

# Digital Discovery

rsc.li/digitaldiscovery

The Royal Society of Chemistry is the world's leading chemistry community. Through our high impact journals and publications we connect the world with the chemical sciences and invest the profits back into the chemistry community.

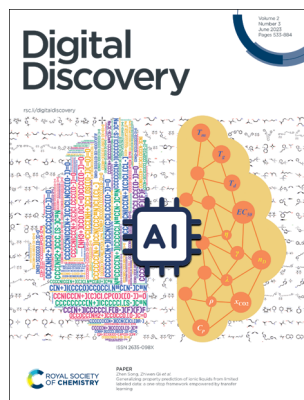
## IN THIS ISSUE

ISSN 2635-098X CODEN DDIIAI 2(3) 533–884 (2023)



### Cover

See Andy S. Anker, Keith T. Butler *et al.*, pp. 578–590. Image reproduced by permission of Andy S. Anker, Keith T. Butler, Manh Duc Le, Toby G. Perring and Jeyan Thiyagalingam from *Digital Discovery*, 2023, 2, 578. Created with the use of Midjourney.



### Inside cover

See Zhen Song, Zhiwen Qi *et al.*, pp. 591–601. Image reproduced by permission of Guzhong Chen and Zhen Song from *Digital Discovery*, 2023, 2, 591.

## PERSPECTIVES

544

### The laboratory of Babel: highlighting community needs for integrated materials data management

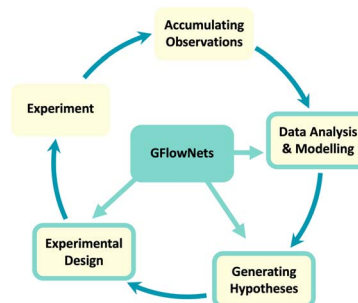
Brenden G. Pelkie and Lilo D. Pozzo\*



557

### GFlowNets for AI-driven scientific discovery

Moksh Jain,\* Tristan Deleu, Jason Hartford, Cheng-Hao Liu, Alex Hernandez-Garcia and Yoshua Bengio



## Editorial Staff

### Editor

Anna Rulka

### Deputy Editor

Audra Taylor

### Editorial Production Manager

Viktoria Titmus

### Assistant Editors

Shwetha Krishna, Michael Whitelaw, Alexander Whiteside

### Editorial Assistant

Samantha Campos

### Publishing Assistant

Brittany Hanlon

### Publisher

Neil Hammond

For queries about submitted articles please contact Viktoria Titmus, Editorial Production Manager in the first instance. E-mail [digitaldiscovery@rsc.org](mailto:digitaldiscovery@rsc.org)

For pre-submission queries please contact Anna Rulka, Editor.

Email [digitaldiscovery-rsc@rsc.org](mailto:digitaldiscovery-rsc@rsc.org)

Digital Discovery (electronic: ISSN 2635-098X) is published 6 times a year by the Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, UK CB4 0WF.

Digital Discovery is a Gold Open Access journal and all articles are free to read. Please email [orders@rsc.org](mailto:orders@rsc.org) to register your interest or contact Royal Society of Chemistry Order Department, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK Tel +44 (0)1223 432398; E-mail: [orders@rsc.org](mailto:orders@rsc.org)

Whilst this material has been produced with all due care, the Royal Society of Chemistry cannot be held responsible or liable for its accuracy and completeness, nor for any consequences arising from any errors or the use of the information contained in this publication. The publication of advertisements does not constitute any endorsement by the Royal Society of Chemistry or Authors of any products advertised. The views and opinions advanced by contributors do not necessarily reflect those of the Royal Society of Chemistry which shall not be liable for any resulting loss or damage arising as a result of reliance upon this material. The Royal Society of Chemistry is a charity, registered in England and Wales, Number 207890, and a company incorporated in England by Royal Charter (Registered No. RC000524), registered office: Burlington House, Piccadilly, London W1J 0BA, UK, Telephone: +44 (0) 207 4378 6556.

### Advertisement sales:

Tel +44 (0) 1223 432246; Fax +44 (0) 1223 426017;

E-mail [advertising@rsc.org](mailto:advertising@rsc.org)

For marketing opportunities relating to this journal, contact [marketing@rsc.org](mailto:marketing@rsc.org)

# Digital Discovery

[rsc.li/digitaldiscovery](http://rsc.li/digitaldiscovery)

Digital Discovery is a gold open access journal publishing top research at the intersection of chemistry, materials science and biotechnology. Blurring the barriers between computation and experimentation, we focus on the integration of digital and automation tools with science, putting data first to ensure reproducibility and faster progress.

