

# 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 5(2) 467–960 (2026)



**Cover**  
See Diandong Tang, Xiaoxia Cai, Xiao Yuan *et al.*, pp. 548–570. Image reproduced by permission of Tianyi Li from *Digital Discovery*, 2026, 5, 548.



**Inside cover**  
See Wei Zhou, Xiao Wang, Xing Lu *et al.*, pp. 523–547. Image reproduced by permission of Wei Zhou from *Digital Discovery*, 2026, 5, 523.

## EDITORIAL

480

### Introduction to the “Accelerate Conference 2023–2024” themed collection

Janine George, Claudiane Ouellet-Plamondon and Kristofer Reyes

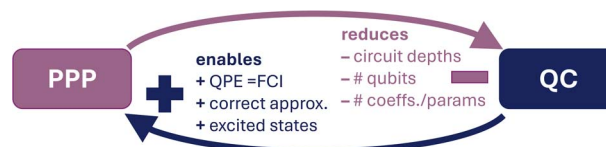


## PERSPECTIVE

482

### The PPP model – a minimum viable parametrisation of conjugated chemistry for modern computing applications

Marcel D. Fabian,\* Nina Glaser and Gemma C. Solomon



# Advance your career in science

with professional recognition that showcases  
your **experience, expertise and dedication**

## Stand out from the crowd

Prove your commitment  
to attaining excellence in  
your field

## Gain the recognition you deserve

Achieve a professional  
qualification that inspires  
confidence and trust

## Unlock your career potential

Apply for our professional  
registers (RSci, RSciTech)  
or chartered status  
(CChem, CSci, CEnv)

## Apply now

[rsc.li/professional-development](https://rsc.li/professional-development)



## OPINION

497

### Accelerating catalytic advancements through the precision of high-throughput experiments & calculations

Jenny G. Vitillo,\* Alán Aspuru-Guzik, Eric Dorskocil, Omar K. Farha, Timur Islamoglu, Heather J. Kulik, Peter M. Margl, Stuart Miller, Jordan Reddel, Aayush R. Singh and Varinia Bernales\*



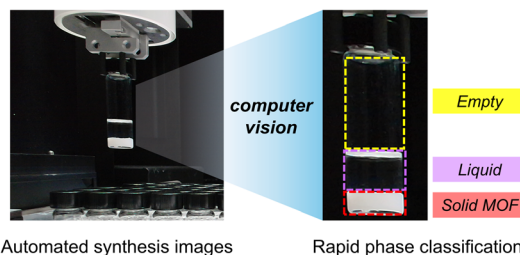
## TUTORIAL REVIEW

510

### Computer vision for high-throughput materials synthesis: a tutorial for experimentalists

Madeleine A. Gaidimas, Abhijoy Mandal, Pan Chen, Shi Xuan Leong, Gyu-Hee Kim, Akshay Talekar, Kent O. Kirlikovali, Kourosh Darvish, Omar K. Farha,\* Varinia Bernales\* and Alán Aspuru-Guzik\*

#### Image-based characterization for high-throughput materials discovery

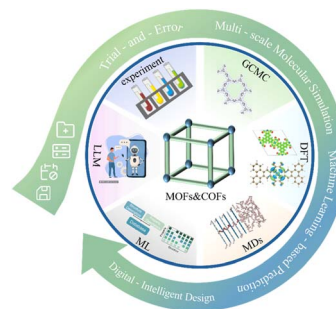


## REVIEW

523

### Advancing metal organic framework and covalent organic framework design via the digital-intelligent paradigm

Bing Ma, Na Qin, Qianqian Yan, Wei Zhou,\* Sheng Zhang, Xiao Wang,\* Lipiao Bao and Xing Lu\*

