

# 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 3(2) 223–438 (2024)



### Cover

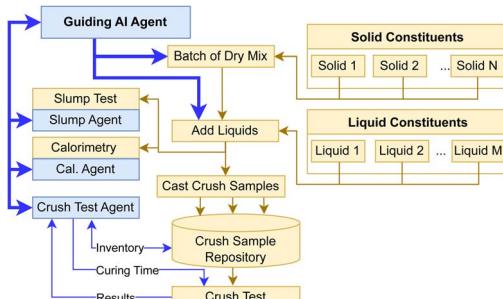
See Michael Levin *et al.*, pp. 249–263. Image reproduced by permission of Patrick McMillen, Levin Lab from *Digital Discovery*, 2024, 3, 249.

## PERSPECTIVES

231

### Autonomous cementitious materials formulation platform for critical infrastructure repair

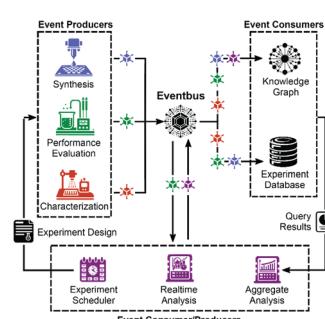
Howie Joress,\* Rachel Cook, Austin McDannald, Mark Kozdras, Jason Hattrick-Simpers, Aron Newman and Scott Jones



238

### Event-driven data management with cloud computing for extensible materials acceleration platforms

Michael J. Statt,\* Brian A. Rohr,\* Dan Guevarra, Santosh K. Suram and John M. Gregoire\*



# Royal Society of Chemistry approved training courses

Explore your options.  
Develop your skills.  
Discover learning  
that suits you.

**Courses in the classroom,  
the lab, or online**

Find something for every  
stage of your professional  
development. Search our  
database by:

- subject area
- location
- event type
- skill level

Members get at least 10% off

Visit [rsc.li/cpd-training](http://rsc.li/cpd-training)

**SAVE  
10%**

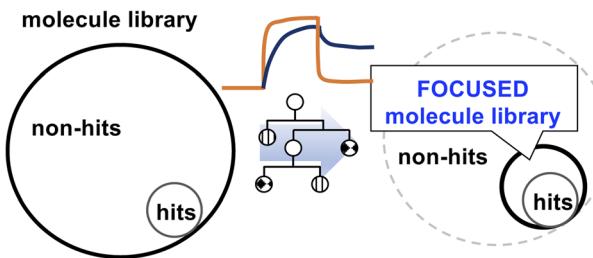


## COMMUNICATION

243

**A machine learning approach toward generating the focused molecule library targeting CAG repeat DNA**

Qingwen Chen, Takeshi Yamada, Asako Murata, Ayako Sugai, Yasuyuki Matsushita and Kazuhiko Nakatani\*

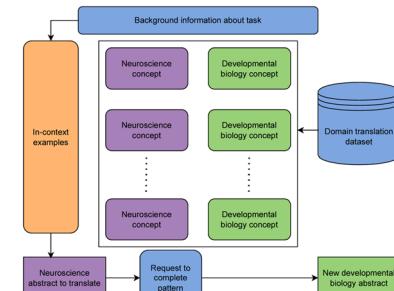


## PAPERS

249

**Machine learning for hypothesis generation in biology and medicine: exploring the latent space of neuroscience and developmental bioelectricity**

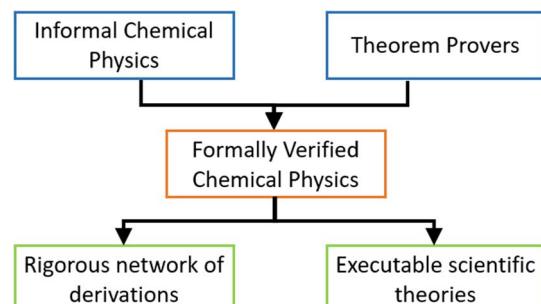
Thomas O'Brien, Joel Stremmel, Léo Pio-Lopez, Patrick McMillen, Cody Rasmussen-Ivey and Michael Levin\*



264

**Formalizing chemical physics using the Lean theorem prover**

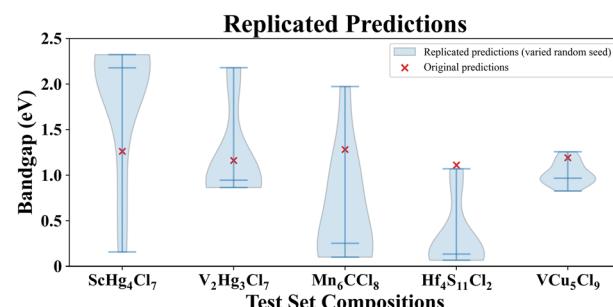
Maxwell P. Bobbin, Samiha Sharlin, Parivash Feyzishendi, An Hong Dang, Catherine M. Wraback and Tyler R. Josephson\*



