DENOPTIM: Software for Computational de Novo Design of Organic and Inorganic Molecules

Marco Foscato,[#,‡] Vishwesh Venkatraman,[†,‡] and Vidar R. Jensen[#,]*

[†]Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

[#]Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway

ABSTRACT: A general-purpose software package, termed DE Novo OPTimization of In/organic Molecules (DENOPTIM), for *de novo* design and virtual screening of functional molecules is described. Molecules of any element and kind, including metastable species and transition states, are handled as chemical objects that go beyond valence-rules representations. Synthetic accessibility of the generated molecules is ensured via detailed control of the kinds of bonds that are allowed to form in the automated molecular building process. DENOPTIM contains a combinatorial explorer, for screening, and a genetic algorithm for global optimization of user-defined properties. Estimates of these properties may be obtained, to form the fitness function (figure of merit or scoring function), from external molecular modeling programs via shell scripts. Examples of a range of different fitness functions and DENOPTIM applications, including an easy-to-do test case, are described. DENOPTIM is available as Open Source from https://github.com/denoptim-project/DENOPTIM.

INTRODUCTION

Computational-driven design and discovery of molecules and materials with desired properties has been a long-sought goal that is, little by little, being realized by spectacular improvements in both computational hardware and algorithms for molecular and materials modeling and prediction. In drug discovery, for instance, a primary goal is the identification of ligands with high binding affinity and selectivity for the protein receptor, and this identification is accelerated by a range of computerized methods.[1] Although there is a similar need for functional non-drug compounds and materials to help achieving sustainable energy and chemical manufacturing, the discovery of these compounds have to a greater extent relied on chemical intuition and serendipity. However, this trend is changing, and modern computational methods, some augmented by artificial intelligence (AI),[2-3] are increasingly being used in the screening of compounds and materials with desired properties.[4-5]

However, such screenings cannot, due to the vastness of chemical space, involve all possible molecules. Instead, structures may be designed in an inverse fashion from the intended properties,[6] or the chemical space may be traversed more efficiently, favoring higher-scoring structures. The latter is the idea behind *de novo* design, which is a standard tool in drug design:[7] Molecules are built, automatically, by assembling building blocks (atoms or fragments).[8] Next, the fitness, or scoring function, of each candidate is calculated and this information is used to traverse the chemical space via global optimization. To ensure that each such candidate is sufficiently realistic, the building blocks may be assembled to synthetically accessible molecules using retrosynthetic- and reaction-based approaches[9-11] or by applying explicit connection rules for bond formation.[12-13]

Most of the existing *de novo* methods have been developed for the assembly of organic, drug-like compounds,[14-15] and constructing reasonable non-drug like molecules is, in general, more challenging due to the variety of elements and their different chemical reactivities and geometrical preferences. This is especially true for transition-metal (TM) compounds, and the automatic building and screening of inorganic compounds in general and TM compounds in particular requires specifically adapted methods.[16-19] To illustrate, HostDesigner is a tool aimed at designing metal ion binding sites and other host-guest systems,[19] MolSimplify is a toolkit for screening of inorganic molecules and intermolecular complexes,[16] and an evolutionary algorithm (EA) has recently been developed for the design of porous organic cages and other supramolecules.[20]

Broad and general applicability to all kinds of molecules and design problems is the main goal behind our development of an EA-based *de novo* design method. A prototype of this method reproduced the known relative performance among ruthenium-based olefin metathesis catalysts,[21] but produced many exotic and synthetically inaccessible molecules in addition to low-quality starting structures (for fitness evaluation). This software has since been completely rewritten and equipped with routines to handle synthetic accessibility,[22] assembly of 3D fragments directly into 3D molecules of high quality,[23] and the closure and rupture of cyclic structures (as in metal chelates).[24] The resulting program, baptized DE Novo OPTimization of In/organic Molecules (DENOPTIM), has been applied to a variety of design problems, and predicted the first *de novo* designed inorganic molecule experimentally verified to reflect the intended property.[25] DENOPTIM is available as Open Source software from https://github.com/denoptim-project.