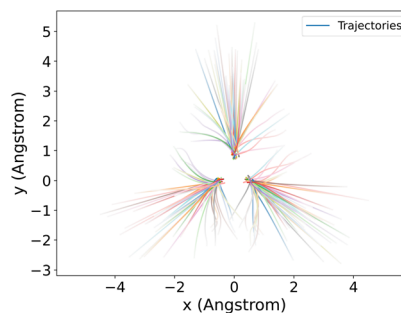


## PAPERS

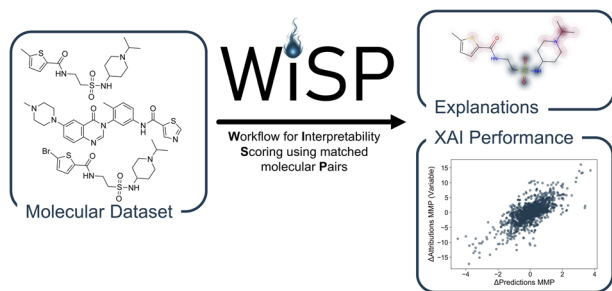
548

### Efficient quantum simulation of non-adiabatic molecular dynamics with precise electronic structure

Tianyi Li, Yumeng Zeng, Qiming Ding, Zixuan Huo, Xiaosi Xu, Jiajun Ren, Diandong Tang,\* Xiaoxia Cai\* and Xiao Yuan\*



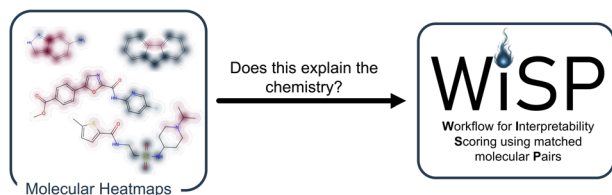
571



### When machine learning models learn chemistry I: quantifying explainability with matched molecular pairs

Kerrin Janssen, Jan M. Wollschläger, Jonny Proppe\* and Andreas H. Göller\*

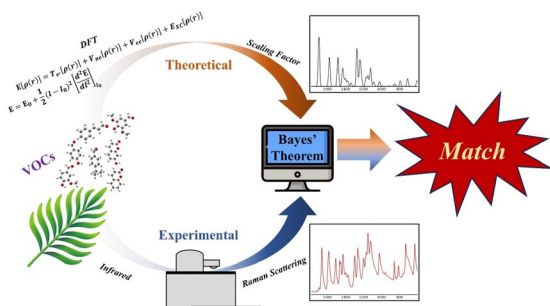
583



### When machine learning models learn chemistry II: applying WISP to real-world examples

Kerrin Janssen, Jan M. Wollschläger, Jonny Proppe\* and Andreas H. Göller\*

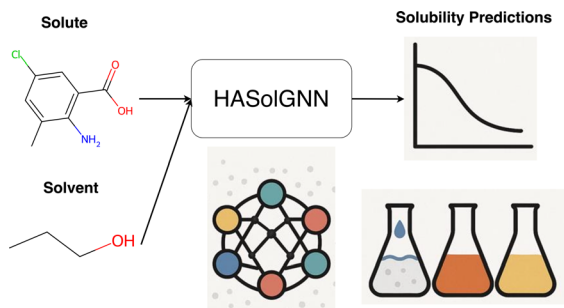
592



### DFT meets Bayesian inference: creating a framework for the assignment of calculated vibrational frequencies

Michael Nicolaou, Hans M. Senn, Emma Gibson,\* Mario González-Jiménez and Laia Vilà-Nadal\*

603



### Hierarchical attention graph learning with LLM enhancement for molecular solubility prediction

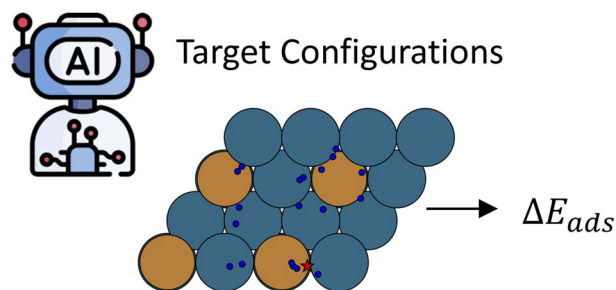
Yangxin Fan, Yinghui Wu, Roger H. French, Danny Perez, Michael G. Taylor and Ping Yang\*