## Editorial Board

### Editor in Chief

Alán Aspuru-Guzik, University of Toronto, Canada

### Associate Editors

Jason E. Hein, University of British Columbia, Canada  
Linda Hung, Toyota Research Institute, USA  
Joshua Schrier, Fordham University, USA  
Kedar Hippalgaonkar, Nanyang Technological University, Singapore  
Cesar de la Fuente, University of Pennsylvania, USA

### Members

Yousung Jung, KAIST, South Korea  
Anat Milo, Ben-Gurion University of the Negev, Israel  
Lilo D. Pozzo, University of Washington, USA  
Ekaterina Skorb, ITMO University, Russia

## Advisory Board

Abigail Doyle, University of California Los Angeles, USA  
Ola Engkvist, AstraZeneca and Chalmers University of Technology, Sweden  
Pablo Carbonell, University of Valencia, Spain  
Ian Foster, University of Chicago, USA  
Cecilia Clementi, Freie Universität Berlin, Germany  
Heather Kulik, MIT, USA

Silvana Botti, Friedrich Schiller University Jena, Germany  
Marwin Segler, Microsoft, Germany  
Jan Jensen, University of Copenhagen, Denmark  
Berend Smit, EPFL, Switzerland  
Conor Coley, MIT, USA  
Koji Tsuda, The University of Tokyo, Japan  
Isao Tanaka, Kyoto University, Japan

Shuye Ping Ong, University of California San Diego, USA  
Alexandre Tkatchenko, University of Luxembourg, Luxembourg  
Juan Alegre, Colorado State University, USA

## Information for Authors

Full details on how to submit material for publication in Digital Discovery are given in the Instructions for Authors (available from <http://www.rsc.org/authors>). Submissions should be made via the journal's homepage: [rsc.li/digitaldiscovery](http://rsc.li/digitaldiscovery)

Authors may reproduce/republish portions of their published contribution without seeking permission from the Royal Society of Chemistry, provided that any such republication is accompanied by an acknowledgement in the form: (Original Citation)–Reproduced by permission of the Royal Society of Chemistry.

This journal is © The Royal Society of Chemistry 2023.

Apart from fair dealing for the purposes of research or private study for non-commercial purposes, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988 and the Copyright and Related Rights Regulation 2003, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the Publishers or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency in the UK. US copyright law is applicable to users in the USA.

Registered charity number: 207890



## Using generative adversarial networks to match experimental and simulated inelastic neutron scattering data

# Generalizing property prediction of ionic liquids from limited labeled data: a one-stop framework empowered by transfer learning

## Digital research environments: a requirements analysis

The diagram illustrates the PSDI Core architecture. At the center is the **PSDI Core**. Surrounding it is a ring of features and data sources, each represented by a colored oval:

- Collaboration & management** (Light blue)
- Other Features** (Purple)
- Domain Features** (Light blue)
- Data Features** (Light blue)
- Equipment Interfaces** (Light blue)
- Literature Services** (Light blue)
- Notebooklike Services** (Light blue)
- Generic Features** (Yellow)
- Publishing & Sharing** (Light blue)

Within this ring, specific services are highlighted in colored ovals:

- Semantic Services** (Red)
- Domain based Databases** (Light blue)
- Domain based Services** (Light blue)
- PLVS LANGS** (Light blue)

## Deep learning metal complex properties with natural quantum graphs

The diagram illustrates a workflow for predicting quantum material properties using Geometric and Electronic Information (GEO) and Graph Neural Networks (GNNs).

**Input Data:**

- Transition Metal Complexes:** Shown as a chemical structure of a transition metal complex (e.g.,  $\text{H}_2\text{C}=\text{CH}_2$  coordinated to a metal center).
- NBO Orbitals & SOPA Donor-Acceptor Interactions:** Shown as molecular orbitals and interaction diagrams (LP(d) to BD\*( $\pi^*$ )).