METHODS

Software Design

DENOPTIM consists of a series of modules, each associated with a specific set of functionalities. Input parameters are defined in the main interface to DENOPTIM, an ASCII file (termed Parameters in Figure 1) containing keywords that may or may not require the specification of a value (syntax: `keyword=value`). This ASCII file also defines the locations of further input files, such as the list of fragments, the connection rules, and the Fitness Provider (see Figure 1). The latter is an external BASH script that is called by a fitness evaluation routine and which defines the figure of merit of each candidate molecule. The output from DENOPTIM is an organized series of SDF files, each containing a candidate molecule, its Cartesian coordinates, connectivity, and associated data fields such as the fitness, SMILES/InCHi encoding and other properties that the user may choose to include via the Fitness Provider. Handling of molecular structures is facilitated by routines from the Chemistry Development Kit (CDK),[26] with additional support for some input/output operations from Apache libraries.[27] DENOPTIM is written in Java, allowing for execution on a variety of operating systems and computer architectures.

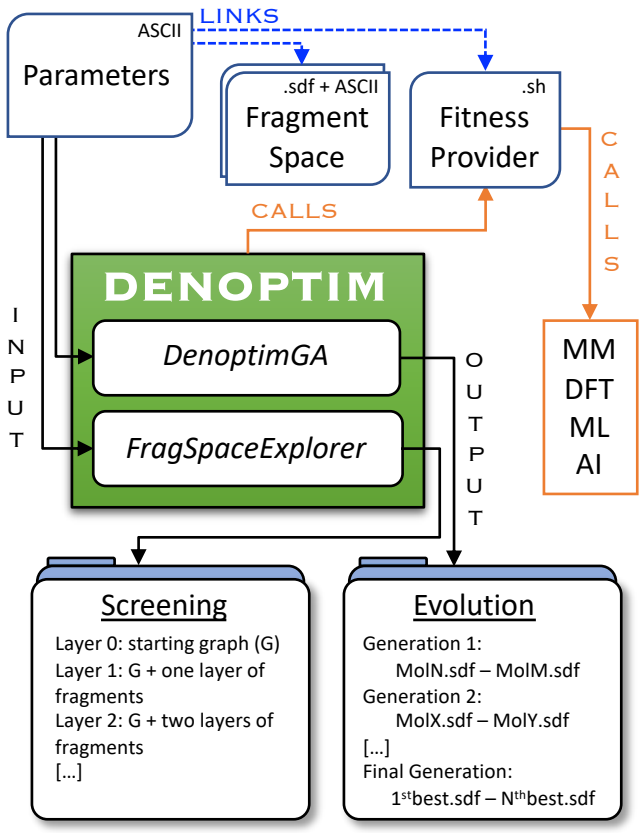

Figure 1: Relation between the main DENOPTIM modules (*DenoptimGA* and *FragSpaceExplorer*), the input parameters file, and the Fitness Provider.

Representation of chemical objects

In cheminformatics, molecules are typically represented as graphs.[28] DENOPTIM represents a chemical object as a graph, G = (T,Fc), resulting from a spanning tree (T), which is a collection of vertices (i.e., building blocks/molecular fragments) and edges (i.e., connections between pairs of fragments), and a collection of fundamental cycles (Fc) corresponding to chords between vertices of the spanning tree. A particular feature of DENOPTIM is that edges may represent either any kind of bond or any spatial relationship between the fragments involved. This allows handling of any molecular and supramolecular entities. For use in DENOPTIM, molecular fragments are generated by the fragmentation of compounds taken from databases (e.g.,