617

### Adsorb-Agent: autonomous identification of stable adsorption configurations via a large language model agent

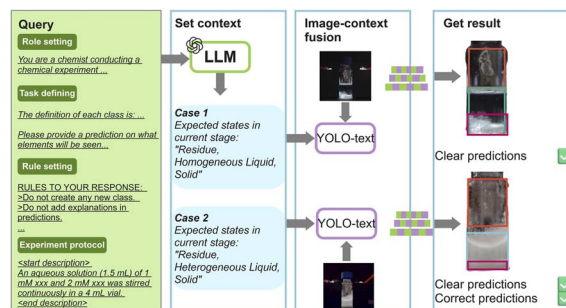
Janghoon Ock, Radheesh Sharma Meda, Tirtha Vinchurkar, Yayati Jadhav and Amir Barati Farimani\*



630

### Context-aware computer vision for chemical reaction state detection

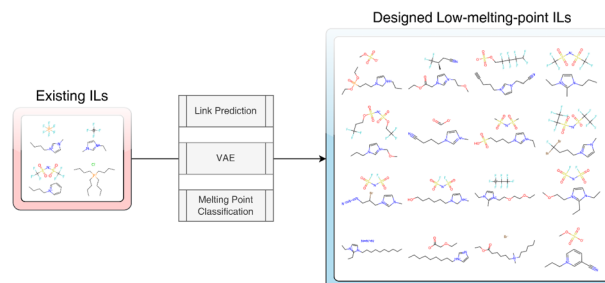
Junru Ren, Abhijoy Mandal, Rama El-khawaldeh, Shi Xuan Leong, Jason Hein, Alán Aspuru-Guzik, Lazaros Nalpantidis and Kourosh Darvish\*



643

### Deep learning-enabled discovery of low-melting-point ionic liquids

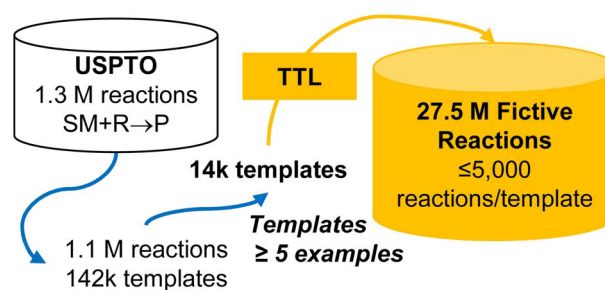
Gaopeng Ren, Austin M. Mroz, Frederik Philippi, Tom Welton and Kim E. Jelfs\*



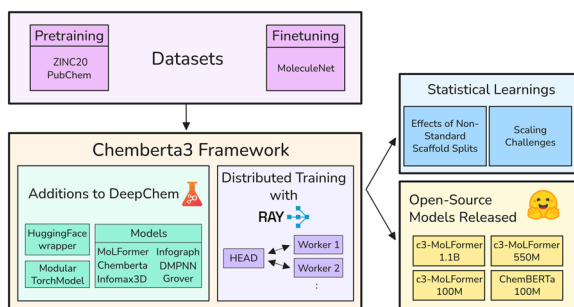
653

### Data augmentation in a triple transformer loop retrosynthesis model

Yves Grandjean, David Kreutter and Jean-Louis Reymond\*



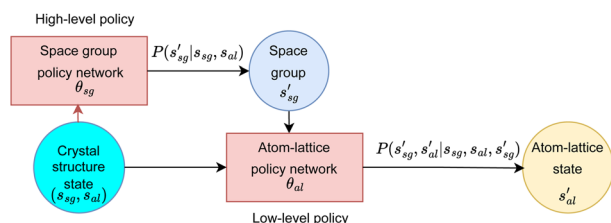
662



### ChemBERTa-3: an open source training framework for chemical foundation models

Riya Singh, Aryan Amit Barsainyan, Rida Irfan, Connor Joseph Amorin, Stewart He, Tony Davis, Arun Thiagarajan, Shiva Sankaran, Seyone Chithrananda, Walid Ahmad, Derek Jones, Kevin McLoughlin, Hyojin Kim, Anoushka Bhutani, Shreyas Vinaya Sathyanarayana, Venkat Viswanathan, Jonathan E. Allen and Bharath Ramsundar\*

