic Approach to Fast Pattern Matching in Time Series Databases. In: *KDD*;. p. 24–30.
57. Ge X, Smyth P. Deformable Markov Model Templates for Time-series Pattern Matching. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'00*. New York, NY, USA: ACM; 2000. p. 81–90. Available from: <http://doi.acm.org/10.1145/347090.347109>.
58. Wehling TO, Black-Schaffer AM, Balatsky AV. Dirac materials. *Advances in Physics*. 2014; 63(1):1–76. doi: [10.1080/00018732.2014.927109](https://doi.org/10.1080/00018732.2014.927109)
59. Uji S, Mori T, Takahashi T. Focus on Organic Conductors. *Science and Technology of Advanced Materials*. 2009; 10(2):020301. doi: [10.1088/1468-6996/10/2/020301](https://doi.org/10.1088/1468-6996/10/2/020301) PMID: [27877272](https://pubmed.ncbi.nlm.nih.gov/27877272/)
60. Johnson KA, Ashcroft N. Corrections to density-functional theory band gaps. *Physical Review B*. 1998; 58(23):15548. doi: [10.1103/PhysRevB.58.15548](https://doi.org/10.1103/PhysRevB.58.15548)
61. Chan MKY, Ceder G. Efficient Band Gap Prediction for Solids. *Phys Rev Lett*. 2010; 105:196403. doi: [10.1103/PhysRevLett.105.196403](https://doi.org/10.1103/PhysRevLett.105.196403) PMID: [21231189](https://pubmed.ncbi.nlm.nih.gov/21231189/)
62. Wang CS, Pickett WE. Density-Functional Theory of Excitation Spectra of Semiconductors: Application to Si. *Phys Rev Lett*. 1983; 51:597–600. doi: [10.1103/PhysRevLett.51.597](https://doi.org/10.1103/PhysRevLett.51.597)
63. Godby RW, Schlüter M, Sham LJ. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys Rev B*. 1988; 37:10159–10175. doi: [10.1103/PhysRevB.37.10159](https://doi.org/10.1103/PhysRevB.37.10159)
64. Materials Project;. [https://materialsproject.org/docs/calculations#Band\\_gaps](https://materialsproject.org/docs/calculations#Band_gaps).
65. Hedin L. New Method for Calculating the One-Particle Green’s Function with Application to the Electron-Gas Problem. *Phys Rev*. 1965; 139:A796–A823. doi: [10.1103/PhysRev.139.A796](https://doi.org/10.1103/PhysRev.139.A796)
66. Heyd J, Peralta JE, Scuseria GE, Martin RL. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *The Journal of Chemical Physics*. 2005; 123(17):174101. doi: [10.1063/1.2085170](https://doi.org/10.1063/1.2085170) PMID: [16375511](https://pubmed.ncbi.nlm.nih.gov/16375511/)
67. Setyawan W, Gaume RM, Lam S, Feigelson RS, Curtarolo S. High-Throughput Combinatorial Database of Electronic Band Structures for Inorganic Scintillator Materials. *ACS Combinatorial Science*. 2011; 13(4):382–390. doi: [10.1021/co200012w](https://doi.org/10.1021/co200012w) PMID: [21644557](https://pubmed.ncbi.nlm.nih.gov/21644557/)

68. Geilhufe RM, Borysov SS, Bouhon A, Balatsky AV. Data Mining for 3D Organic Dirac Materials: Focus on Space Group #19. arXiv:161104316. 2016;
69. Geilhufe RM, Bouhon A, Borysov SS, Balatsky AV. Three-dimensional organic Dirac-line materials due to nonsymmorphic symmetry: A data mining approach. *Physical Review B*. 2017; 95(4):041103. doi: [10.1103/PhysRevB.95.041103](https://doi.org/10.1103/PhysRevB.95.041103)