281

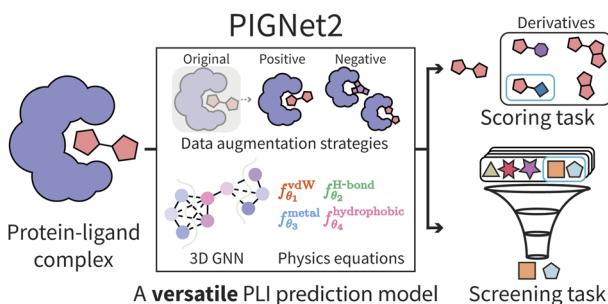
**Reproducibility in materials informatics: lessons from 'A general-purpose machine learning framework for predicting properties of inorganic materials'**

Daniel Persaud, Logan Ward and Jason Hattrick-Simpers\*



## PAPERS

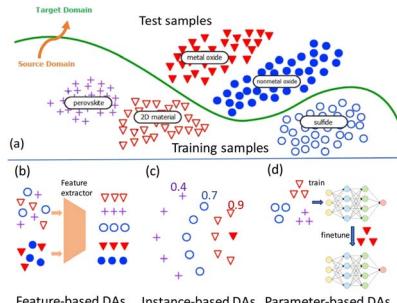
287



## PIGNet2: a versatile deep learning-based protein–ligand interaction prediction model for binding affinity scoring and virtual screening

Seokhyun Moon, Sang-Yeon Hwang, Jaechang Lim and Woo Youn Kim\*

300



## Realistic material property prediction using domain adaptation based machine learning

Jeffrey Hu, David Liu, Nihang Fu and Rongzhi Dong\*

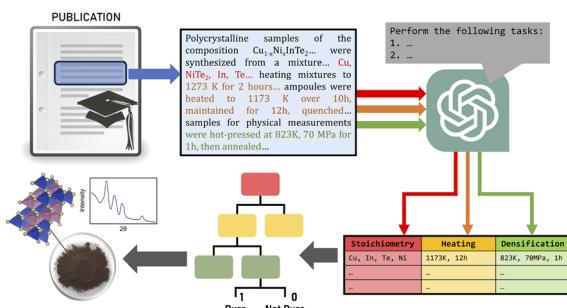
313



## MaScQA: investigating materials science knowledge of large language models

Mohd Zaki, Jayadeva, Mausam and N. M. Anoop Krishnan\*

328



## Harnessing GPT-3.5 for text parsing in solid-state synthesis – case study of ternary chalcogenides

Maung Thway, Andre K. Y. Low, Samyak Khetan, Haiwen Dai, Jose Recatala-Gomez, Andy Paul Chen and Kedar Hippalgaonkar\*

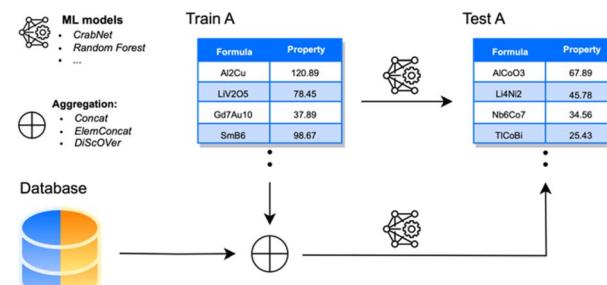


## PAPERS

337

## Not as simple as we thought: a rigorous examination of data aggregation in materials informatics

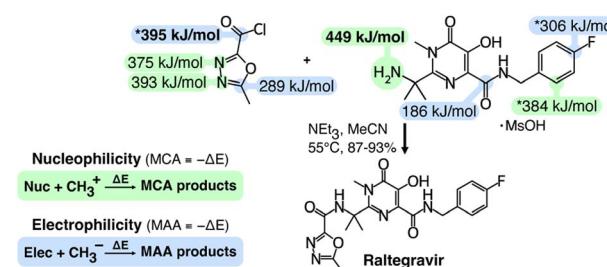
Federico Ottomano, Giovanni De Felice,\* Vladimir V. Gusev and Taylor D. Sparks\*



347

## Automated quantum chemistry for estimating nucleophilicity and electrophilicity with applications to retrosynthesis and covalent inhibitors