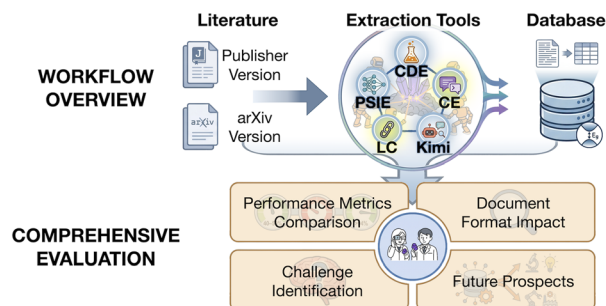
686



### Efficient symmetry-aware materials generation via hierarchical generative flow networks

Tri Minh Nguyen,\* Sherif Abdulkader Tawfik, Truyen Tran, Sunil Gupta, Santu Rana and Svetha Venkatesh

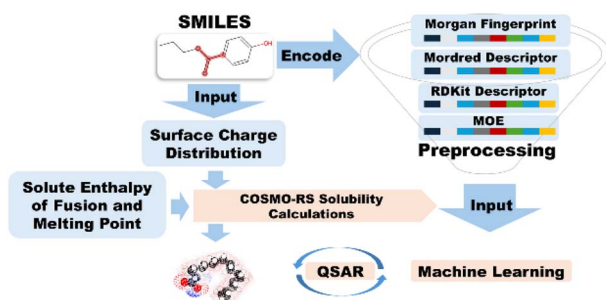
698



### Optimizing data extraction from materials science literature: a study of tools using large language models

Wenkai Ning, Musen Li,\* Jeffrey R. Reimers\* and Rika Kobayashi\*

716



### A case study on hybrid machine learning and quantum-informed modelling for solubility prediction of drug compounds in organic solvents

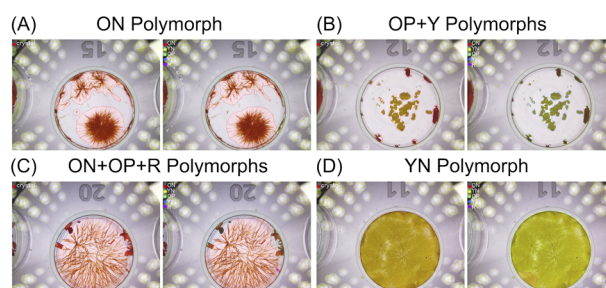
Weiling Wang, Isabel Cooley, Morgan R. Alexander, Ricky D. Wildman, Anna K. Croft and Blair F. Johnston\*



734

### AI-driven robotic crystal explorer for rapid polymorph identification

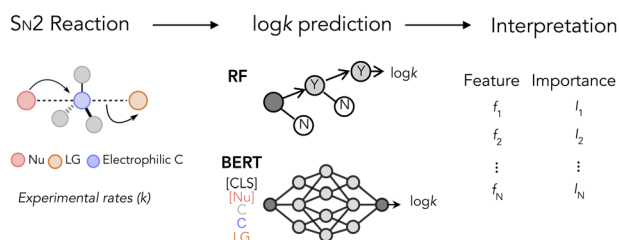
Edward C. Lee, Daniel Salley, Abhishek Sharma and Leroy Cronin\*



743

### Kinetic predictions for S<sub>N</sub>2 reactions using the BERT architecture: comparison and interpretation

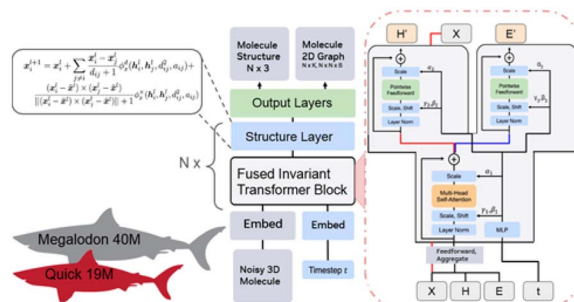
Chloe Wilson,\* María Calvo, Stamatia Zavitsanou, James D. Somper, Ewa Wieczorek, Tom Watts, Jason Crain and Fernanda Duarte\*



754

### Applications of modular co-design for de novo 3D molecule generation

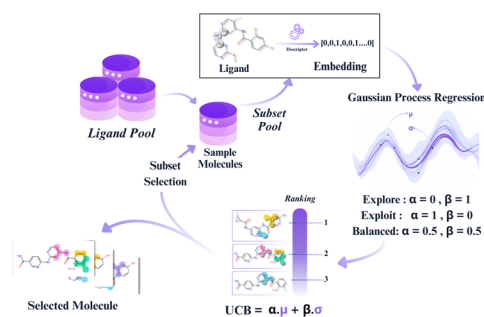
Danny Reidenbach,\* Filipp Nikitin,\* Olexandr Isayev and Saeed Gopal Paliwal