ChEMBL[29] and crystal structure databases)[30] or computational studies. Each fragment is characterized by its chemical constitution (i.e., the atoms, bonds, and charges) and a set of attachment points (APs) that define how (i.e., formal bond order and geometry) and where (i.e., from which atoms/pseudo-atoms) the fragment can form connections. DENOPTIM comes with an accessory tool for generation of fragments (https://github.com/denoptim-project/GM3DFragmenter) particularly suited (e.g., allowing for multihapto bonding) for organometallic chemistry, but fragments may also be generated by other tools, such as RDKit[31] and eMolFrag.[32]

DENOPTIM fragments, while all being formatted equally, have four different uses:

- *Scaffolds* are the roots of the tree-like molecular structures.
- *Standard fragments* are used to expand or modify molecules but not to start new ones.
- *Ring-closing placeholders* serve to saturate APs that may set up chords to form multi-fragment rings.
- *Capping groups* are fragments with only one AP that are used to saturate open valences.

DENOPTIM currently supports two strategies for handling APs: In the valence-based approach, each AP is one of the user-defined free valences of a given atom. Combination of APs are controlled only by the number of open valences on the atoms. In the class-based approach, each AP is annotated with a label (a string) specifying the so-called AP class (APClass).[22] The AP classes are typically used to encode information about the connected chemical environment during fragmentation of existing molecules. Combination of APs is controlled by a user-defined APClass compatibility matrix. Hence, valence rules do not apply, and atypical molecules and supramolecules can be formed.

Fragments may contain cycles, but since such cycles are intra-vertex, they do not result in a cyclic graph G. However, chains of vertices (i.e., cyclic or acyclic fragments) can form new, multi-fragment rings when chords are defined accordingly in the graph G.[24] This formalism allows for definition of any type of ring, as well as on-the-fly ring formation and rupture during design studies.

Molecular Assembly

Fragments are attached layer-by-layer on a scaffold generating a tree-like structure. DENOPTIM detects APs that are related by high topological similarity (same APClass, AP source element, type and number of nearest-neighbor atoms) and performs the same action on all such symmetry-related APs. The user, in addition to controlling the probability of symmetric operations, may introduce or prevent this symmetry by using identical or different AP classes for otherwise topologically identical APs.

Synthetic accessibility is controlled by a set of pre-defined connection rules, specified in the above-mentioned APClass compatibility matrix. Similar compatibility rules also apply to the ring-closure placeholders, between which chords may be added to the spanning tree to form cycles. After definition of the rings, the class-based building scheme finalizes a molecule by appending a capping fragment with a single AP of a matching class (as defined by the compatibility matrix) to each of the free APs and replacing unused ring-closure placeholders. Finally, a unique identifier, such as an InCHi key or a custom identifier, is assigned to each completed molecule to allow enumeration and removal of duplicates.

Fitness Evaluation

DENOPTIM handles the calculation of fitness, or figure of merit. A fitness, or scoring, function may, for instance, be a calculated measure of the activation energy of a candidate catalyst[21] or the spin instability of a candidate spin-crossover compound; see the Application Examples section below and the Supporting Information (SI).[24] A single fitness evaluation may involve several computational chemistry tasks and individual calculations, such as a conformational search followed by geometry optimization, refinement and property calculations of several of the identified conformers. To ensure generality and flexibility with respect to the nature of the fitness, DENOPTIM calls an external BASH script, here referred to as the Fitness Provider, in which the user defines the details of the fitness calculation. Thus, any kind of fitness evaluations, ranging from estimates based on machine learning to density functional theory calculations, is supported. Each Fitness Provider job is entered in the task list and monitored for completion. Once completed, the Fitness Provider reports the fitness value in a dedicated data field ("<FITNESS>") in the SDF file. Alternatively, a failed fitness evaluation will be indicated by the "<MOL_ERROR>" field.