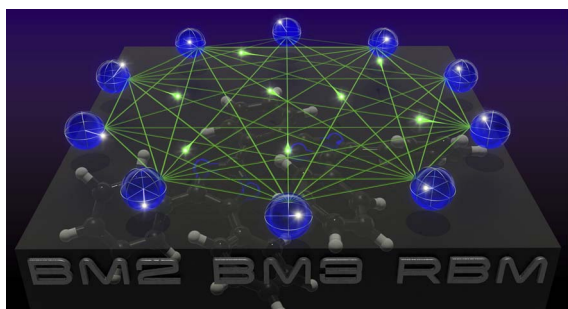
**Processing:**

- The input data is processed into **Natural Quantum Graphs (NatQG)**, which are categorized into *u-NatQG* (undirected) and *d-NatQG* (directed).
- The NatQG are then processed into the **tmQMg dataset**, which includes various material properties (e.g.,  $E_{\text{ZPE}}$ ,  $H$ ,  $S$ ,  $G$ ,  $C_v$ ,  $\epsilon_0$ ,  $\alpha$ ,  $\mu$ ,  $V$ ,  $\dots$ ,  $\text{CSV}$ ).

**Output:**

- The tmQMg dataset is used to train **GNNs Predicting Quantum Properties**, which are visualized as heatmaps showing the relationship between input features (e.g.,  $\alpha$ ,  $\epsilon$ ) and output properties.

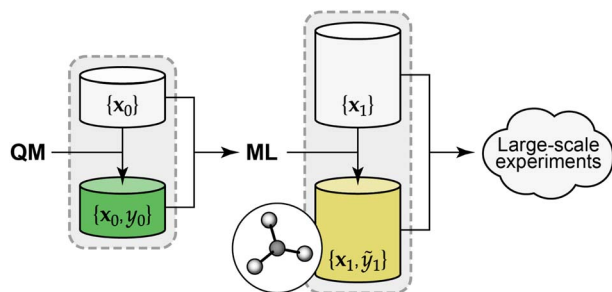
634



### Artificial neural network encoding of molecular wavefunctions for quantum computing

Masaya Hagai,<sup>\*</sup> Mahito Sugiyama, Koji Tsuda and Takeshi Yanai<sup>\*</sup>

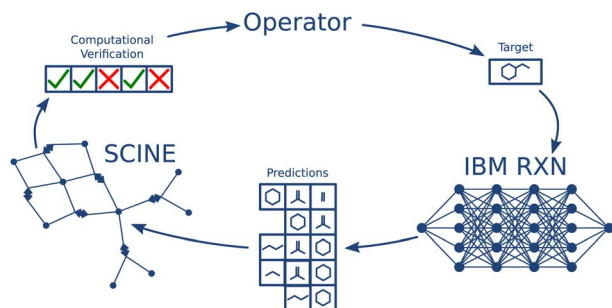
651



### Synthetic data enable experiments in atomistic machine learning

John L. A. Gardner, Zoé Faure Beaulieu and Volker L. Deringer<sup>\*</sup>

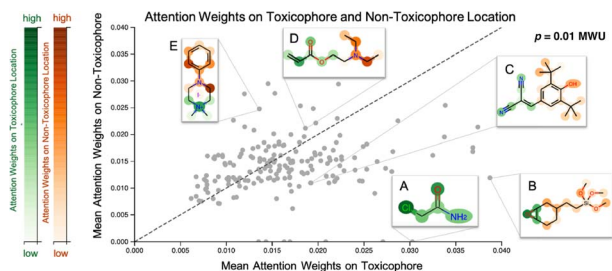
663



### Quantum chemical data generation as fill-in for reliability enhancement of machine-learning reaction and retrosynthesis planning

Alessandra Toniato, Jan P. Unsleber, Alain C. Vaucher, Thomas Weymuth, Daniel Probst, Teodoro Laino<sup>\*</sup> and Markus Reiher<sup>\*</sup>

674



### Chemical representation learning for toxicity prediction

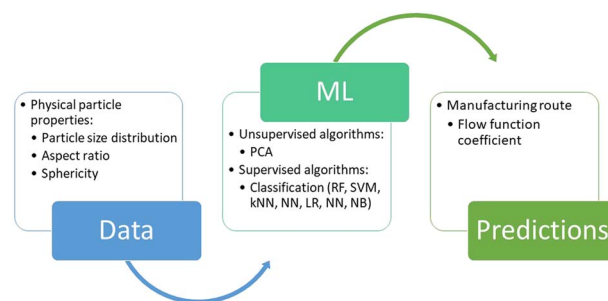
Jannis Born,<sup>\*</sup> Greta Markert, Nikita Janakaraman, Talia B. Kimber, Andrea Volkamer, María Rodríguez Martínez and Matteo Manica