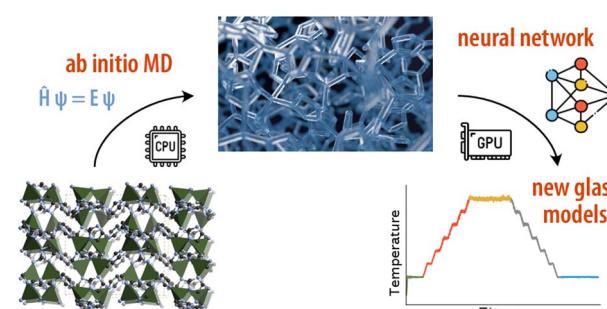
Nicolai Ree, Andreas H. Gölle\* and Jan H. Jensen\*



355

## Machine learning interatomic potentials for amorphous zeolitic imidazolate frameworks

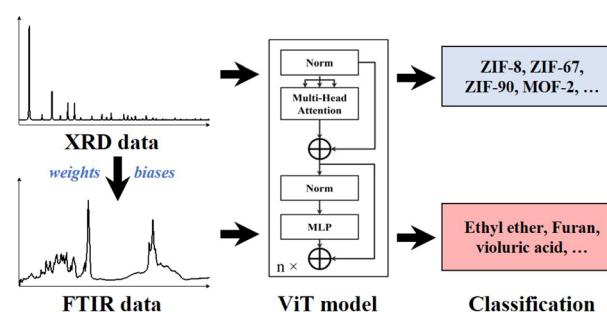
Nicolas Castel, Dune André, Connor Edwards, Jack D. Evans\* and François-Xavier Coudert\*



369

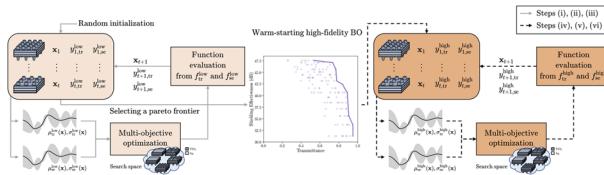
## An interpretable and transferrable vision transformer model for rapid materials spectra classification

Zhenru Chen, Yunchao Xie,\* Yuchao Wu, Yuyi Lin, Shigetaka Tomiya and Jian Lin\*



## PAPERS

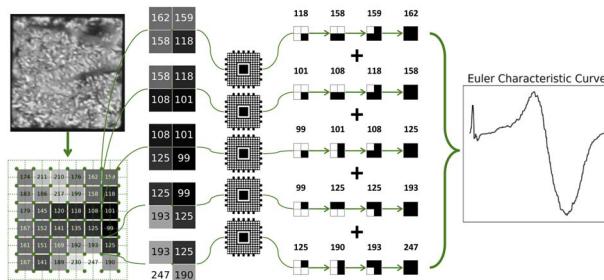
381



## Multi-BOWS: multi-fidelity multi-objective Bayesian optimization with warm starts for nanophotonic structure design

Jungtaek Kim, Mingxuan Li, Yirong Li, Andrés Gómez, Oliver Hinder and Paul W. Leu\*

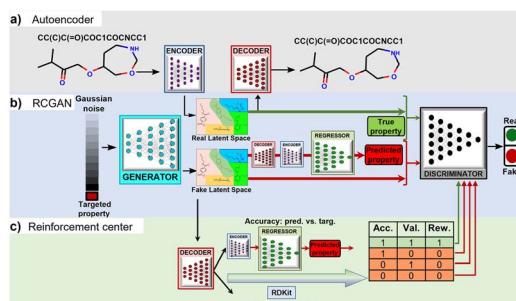
392



## A fast and scalable computational topology framework for the Euler characteristic

Daniel J. Laky and Victor M. Zavala\*

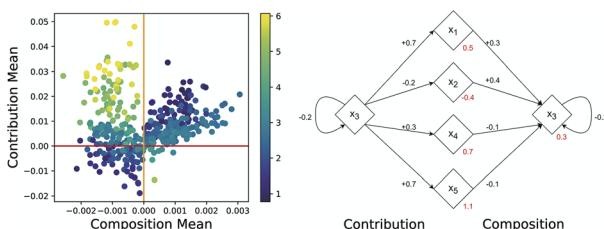
410



## De novo molecule design towards biased properties via a deep generative framework and iterative transfer learning

Kianoosh Sattari, Dawei Li, Bhupalee Kalita, Yunchao Xie, Fatemeh Barmaleki Lighvan, Olexandr Isayev and Jian Lin\*

422



## Understanding the importance of individual samples and their effects on materials data using explainable artificial intelligence

Tommy Liu,\* Zhi Yang Tho and Amanda S. Barnard