Finally, DENOPTIM currently only supports a single fitness value. Cases in which the fitness reflects different properties must be resolved by summarizing the different fitness components into a single number.[33]

Evolutionary Optimization

The *DenoptimGA* module contains a genetic algorithm for global optimization of molecules. An initial population of candidates is evolved using crossover and mutation to modify existing molecular graphs. Mutation replaces a vertex v_i and all vertices that can be reached from v_i by a directed path in the spanning tree (i.e., the sub-branches of v_i) either by deletion (i.e., no vertex)

or by a different vertex and its sub-branches. The crossover operation swaps such branches between the graphs of two parent molecules.

A third operator, growth, extends the graph by another fragment. In order to prevent molecules from growing infinitely, this extension is associated with a probability p of addition to level L of the spanning-tree, controlled by the parameters λ, σ_1, and σ_2:

- EXP_DIFF: $p = 1 - \frac{1-e^{-\lambda L}}{1+e^{-\lambda L}}$
- TANH: $p = 1 - tanh(\lambda L)$.
- SIGMA: $p = 1 - \frac{1}{1+exp(-\sigma_1(L-\sigma_2))}$.

Additional constraints, such as a maximum number of atoms, rotatable bonds, and molecular weight, can be introduced to filter the candidates prior to their fitness evaluation.

Finally, fitness calculations may be expedited on multiprocessor platforms via two multi-threading schemes. In the first such scheme, fitness calculations for a group of candidates are submitted in parallel. No new fitness task is submitted until all the tasks of the previous group are completed. A second scheme, termed asynchronous mode, reduces the waiting time by continuous submission so that new fitness tasks may be launched as soon as idle CPUs are available.

Fragment Space Explorer

The *FragSpaceExplorer* module allows for systematic exploration of the chemical space, and iterates over all fragment combinations that can be obtained from a given set of starting points (fragments or graphs) and a given fragment space. Only one fragment is appended to each AP, thus each root is decorated by fragments one layer at a time. The graphs that represent finished

entities are then submitted to fitness evaluation following the asynchronous parallelization mode. Next, the graphs belonging to level *L* are used as roots when creating the graphs of level *L+1*.

APPLICATION EXAMPLES

Examples where DENOPTIM or earlier versions of the software have been used in *de novo* molecular design are described briefly in the following. More detailed descriptions, including graphical illustrations, are available in the SI.

Olefin metathesis catalysts. In its first implementation, lacking control of synthetic accessibility,[22] advanced 3D features,[23] and handling of rings,[24] the genetic algorithm still reproduced the historical transition from the first-generation, phosphine-based Grubbs-type olefin metathesis catalysts to the more active second-generation, *N*-heterocyclic carbene-based catalysts.[21] The catalytic activity was reflected in the fitness via a quantitative structure–activity (QSAR) model that correlated semi-empirically calculated properties of the catalytically active ruthenium complex with the barrier height of the reaction. Each fitness calculation thus consisted of a semi-empirical (PM6) geometry optimization.

Metal-free dyes for dye-sensitized solar cells.[34] DENOPTIM was used to design phenothiazine dyes for dye-sensitized solar cells.[35] The power conversion efficiencies (PCEs) of the candidate dyes were estimated by a linear quantitative structure–property relationship (QSPR) model built from over 100 known phenothiazines. Several dyes with efficiencies greater than 9% could be identified, a close to 2% increase with respect to the experimentally determined PCEs of the best metal-free sensitizers reported at the time. Similarly, improved coumarin dyes have been designed using DENOPTIM in conjunction with a fitness function

based on a QSPR model fitted to the product of the current and voltage observed for 49 molecules.[36]

Polymers with high refractive index. Optical applications require polymers with high refractive index ($n_D > 1.70$), good thermal stability, and solubility in selected solvents. Accordingly, designing such polymers requires simultaneous optimization of multiple properties. New monomers were designed using the refractive index as the primary figure of merit, while machine-learning (ML) estimates of the temperature of glass transition (T_g) and thermal decomposition (T_d) were used to exclude candidates. Subsequent solubility-based filtering led to candidate monomers matching several complementary properties.[37]