692

## Machine learning approaches to the prediction of powder flow behaviour of pharmaceutical materials from physical properties

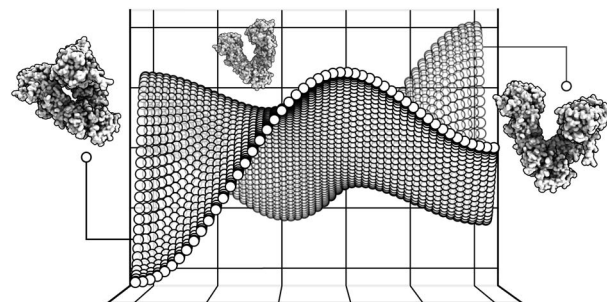
Laura Pereira Diaz, Cameron J. Brown, Ebenezer Ojo, Chantal Mustoe and Alastair J. Florence\*



702

## Beyond ManifoldEM: geometric relationships between manifold embeddings of a continuum of 3D molecular structures and their 2D projections

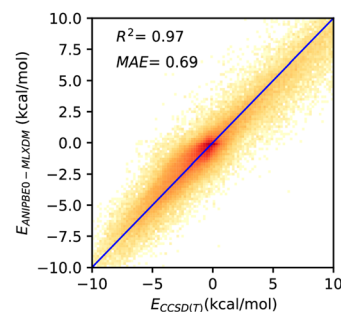
Evan Seitz, Joachim Frank\* and Peter Schwander\*



718

## A neural network potential with rigorous treatment of long-range dispersion

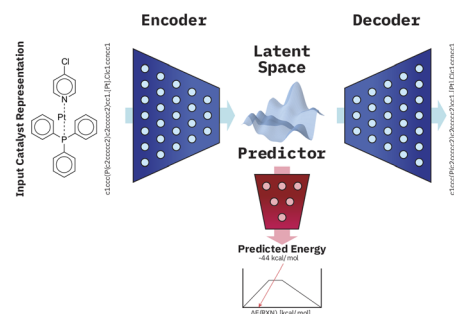
Nguyen Thien Phuc Tu,\* Nazanin Rezajooei, Erin R. Johnson\* and Christopher N. Rowley\*



728

## Designing catalysts with deep generative models and computational data. A case study for Suzuki cross coupling reactions

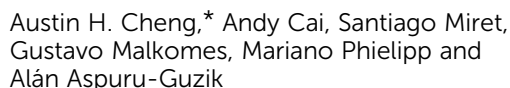
Oliver Schilter,\* Alain Vaucher, Philippe Schwaller and Teodoro Laino



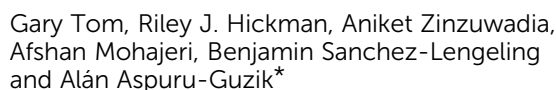




748



759



775

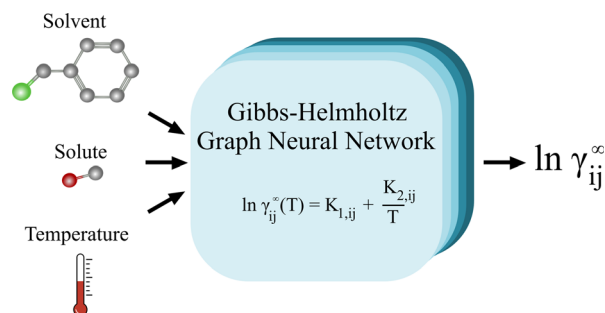


## PAPERS

781

### Gibbs–Helmholtz graph neural network: capturing the temperature dependency of activity coefficients at infinite dilution

Edgar Ivan Sanchez Medina, Steffen Linke, Martin Stoll and Kai Sundmacher\*



799

### Automated electrolyte formulation and coin cell assembly for high-throughput lithium-ion battery research

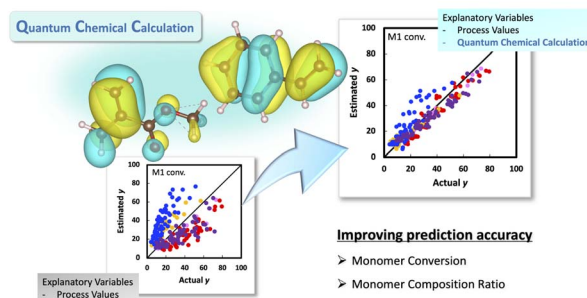
Jackie T. Yik,\* Leiting Zhang,\* Jens Sjölund, Xu Hou, Per H. Svensson, Kristina Edström and Erik J. Berg



