

# Data-driven discovery in the chemical sciences

10-12 September 2024 | Oxford, UK



Faraday  
Discussions

## Tuesday 10 September

11:00	Registration and refreshments
12:00	Lunch
12:45	<b>Welcome and introductions</b> Volker Deringer & Fernanda Duarte, <i>Co-chairs of Scientific Committee</i>
12:55	<b>Outline of Discussion format</b> Emma Gorrell & Lauren Yarrow-Wright, <i>Royal Society of Chemistry</i>
13:30	<b>Introductory lecture – Spiers memorial lecture</b> (Session chair: Volker Deringer) Alán Aspuru-Guzik <i>University of Toronto, Canada</i>
14:30	Short break (no refreshments)
	<b>Session 1: Discovering chemical structure</b> (Session chair: Graeme Day)
14:45	<b>Beyond theory-driven discovery: introducing hot random search and datum-derived structures</b> Chris Pickard <i>University of Cambridge, UK</i>
14:50	<b>Integration of generative machine learning with the heuristic crystal structure prediction code FUSE</b> Chris Collins <i>University of Liverpool, UK</i>
14:55	<b>Large property models: a new generative machine-learning formulation for molecules</b> Brett Savoie <i>Purdue University, USA</i>
15:00	Discussion
16:15	Refreshments
	<b>Session 1 continued: Discovering chemical structure</b> (Session chair: Janine George)
16:45	<b>Data-efficient fine-tuning of foundational models for first-principles quality sublimation enthalpies</b> Venkat Kapil <i>University of Cambridge, UK</i>
16:50	<b>Knowledge distillation of neural network potential for molecular crystals</b> Takuya Taniguchi <i>Waseda University, Japan</i>
16:55	<b>Modelling ligand exchange in metal complexes with machine learning potentials</b> Veronika Jurásková <i>University of Oxford, UK</i>
17:00	Discussion
18:15	Poster session and wine reception
19:45	Close

Wednesday 11 September

	<b>Session 2: Discovering structure–property correlations</b> (Session chair: Nadine Schneider)
09:00	<b>Web-BO: towards increased accessibility of Bayesian optimisation (BO) for chemistry</b> Kim Jelfs <i>Imperial College London, UK</i>
09:05	<b>Sequence determinants of protein phase separation and recognition by protein phase-separated condensates through molecular dynamics and active learning</b> Arya Changiarath Sivadasan <i>Johannes Gutenberg University of Mainz, Germany</i>
09:10	<b>Discovery of highly anisotropic dielectric crystals with equivariant graph neural networks</b> Yuchen Lou, Alex Ganose <i>Imperial College London, UK</i>
09:15	Discussion
10:30	Refreshments
	<b>Session 2 continued: Discovering structure–property correlations</b> (Session chair: Fernanda Duarte)
11:00	<b>Leveraging natural language processing to curate the tmCAT, tmPHOTO, tmBIO, and tmSCO datasets of functional transition metal complexes</b> Heather Kulik <i>Massachusetts Institute of Technology, USA</i>
11:05	<b>Are we fitting data or noise? Analysing the predictive power of commonly used datasets in drug-, materials-, and molecular-discovery</b> Daniel Crusius <i>University of Oxford, UK</i>
11:10	<b>Prediction rigidities for data-driven chemistry</b> Sanggyu Chong <i>EPFL, Switzerland</i>
11:15	<b>Accurate and reliable thermochemistry by data analysis of complex thermochemical networks using active thermochemical tables: the case of glycine thermochemistry</b> Branko Ruscic <i>Argonne National Laboratory, USA</i>
11:20	Discussion
13:00	Lunch
	<b>Session 3: Discovering trends in big data</b> (Session chair: Philippe Schwaller)
13:45	<b>Specialising and analysing instruction-tuned and byte-level language models for organic reaction prediction</b> Jiayun Pang <i>University of Greenwich, UK</i>
13:50	<b>Predictive crystallography at scale: mapping, validating, and learning from 1000 crystal energy landscapes</b> Christopher Taylor <i>University of Southampton, UK</i>
13:55	<b>Optical materials discovery and design with federated databases and machine learning</b> Matthew Evans <i>UCLouvain, Belgium</i>
14:00	Discussion
15:15	Refreshments
	<b>Session 3 continued: Discovering trends in big data</b> (Session chair: Janine George)
15:45	<b>How big is big data?</b> Claudia Draxl

	<i>Humboldt-Universität zu Berlin, Germany</i>
15:50	<b>Making the InChI FAIR and sustainable while moving to inorganics</b> Gerd Blanke <i>RWTH Aachen University, Germany</i>
15:55	Discussion
16:45	<b>Panel Discussion - EDI in Digital Chemistry</b> Panel members: Alán Aspuru-Guzik, <i>University of Toronto, Canada</i> Stephen Hendry, <i>Royal Society of Chemistry</i> Kim Jelfs, <i>Imperial College London, UK</i> Chaired by Volker Deringer & Fernanda Duarte
17:30	Poster session and wine reception
19:00	Close
19:30	Conference dinner

#### Thursday 12 September

	<b>Session 4: Discovering synthesis targets</b> (Session chair: Philippe Schwaller)
09:00	<b>Analysis of uncertainty of neural fingerprint-based models</b> Miriam Mathea <i>BASF, Germany</i>
09:05	<b>Re-evaluating retrosynthesis algorithms with Syntheseus</b> Marwin Segler <i>Microsoft Research AI for Science, UK</i>
09:10	<b>Embedding human knowledge in material screening pipeline as filters to identify novel synthesizable inorganic materials</b> Basita Das <i>Massachusetts Institute of Technology, USA</i>
09:15	Discussion
10:30	Refreshments
	<b>Session 4 continued: Discovering synthesis targets</b> (Session chair: Volker Deringer)
11:00	<b>Mapping inorganic crystal chemical space</b> Aron Walsh <i>Imperial College London, UK</i>
11:05	<b>A critical reflection on attempts to machine-learn materials synthesis insights from text-mined literature recipes</b> Wenhao Sun <i>University of Michigan, Ann Arbor, USA</i>
11:10	Discussion
12:00	<b>Concluding remarks lecture</b> (Session chair: Fernanda Duarte) Andy Cooper <i>University of Liverpool, UK</i>
12:30	<b>Acknowledgements</b> Volker Deringer & Fernanda Duarte, <i>Co-chairs of Scientific Committee</i>
12:45	Close of meeting and lunch