Solvents for CO_2 capture. To identify alternatives to traditional aqueous amine solutions for CO_2 capture, DENOPTIM was used to propose new imidazole-based compounds.[38] The extent to which these solvents are efficient in CO_2 capture to a large extent is determined by their basicity. Thus, the acid dissociation constant (pK_a), estimated by a QSPR model, was used as fitness. More than 8000 unique imidazole derivatives were obtained, and candidates with a QSPR-predicted $pK_a > 9$ were further filtered for other important properties, such as density, viscosity, vapor pressure, and biodegradability.

Tuning of excitation energies. DENOPTIM has been used to fine-tune azobenzene excitation energies.[39] The absorption maximum of each candidate was estimated using time-dependent density functional theory (TD-DFT) and used as fitness function. Starting from 300 known azobenzenes (λ_{max} 318−575 nm), DENOPTIM identified novel compounds with a predicted λ_{max} higher than 600 nm.

Iron spin crossover (SCO) compounds. Using a fitness function consisting of ligand field molecular mechanics (LFMM)[40-41] calculated energies of high- and low-spin $[Fe^{II}]^{2+}$ complexes,

chelating amine ligands were designed to promote spin instability for such complexes.[24] The most promising candidate ligand, 1,1,1-tris(aminomethyl)ethane (TAME), though commercially available, had never been tested for this property. In a subsequent experimental follow-up, the chloride salt of the $[Fe(TAME)_2]^{2+}$ complex was confirmed to possess the predicted spin-crossover property.[42] To our knowledge, this is the first *de novo* designed non-drug compound experimentally confirmed to reflect the intended, in silico optimized property.

Test Case: Ligand Design in Organometallic Complexes. In addition to the above-described, already published applications, a test case was designed to assess the performance of DENOPTIM's genetic algorithm. In practical molecular design, the size of the chemical space precludes the screening of all possible solutions to find the optimum. Instead, near-optimal solutions may be identified using global-optimization methods such as genetic algorithms. The performance of a genetic algorithm can be evaluated as the speedup associated with identifying high-fitness candidates relative to brute-force screening to arrive at the optimal candidate.[43] Thus, we have compared the genetic algorithm (the *DenoptimGA* module) implemented in DENOPTIM with the brute-force method (the *DenoptimRND* module, not shown in Figure 1). The test case does not address a specific research challenge, but illustrates how DENOPTIM may be used to design organometallic complexes. In particular, the test case illustrates the handling of a fitness function dominated by electronic properties, which are often more challenging to evaluate than shape and steric properties. More precisely, as described in more detail in the SI, we set out to design square-planar complexes *trans*-$Pt(X)_2(L)(CO)$ with weak C–O bonds. Since the strength of the C–O bond is determined by the electronic properties of *trans*-$Pt(X)_2(L)$, we used the C–O bond distance as fitness while exploring ligand sets [X,L] built from a selection of four anionic ligands (X) and about a thousand *de novo* generated dative

ligands (L). As **Figure 2** shows, random selection (brute force) takes much longer to reach high fitness values than the genetic algorithm.