809

### Extrapolation performance improvement by quantum chemical calculations for machine-learning-based predictions of flow-synthesized binary copolymers

Shogo Takasuka, Shunto Oikawa, Takayoshi Yoshimura, Sho Ito, Yosuke Harashima, Tomoaki Takayama, Shigehito Asano, Akira Kurosawa, Tetsunori Sugawara, Miho Hatanaka, Tomoyuki Miyao, Takamitsu Matsubara, Yu-ya Ohnishi, Hiroharu Ajiro and Mikiya Fujii\*



819

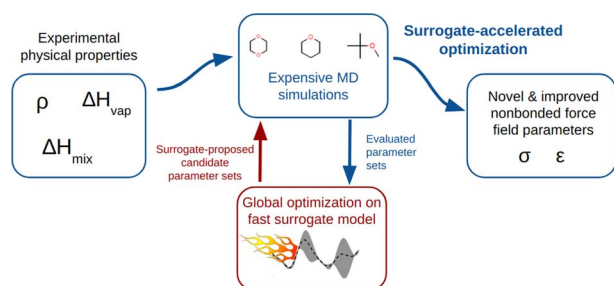
### Predicting ruthenium catalysed hydrogenation of esters using machine learning

Challenger Mishra,\* Niklas von Wolff,\* Abhinav Tripathi, Claire N. Brodie, Neil D. Lawrence, Aditya Ravuri, Éric Brémond, Annika Preiss and Amit Kumar\*



## PAPERS

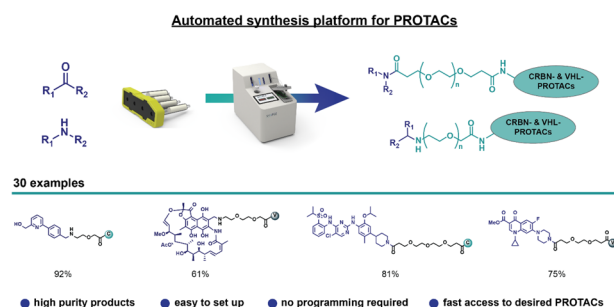
828



## Using physical property surrogate models to perform accelerated multi-fidelity optimization of force field parameters

Owen C. Madin and Michael R. Shirts\*

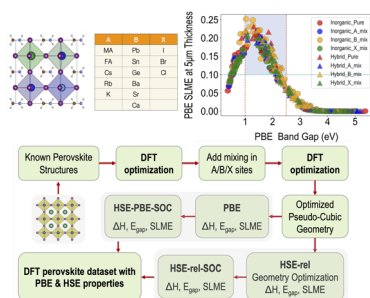
848



## Capsule-based automated synthesis for the efficient assembly of PROTAC like molecules

Samuele Bordi, Tuo Jiang, Anna Konopka, Guillaume Coin, Paula L. Nichols and Benedikt M. Wanner\*

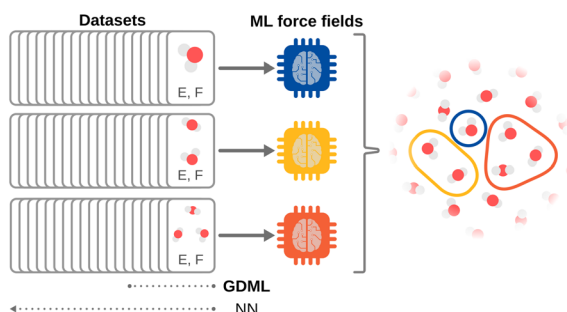
856



## A high-throughput computational dataset of halide perovskite alloys

Jiaqi Yang, Panayotis Manganaris and Arun Mannodi-Kanakkithodi\*

871



## Modeling molecular ensembles with gradient-domain machine learning force fields

Alex M. Maldonado, Igor Poltavsky, Valentin Vassilev-Galindo, Alexandre Tkatchenko\* and John A. Keith\*





## CORRECTIONS

881

**Correction: Molecular sonification for molecule to music information transfer**

Babak Mahjour, Jordan Bench, Rui Zhang, Jared Frazier and Tim Cernak\*

882

**Correction: Latent spaces for antimicrobial peptide design**

Samuel Renaud and Rachael A. Mansbach\*