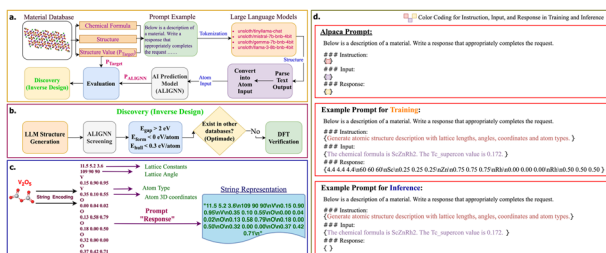
769

### Explainable active learning framework for ligand binding affinity prediction

Satya Pratik Srivastava, Rohan Gorantla, Sharath Krishna Chundru, Claire J. R. Winkelman, Antonia S. J. S. Mey\* and Rajeev Kumar Singh\*



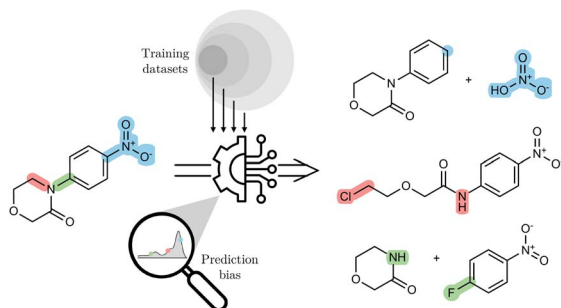
780



## Evaluating large language models for inverse semiconductor design

Muhammed Nur Talha Kilic, Daniel Wines,\*  
Kamal Choudhary,\* Vishu Gupta, Youjia Li,  
Sayak Chakrabarty, Wei-Keng Liao, Alok Choudhary  
and Ankit Agrawal\*

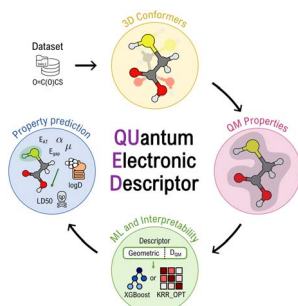
793



## An exploration of dataset bias in single-step retrosynthesis prediction

Sara Tanovic, Ewa Wiczorek and Fernanda Duarte\*

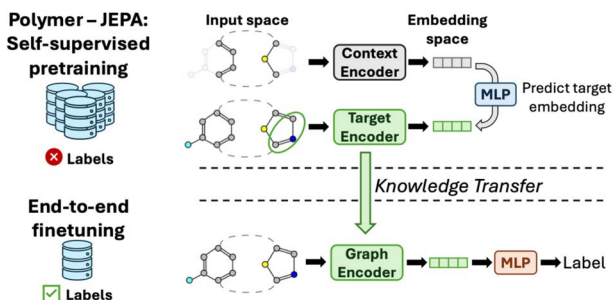
803



## Assessing the performance of quantum-mechanical descriptors in physicochemical and biological property prediction

Alejandra Hinostroza Caldas, Artem Kokorin,  
Alexandre Tkatchenko\* and Leonardo Medrano  
Sandona\*

819



## Joint embedding predictive architecture for self-supervised pretraining on polymer molecular graphs

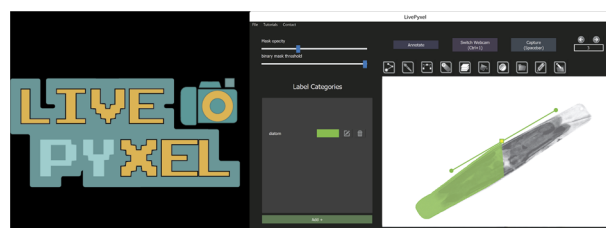
Francesco Piccoli, Gabriel Vogel and Jana M. Weber\*