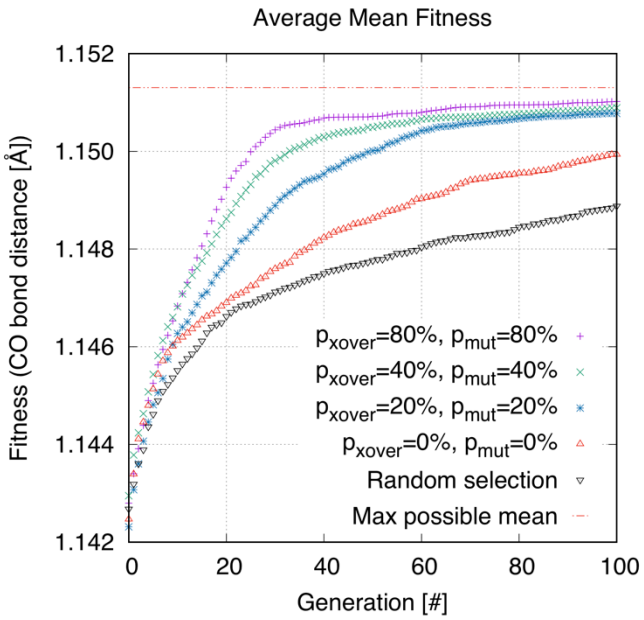

Figure 2: Average mean fitness from ten independent replicas of evolutionary experiments where new candidates were generated either using *DenoptimGA* (probability of crossover, p_{xover}, and probability of mutation, p_{mut}), or by random selection from the full list of candidates (*DenoptimRND*).

Conclusions

DENOPTIM is a software package for *de novo* design and virtual screening of functional molecules of any kind. Handling of organic, inorganic and organometallic molecules as well as metastable species and transition states is achieved via chemical objects that go beyond standard valence-rules representation, with the synthetic accessibility controlled by a compatibility matrix determining what kinds of bonds are allowed to form. The fitness evaluation is also flexible, and is easily extended to use any kind of fitness function from external programs. DENOPTIM has

been used in a range of different design problems involving organic, inorganic and organometallic compounds. Moreover, a test case demonstrates the superior performance of the genetic algorithm over brute-force screening. DENOPTIM is distributed under the GNU Affero GPL v3 license and is available as Open Source software from https://github.com/denoptim-project/DENOPTIM.

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge.

Supplementary details for each application example, description of the test case settings, strategy and results (PDF).

AUTHOR INFORMATION

Corresponding Author

*E-mail: vidar.jensen@uib.no

ORCID

Marco Foscato: 0000-0001-7762-6931

Vishwesh Venkatraman: 0000-0001-7609-2245

Vidar R. Jensen: 0000-0003-2444-3220

Author Contributions

‡V.V. and M.F. contributed equally. All of the authors approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

DEDICATION

Dedicated to the memory of Prof. Bjørn K. Alsberg.

ACKNOWLEDGMENT

The authors are grateful for the many inspiring discussions and the fruitful collaboration with Prof. Bjørn K. Alsberg on *de novo* design in general and DENOPTIM in particular. Prof. Alsberg took part in the early planning of this contribution but sadly passed away in December-2017. The Research Council of Norway (RCN) is acknowledged for financial support via the eVITA (Grant 205273), CLIMIT (233776), and FRINATEK (262152 and 262370) programs and for CPU and storage resources granted through the NOTUR (NN2506K) and NORSTORE (NS2506K) supercomputing programs. Ms. Inger Johanne Fjellanger (M.Sc.) is thanked for help with the graphical material.

REFERENCES

1. Jorgensen, W. L., The Many Roles of Computation in Drug Discovery. *Science* **2004,** *303*, 1813-1818.

2. Warren, J. A., The Materials Genome Initiative and Artificial Intelligence. *MRS Bulletin* **2018,** *43*, 452-457.

3. Sanchez-Lengeling, B.; Aspuru-Guzik, A., Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018,** *361*, 360-365.

4. Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O., The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013,** *12*, 191-201.

5. Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O., Rapid Virtual Screening of Enantioselective Catalysts Using CatVS. *Nat. Catal.* **2019,** *2*, 41-45.

6. Weymuth, T.; Reiher, M., Inverse Quantum Chemistry: Concepts and Strategies for Rational Compound Design. *Int. J. Quantum Chem.* **2014,** *114*, 823-837.

7. Schneider, G., Future de Novo Drug Design. *Mol. Inf.* **2014,** *33*, 397-402.

8. Kutchukian, P. S.; Shakhnovich, E. I., De Novo Design: Balancing Novelty and Confined Chemical Space. *Expert Opin. Drug Discov* **2010,** *5*, 789-812.