835

### LivePyxel: accelerating image annotations with a Python-integrated webcam live streaming

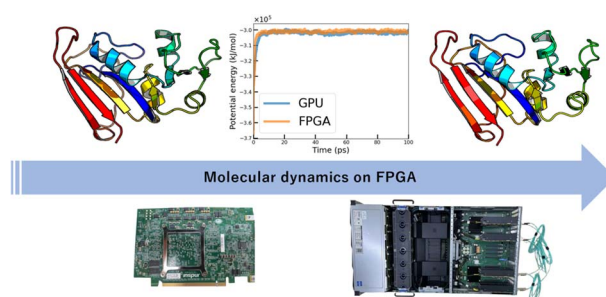
Uriel Garcilazo-Cruz,<sup>\*</sup> Joseph O. Okeme and Rodrigo A. Vargas-Hernández<sup>\*</sup>



844

### Molecular dynamics simulations accelerated on FPGA with high-bandwidth memory

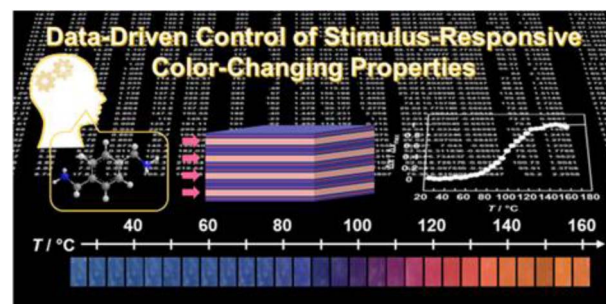
Jing Xiao, Jinfeng Chen, Ye Ding, You Xu and Jing Huang<sup>\*</sup>



862

### A data-driven approach to control stimulus responsivity of functional polymer materials: predicting thermoresponsive color-changing properties of polydiacetylene

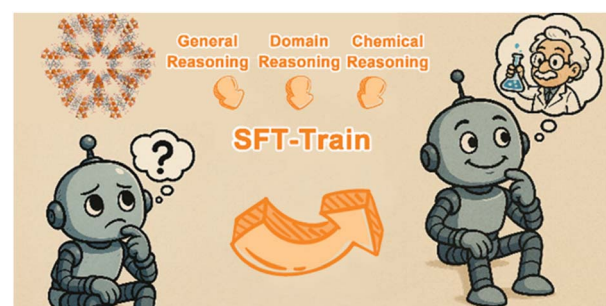
Risako Shibata, Nano Shioda, Hiroaki Imai, Yasuhiko Igarashi and Yuya Oaki<sup>\*</sup>



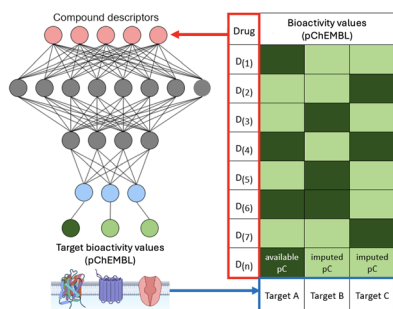
869

### MOFReasoner: think like a scientist—a reasoning large language model via knowledge distillation

Xuefeng Bai, Zhiling Zheng, Xin Zhang,<sup>\*</sup> Hao-Tian Wang, Rui Yang and Jian-Rong Li<sup>\*</sup>



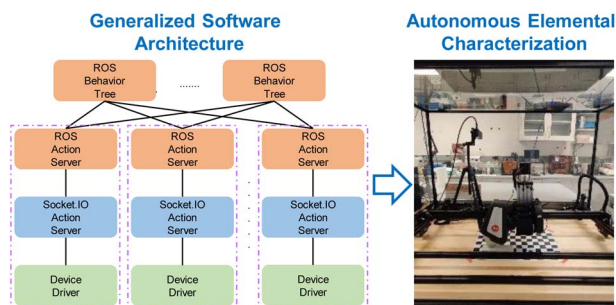
878



## A multi-task learning approach for prediction of missing bioactivity values of compounds for the SLC transporter superfamily

Tarik Ćerimagić, Sergey Sosnin and Gerhard F. Ecker\*

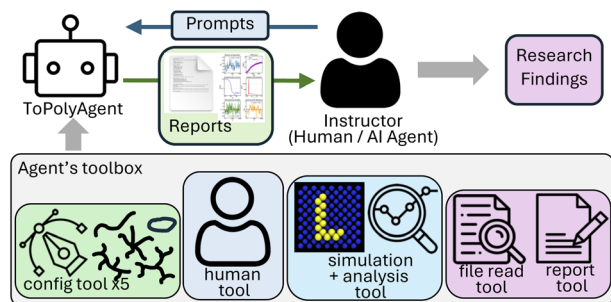
891



## Autonomous elemental characterization enabled by a low cost robotic platform built upon a generalized software architecture

Xuan Cao,\* Yuxin Wu and Michael L. Whittaker\*

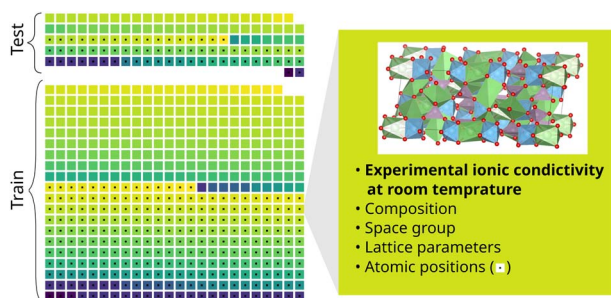
901



## ToPolyAgent: AI agents for coarse-grained bead-spring topological polymer simulations

Lijie Ding,\* Jan-Michael Carrillo and Changwoo Do\*

910



## OBELiX: a curated dataset of crystal structures and experimentally measured ionic conductivities for lithium solid-state electrolytes

Félix Therrien,\* Jamal Abou Haibeh, Divya Sharma, Rhiannon Hendley, Leah Wairimu Mungai, Sun Sun, Alain Tchagang, Jiang Su, Samuel Huberman, Yoshua Bengio, Hongyu Guo,\* Alex Hernández-García\* and Homin Shin\*

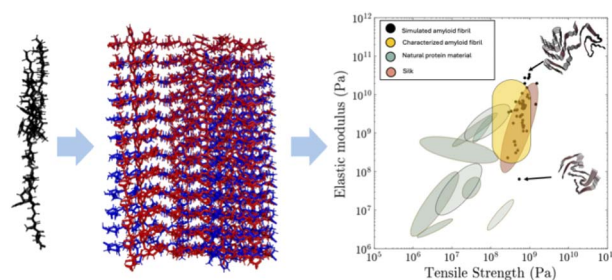


## PAPERS

919

**FiberForge: enabling high-throughput simulations of the mechanical properties of helical fibrils**

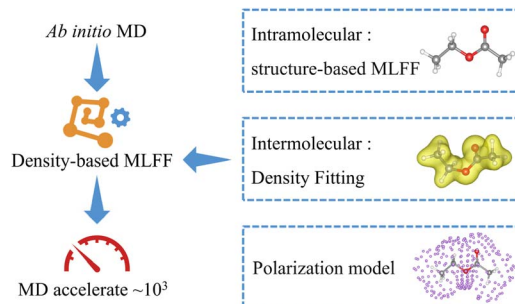
Kieran Nehil-Puleo and Zhongyue John Yang\*



931

**DBMLFF: linear scaling machine learning force fields via electron density decomposition for molecular electrolytes**

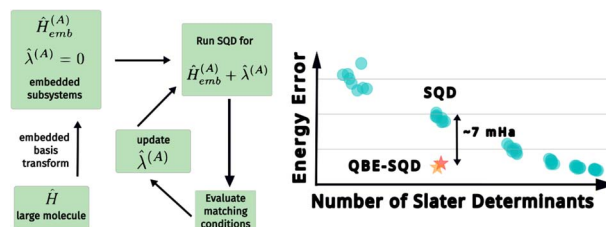
Jie Shen, Chenyu Wang, Libin Chen, Shaoqin Jiang,\* Jianhui Chen, Cuilian Wen, Bo Wu, Baisheng Sa\* and Lin-Wang Wang\*



945

**Towards utility-scale electronic structure with sample-based quantum bootstrap embedding**

Joel Bierman and Yuan Liu\*



## CORRECTION

957

**Correction: A case study on hybrid machine learning and quantum-informed modelling for solubility prediction of drug compounds in organic solvents**

Weiling Wang, Isabel Cooley, Morgan R. Alexander, Ricky D. Wildman, Anna K. Croft and Blair F. Johnston\*