9. Pirok, G.; Mate, N.; Varga, J.; Szegezdi, J.; Vargyas, M.; Dorant, S.; Csizmadia, F., Making “Real" Molecules in Virtual Space. *J. Chem. Inf. Model.* **2006,** *46*, 563-568.

10. Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G., DOGS: Reaction-Driven de Novo Design of Bioactive Compounds. *PLOS Comput. Biol.* **2012,** *8*, e1002380.

11. Patel, H.; Bodkin, M. J.; Chen, B. N.; Gillet, V. J., Knowledge-Based Approach to de Novo Design Using Reaction Vectors. *J. Chem. Inf. Model.* **2009,** *49*, 1163-1184.

12. Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M., RECAP - Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **1998,** *38*, 511-522.

13. Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K. H.; Schneider, G.; Jacoby, E.; Renner, S., Probing the Bioactivity-Relevant Chemical Space of Robust Reactions and Common Molecular Building Blocks. *J. Chem. Inf. Model.* **2012,** *52*, 1167-1178.

14. Devi, R. V.; Sathya, S. S.; Coumar, M. S., Evolutionary Algorithms for de Novo Drug Design – A Survey. *Applied Soft Computing* **2015,** *27*, 543-552.

15. Le, T. C.; Winkler, D. A., A Bright Future for Evolutionary Methods in Drug Design. *ChemMedChem* **2015,** *10*, 1296-1300.

16. Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J., molSimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016,** *37*, 2106-2117.

17. Andronico, A.; Randall, A.; Benz, R. W.; Baldi, P., Data-Driven High-Throughput Prediction of the 3-D Structure of Small Molecules: Review and Progress. *J. Chem. Inf. Model.* **2011,** *51*, 760-776.

18. Sadowski, P.; Baldi, P., Small-Molecule 3D Structure Prediction Using Open Crystallography Data. *J. Chem. Inf. Model.* **2013,** *53*, 3127-3130.

19. Hay, B. P.; Firman, T. K., HostDesigner: A Program for the de Novo Structure-Based Design of Molecular Receptors with Binding Sites that Complement Metal Ion Guests. *Inorg. Chem.* **2002,** *41*, 5502-5512.

20. Berardo, E.; Turcani, L.; Miklitz, M.; Jelfs, K. E., An Evolutionary Algorithm for the Discovery of Porous Organic Cages. *Chem. Sci.* **2018,** *9*, 8513-8527.

21. Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg, B. K., An Evolutionary Algorithm for de Novo Optimization of Functional Transition Metal Compounds. *J. Am. Chem. Soc.* **2012,** *134*, 8885-8895.

22. Foscato, M.; Occhipinti, G.; Venkatraman, V.; Alsberg, B. K.; Jensen, V. R., Automated Design of Realistic Organometallic Molecules from Fragments. *J. Chem. Inf. Model.* **2014,** *54*, 767-780.

23. Foscato, M.; Venkatraman, V.; Occhipinti, G.; Alsberg, B. K.; Jensen, V. R., Automated Building of Organometallic Complexes from 3D Fragments. *J. Chem. Inf. Model.* **2014,** *54*, 1919-31.

24. Foscato, M.; Houghton, B. J.; Occhipinti, G.; Deeth, R. J.; Jensen, V. R., Ring Closure To Form Metal Chelates in 3D Fragment-Based de Novo Design. *J. Chem. Inf. Model.* **2015,** *55*, 1844-1856.

25. Bernhardt, P. V.; Bilyj, J. K.; Brosius, V.; Chernyshov, D.; Deeth, R. J.; Foscato, M.; Jensen, V. R.; Mertes, N.; Riley, M. J.; Tornroos, K. W., Spin Crossover in a Hexaamineiron(II) Complex: Experimental Confirmation of a Computational Prediction. *Chem.Eur. J.* **2018,** *24*, 5082-5085.

26. Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E., The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003,** *43*, 493-500.

27. Apache Commons. https://commons.apache.org/ (accessed 4 June 2019).

28. Engel, T., Representation of Chemical Compounds. In *Chemoinformatics*, John Wiley & Sons, Ltd: 2003.

29. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R., The ChEMBL Database in 2017. *Nucleic Acids Res.* **2016,** *45*, D945-D954.

30. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Cryst. B* **2016,** *72*, 171-179.

31. RDKit: Open-source cheminformatics. https://www.rdkit.org/ (accessed 2-June-2019).

32. Liu, T.; Naderi, M.; Alvin, C.; Mukhopadhyay, S.; Brylinski, M., Break Down in Order To Build Up: Decomposing Small Molecules for Fragment-Based Drug Design with eMolFrag. *J. Chem. Inf. Model.* **2017,** *57*, 627-631.

33. Costa, N. R.; Lourenço, J., Multiresponse Problems: Desirability and Other Optimization Approaches. *J. Chemometr.* **2016,** *30*, 702-714.

34. Ji, J.-M.; Zhou, H.; Kim, H. K., Rational Design Criteria for D–π–A Structured Organic and Porphyrin Sensitizers for Highly Efficient Dye-Sensitized Solar Cells. *J. Mater. Chem. A* **2018,** *6*, 14518-14545.

35. Venkatraman, V.; Foscato, M.; Jensen, V. R.; Alsberg, B. K., Evolutionary de Novo Design of Phenothiazine Derivatives for Dye-Sensitized Solar Cells. *J. Mater. Chem. A* **2015,** *3*, 9851-9860.

36. Venkatraman, V.; Abburu, S.; Alsberg, B. K., Artificial Evolution of Coumarin Dyes for Dye Sensitized Solar Cells. *Phys. Chem. Chem. Phys.* **2015,** *17*, 27672-27682.

37. Venkatraman, V.; Alsberg, K. B., Designing High-Refractive Index Polymers Using Materials Informatics. *Polymers* **2018,** *10*.

38. Venkatraman, V.; Gupta, M.; Foscato, M.; Svendsen, H. F.; Jensen, V. R.; Alsberg, B. K., Computer-Aided Molecular Design of Imidazole-Based Absorbents for CO_2 Capture. *Int. J. Greenh. Gas Con.* **2016,** *49*, 55-63.

39. Abburu, S.; Venkatraman, V.; Alsberg, B. K., TD-DFT Based Fine-Tuning of Molecular Excitation Energies Using Evolutionary Algorithms. *RSC Advances* **2016,** *6*, 3661-3670.

40. Deeth, R. J., The Ligand Field Molecular Mechanics Model and the Stereoelectronic Effects of d and s Electrons. *Coord. Chem. Rev.* **2001,** *212*, 11-34.

41. Foscato, M.; Deeth, R. J.; Jensen, V. R., Integration of Ligand Field Molecular Mechanics in Tinker. *J. Chem. Inf. Model.* **2015**.

42. Bernhardt, P. V.; Bilyj, J. K.; Brosius, V.; Chernyshov, D.; Deeth, R. J.; Foscato, M.; Jensen, V. R.; Mertes, N.; Riley, M. J.; Törnroos, K. W., Spin Crossover in a Hexaamineiron(II) Complex: Experimental Confirmation of a Computational Prediction. *Chem. Eur. J.* **2018,** *24*, 5082-5085.

43. Sugihara, K., Measures for Performance Evaluation of Genetic Algorithms. *Proc. 3rd Joint Conference on Information Sciences, Research Triangle Park* **1997**, 172-175.

For Table of Contents Use Only

DENOPTIM: Software for Computational de Novo Design of Organic and Inorganic Molecules

Marco Foscato,[#,‡] *Vishwesh Venkatraman,*[†,‡] *and Vidar R. Jensen*[#,*]

[†]Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

[#]Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway