Electronic Supplementary Information

Artificial Intelligence in the Discovery and Design of Molecular Semiconductors: A Systematic Review

Malin Zollner,[a] Yashar Moshfeghi,[b] and Tahereh Nematiaram[*a]

[a] Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, U.K. E-mail: tahereh.nematiaram@strath.ac.uk

[b] Department of Computer and Information Sciences, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, U.K.

The Supporting Information includes the PRISMA 2020 flow diagram (Figure S1) and PRISMA 2020 checklist (Figure S2), the exact Boolean search strings and filtering criteria applied to all queried databases (Section 2) and tables of representative predictive (Table S1) and generative (Table S2) studies.

1 PRISMA checklist and flow diagram

Figure S1 presents the PRISMA 2020 flow diagram for this systematic review, summarising the study identification, screening, and inclusion process across all databases. The diagram reports the number of records identified, duplicates removed, records excluded at title/abstract and full-text screening stages, and the final set of studies included in the qualitative synthesis, in accordance with PRISMA guidelines. Studies excluded at the full-text screening stage comprise works focused on polymeric systems,[1–24] review articles,[25–33] or studies not addressing organic semiconductor applications.[34] Figure S2 provides the completed PRISMA 2020 checklist, indicating where each reporting item is addressed in the main text or Supporting Information.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only
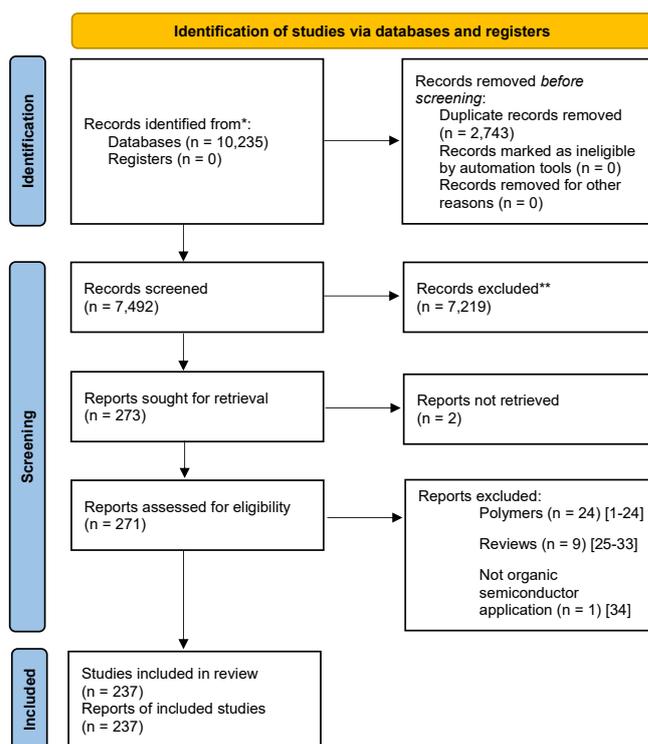
*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

Figure S1: PRISMA 2020 flow diagram illustrating the study identification, screening, and inclusion process for the systematic review across all searched databases. Figure adapted from Page *et al.*[35].

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	-
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	1) Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	1) Introduction
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	2.1) Search Strategy and Inclusion Criteria
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	2.1) Search Strategy and Inclusion Criteria
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Supporting information – Section 1)
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	2.2) Study selection and Data Extraction
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	2.2) Study selection and Data Extraction
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	2.3) Quality assessment
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	-
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	2.3) Quality assessment
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	2.2) Study selection and Data Extraction
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	2.3) Quality assessment
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	-
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	2.2) Study selection and Data Extraction
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	2.2) Study selection and Data Extraction
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	-
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	2.2) Study selection and Data Extraction
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	-
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	2.3) Quality assessment

PRISMA 2020 Checklist

RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Supporting information – Prisma_flow_diagram.pdf
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Supporting information – Prisma_flow_diagram.pdf
Study characteristics	17	Cite each included study and present its characteristics.	Supporting information – Database.csv
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Supporting information – Database.csv
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Supporting information – Database.csv
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Supporting information – Database.csv
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Supporting information – Database.csv
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	3) Landscape of AI Application in Molecular Semiconductor Research
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Supporting information – Database.csv
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	-
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Supporting information – Database.csv
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	3) Landscape of AI Application in Molecular Semiconductor Research
	23b	Discuss any limitations of the evidence included in the review.	Supporting information – Database.csv
	23c	Discuss any limitations of the review processes used.	2.3) Quality assessment
	23d	Discuss implications of the results for practice, policy, and future research.	3.5) Challenges and Opportunities
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	-
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	-
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	-
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	-
Competing interests	26	Declare any competing interests of review authors.	Conflicts of Interest
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data availability

Figure S2: Completed PRISMA 2020 checklist indicating the location of each reporting item within the main text and Supporting Information. Figure adapted from Page *et al.*[35].

2 Boolean searches

2.1 Scopus

Boolean: TITLE-ABS-KEY(("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "reinforcement learning" OR "data-driven") AND ("organic semiconductor*" OR "organic electronic*" OR "organic semiconducting material*" OR "small-molecule" OR "molecular semiconductor*" OR "OLED" OR "Organic Light-Emitting Diodes" OR "organic photovoltaic*")) AND PUBYEAR ¿ 2009 AND PUBYEAR ¡ 2025 AND (LIMIT-TO(DOCTYPE,"ar") OR LIMIT-TO(DOCTYPE,"re")) AND LIMIT-TO(LANGUAGE,"English")

2.2 Web of science

Boolean: (TS=(("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "reinforcement learning" OR "data-driven") AND ("organic semiconductor*" OR "organic electronic*" OR "organic semiconducting material*" OR "small-molecule" OR "molecular semiconductor*" OR "OLED" OR "Organic Light-Emitting Diodes" OR "organic photovoltaic*"))) AND LA=(English)
  Filters: custom date range 2010-2025

2.3 ACS

Boolean: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "reinforcement learning" OR "data-driven") AND ("organic semiconductor*" OR "organic electronics*" OR "organic semiconducting material*" OR "small-molecule" OR "molecular semiconductor*" OR "OLED" OR "Organic Light-Emitting Diodes" OR "organic photovoltaic*") NOT "Drug")
  Filters: custom date range 2010-2025

2.4 RSC

Boolean: each combination individually and each with including NOT "drug" e.g. "artificial intelligence" AND "organic semiconductor" NOT "drug", "artificial intelligence" AND "organic electronic" NOT "drug"
  Filters: custom date range 2010-2025

2.5 Springer

Boolean: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "neural networks" OR "reinforcement learning" OR "data-driven") AND ("organic semiconductors" OR "organic semiconductor" OR "organic electronics" OR "organic electronic" OR "organic semiconducting materials" OR "organic semiconducting material" OR "small-molecule" OR "molecular semiconductor" OR "molecular semiconductors" OR "OLED" OR "Organic Light-Emitting Diodes" OR "organic photovoltaic" OR "organic photovoltaics")
  Filters: custom date range 2010-2025, English language, sorted by relevance
  Note: Only the first 1000 papers are shown on Springer

2.6 PubMed

Boolean: each combination individually e.g. "artificial intelligence" AND "organic semiconductor", "artificial intelligence" AND "organic electronic"

Filters: custom date range 2010-2025, English language

2.7 Wiley Online Library

Boolean: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "reinforcement learning" OR "data-driven") AND ("organic semiconductor*" OR "organic electronic*" OR "organic semiconducting material*" OR "small-molecule" OR "molecular semiconductor*" OR "OLED" OR "Organic Light-Emitting Diodes" OR "organic photovoltaic*")

Filters: custom date range 2010-2025, journal filter

Note: title, abstract, keywords were searched separately

2.8 Science Direct

Boolean: TITLE-ABS-KEY() with shorter combinations due to Boolean limit e.g. TITLE-ABS-KEY(("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "neural networks" OR "reinforcement learning" OR "data-driven") AND ("Organic Light-Emitting Diodes")), TITLE-ABS-KEY (("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "neural networks" OR "reinforcement learning" OR "data-driven") AND ("organic electronics" OR "organic electronic"))

Filters: custom date range 2010-2025, limited to review articles, limited to research article, limited to mini reviews

2.9 Arxiv

Boolean: each combination individually e.g. "artificial intelligence" AND "organic semiconductor", "artificial intelligence" AND "organic electronic"

Filters: custom date range February 2025-April 2025, materials science category, physics category, computer science category

2.10 Nature

Boolean: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "reinforcement learning" OR "data-driven") AND ("organic semiconductor*" OR "organic electronic*" OR "organic semiconducting material*" OR "small-molecule" OR "molecular semiconductor*" OR "OLED" OR "Organic Light-Emitting Diodes" OR "organic photovoltaic*")

Filters: custom date range 2010-2025

3 Representative studies

The tables presented in this section are intended to highlight a subset of representative studies selected from the full set of analysed studies in this review. Tables S1 and S2 are designed to illustrate key methodological approaches, application domains, and validation strategies across predictive and generative categories, rather than to

provide exhaustive coverage. To ensure transparency and reproducibility, a complete comma-separated values file containing all 237 included studies is provided as additional Supporting Information. This dataset reports, for each study, the DOI, title, year of publication, predictive or generative classification, application domain (photoactive materials, electronic structure and spectroscopic properties, emissive materials, or charge transport), target property, data provenance (experimental, computational, or hybrid), validation strategy (internal or external), presence of experimental validation, and the availability of open databases and open-source code.

Table S1: Representative studies employing predictive modelling approaches for molecular semiconductors. Highlighted examples illustrate the diversity of machine learning methods used for property prediction and materials screening. Abbreviations: GNN = Graph Neural Network; GCN = Graph Convolutional Network; GCN-FNN = Graph Convolutional Network - Feedforward Neural Network; SVM = Support Vector Machine.

Study (Year)	Application	ML Method & Data	Key Results
Jeong *et al.* (2022)[36]	Electronic structure and spectroscopic properties	Transfer learning; pretrained on spectroscopic data, fine-tuned on 3,000 experimental HOMO/LUMO values	First to apply transfer learning to HOMO/LUMO prediction in OLED molecules; achieved MAEs of 0.050–0.065 eV and guided high-EQE emitter discovery.
Xie *et al.* (2024)[37]	Electronic structure and spectroscopic properties	Fine-tuned GPT-3 (compared to GNN and SVM); 48,0000 tokenised SMILES (compared to RDKit descriptors and GNN)	Usage of large language models for prediction tasks; GPT-3 achieved accuracies of 0.92 (HOMO) and 0.94 (LUMO).
Pyzer-Knapp *et al.* (2015) [38]	Photoactive materials	Neural network; predicted PCE, HOMO and LUMO for 250,000 molecules from the Harvard Clean Energy Project	One of the first to predict PCE using ML; achieved MAEs of 0.28% (PCE), 0.028 eV (HOMO) and 0.032 eV (LUMO).
Bornschlegl *et al.* (2025)[39]	Photoactive materials	Predicted UV-C photostability of 138 molecules using Gaussian Process Regression	First to investigate UV-C photostability; identified stabilising and destabilising motifs achieving $R^2 = 0.631$.

Study (Year)	Application	ML Method & Data	Key Results
Lee *et al.* (2024)[40]	Emissive materials	Crystal-structure-informed GCN trained on 1117 multilayer OLED stacks	First to model full OLED device stacks using GNN; achieved MAE = 2.6% and R^2 = 0.83 for current efficiency, outperforming MLP baselines.
Barneschi *et al.* (2024)[41]	Emissive materials	3D graph neural network trained on 85,000 DFT and X-ray structures	First ML model for predicting inverted singlet–triplet gaps; achieved MAE = 0.02 eV; validated performance on experimental geometries.
Niu *et al.* (2023)[42]	Charge transport	GCN-FNN trained on 3D conformers of fused-ring systems; stratified sampling by ring number	Achieved R^2 $\approx$ 0.99, RMSE = 2.6 meV; first to show stratified GCNs significantly outperform descriptor-based models for reorganisation energy.
Bhat *et al.* (2024)[43]	Charge transport	3D message-passing neural network trained on 438,000 dimers from 25,000 crystals	Enabled rapid prediction of electronic couplings (MAE $\approx$ 3 meV) and mobility screening across 60,000 crystals in minutes.

Table S2: Representative studies on **generative design and discovery** of molecular semiconductors. Highlighted examples demonstrate the range of AI-driven approaches used for inverse molecular design and optimisation. Abbreviations: DFTB = Density-Functional Tight-Binding, DNN = Deep Neural Network; GA = Genetic Algorithm; RL = Reinforcement Learning; RNN = Recurrent Neural Network; TD-DFT = Time-Dependent Density Functional Theory.

Study (Year)	Application	ML Method & Data	Key Results
Li & Tabor (2023) [44]	Electronic structure and spectroscopic properties	RL (REINVENT) combined with QChem	Discovered diverse singlet-fission molecules (e.g., multi-anthracenes) with ideal singlet–triplet energy gaps and balanced synthetic accessibility.

Study (Year)	Application	ML Method & Data	Key Results
Weiss *et al.* (2023) [45]	Electronic structure and spectroscopic properties	Guided Diffusion Model incorporating property-predictor gradients	Integration of property-predictor gradients yielded diverse polybenzenoid hydrocarbons with optimised electron affinity, ionisation potential, and bandgap beyond the training distribution.
Khazaal *et al.* (2020)[46]	Photoactive materials	EA combined with DFTB	Identified 20 novel branched oligothiophenes optimised for light-harvesting efficiency and synthetic accessibility.
Greenstein *et al.* (2022)[47]	Photoactive materials	GA guided by TD-DFT fitness functions	Generated 5,426 NFAs; 159 candidates predicted to exceed 20% PCE, highlighting recurrent structural motifs in high-performance NFAs.
Kim *et al.* (2018)[48]	Emissive materials	Deep Generative Network combining a DNN encoder and a RNN decoder	Early example of generative design for OLED hosts; successfully proposed new molecules with $T_1 > 3\,\text{eV}$.
Kwak *et al.* (2022)[44]	Emissive materials	REINVENT with multi-objective property optimisation	Simultaneous optimisation of singlet–triplet gap, oscillator strength, and molecular stability, illustrating synthesis-aware generative optimisation.
Kunkel *et al.* (2021)[49]	Charge transport	AL using Bayesian Optimisation	Active learning rapidly identified 20 novel structures with mobility descriptors exceeding known benchmarks.
Kwak *et al.* (2022)[50]	Charge transport	RNN combined with Deep RL	Designed hole-transport materials with low λ and high glass-transition temperatures, balancing mobility with morphological stability.

Bibliography

[1] J. Sun, D. Li, J. Zou, S. Zhu, C. Xu, Y. Zou, Z. Zhang and H. Lu, *npj Computational Materials*, 2024, **10**, 181.

[2] W. K. Tatum, D. Torrejon, A. B. Resing, J. W. Onorato and C. K. Luscombe, *Computational Materials Science*, 2021, **197**, 110599.

[3] M. Chi, R. Gargouri, T. Schrader, K. Damak, R. Maâlej and M. Sierka, *Polymers*, 2021, **14**, 26.

[4] N. E. Jackson, *The Journal of Physical Chemistry B*, 2020, **125**, 485–496.

[5] N. Ahmad, M. A. Ibrahim, S. R. Sayed, S. S. A. Shah, M. H. Tahir and Y. Zou, *Solar Energy*, 2025, **287**, 113223.

[6] S. Zhang, S. Li, S. Song, Y. Zhao, L. Gao, H. Chen, H. Li and J. Lin, *Advanced Materials*, 2025, **37**, 2407613.

[7] G. J. Moore, O. Bardagot and N. Banerji, *Advanced Theory and Simulations*, 2022, **5**, 2100511.

[8] T. W. David and J. Kettle, *The Journal of Physical Chemistry C*, 2022, **126**, 4774–4784.

[9] M. Keshav and F. Kaka, *Solar Energy Materials and Solar Cells*, 2023, **254**, 112258.

[10] S. Zhang, X. He, P. Xiao, X. Xia, F. Zheng, S. Xiang and Q. Lu, *Advanced Functional Materials*, 2024, **34**, 2409143.

[11] L. Fu, H. Hu, Q. Zhu, L. Zheng, Y. Gu, Y. Wen, H. Ma, H. Yin and J. Ma, *Nano Research*, 2023, **16**, 3588–3596.

[12] Y. Wu, J. Guo, R. Sun and J. Min, *npj Computational Materials*, 2020, **6**, 120.

[13] M. Li, C.-R. Zhang, M.-L. Zhang, J.-J. Gong, X.-M. Liu, Y.-H. Chen, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, *physica status solidi (a)*, 2024, **221**, 2400008.

[14] J.-H. Li, C.-R. Zhang, M.-L. Zhang, X.-M. Liu, J.-J. Gong, Y.-H. Chen, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, *Organic Electronics*, 2024, **125**, 106988.

[15] O. D. Abarbanel and G. R. Hutchison, *The Journal of Chemical Physics*, 2021, **155**, year.

[16] N. G. An, J. Y. Kim and D. Vak, *Energy & Environmental Science*, 2021, **14**, 3438–3446.

[17] G.-H. Kim, K. Yoon, C. Lee, M. Nam and D.-H. Ko, *Bulletin of the Korean Chemical Society*, 2024, **45**, 125–130.

[18] G.-H. Kim, C. Lee, K. Kim and D.-H. Ko, *Nano Energy*, 2023, **106**, 108108.

[19] H. Mosalam, S. Hussien and S. O. Abdellatif, *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 2024, **37**, e3132.

[20] N. Aleb, *International Journal of Advances in Soft Computing & Its Applications*, 2023, **15**, year.

[21] M. Marzouglal, A. Souahlia, L. Bessissa, D. Mahi, A. Rabehi, Y. Z. Alharthi, A. K. Bojer, A. Flah, M. M. Alharthi and S. S. Ghoneim, *Scientific Reports*, 2024, **14**, 25931.

[22] A. Paul, A. Furmanchuk, W.-k. Liao, A. Choudhary and A. Agrawal, *Molecular informatics*, 2019, **38**, 1900038.

[23] X. Du, L. Lüer, T. Heumueller, A. Classen, C. Liu, C. Berger, J. Wagner, V. M. Le Corre, J. Cao, Z. Xiao *et al.*, *InfoMat*, 2024, **6**, e12554.

[24] W. Sun, M. Li, Y. Li, Z. Wu, Y. Sun, S. Lu, Z. Xiao, B. Zhao and K. Sun, *Advanced Theory and Simulations*, 2019, **2**, 1800116.

[25] J. C. Fromer and C. W. Coley, *Patterns*, 2023, **4**, year.

[26] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao *et al.*, *Accounts of Chemical Research*, 2021, **54**, 849–860.

[27] F. Häse, L. M. Roch, P. Friederich and A. Aspuru-Guzik, *Nature Communications*, 2020, **11**, 4587.

[28] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer *et al.*, *Communications Materials*, 2022, **3**, 93.

[29] F. Mayr, M. Harth, I. Kouroudis, M. Rinderle and A. Gagliardi, *The Journal of Physical Chemistry Letters*, 2022, **13**, 1940–1951.

[30] Y. Zhuo and J. Brgoch, *The Journal of Physical Chemistry Letters*, 2021, **12**, 764–772.

[31] G. Gryn'ova, K.-H. Lin and C. Corminboeuf, *Journal of the American Chemical Society*, 2018, **140**, 16370–16386.

[32] R. Yu, S. Li, H. Yuan, Z. Yang, S. Jin and Z. Tan, *The Journal of Physical Chemistry Letters*, 2024, **15**, 2781–2803.

[33] R. Yu, S. Li, H. Yuan, Z. Yang, S. Jin and Z. Tan, *The Journal of Physical Chemistry Letters*, 2024, **15**, 2781–2803.

[34] S. Ranu, B. T. Calhoun, A. K. Singh and S. J. Swamidass, *Molecular Informatics*, 2011, **30**, 809–815.

[35] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, *bmj*, 2021, **372**, n71.

[36] M. Jeong, J. F. Joung, J. Hwang, M. Han, C. W. Koh, D. H. Choi and S. Park, *npj Computational Materials*, 2022, **8**, 147.

[37] Z. Xie, X. Evangelopoulos, Ö. H. Omar, A. Troisi, A. I. Cooper and L. Chen, *Chemical science*, 2024, **15**, 500–510.

[38] E. O. Pyzer-Knapp, K. Li and A. Aspuru-Guzik, *Advanced Functional Materials*, 2015, **25**, 6495–6502.

[39] A. J. Bornschlegl, P. Duchstein, J. Wu, J. S. Rocha-Ortiz, M. Caicedo-Reina, A. Ortiz, B. Insuasty, D. Zahn, L. Luer and C. J. Brabec, *Journal of the American Chemical Society*, 2025, **147**, 1957–1967.

[40] T. Lee, J. Choi, I. Na, I. Yoo, S. Woo, K. J. Kim, M. Park, J. Yang, J. Min, S. Lee *et al.*, *Advanced Intelligent Systems*, 2024, 2400598.

[41] L. Barneschi, L. Rotondi and D. Padula, *The Journal of Physical Chemistry A*, 2024, **128**, 2417–2426.

[42] X. Niu, Y. Dang, Y. Sun and W. Hu, *Journal of Energy Chemistry*, 2023, **81**, 143–148.

[43] V. Bhat, B. Ganapathysubramanian and C. Risko, *The Journal of Physical Chemistry Letters*, 2024, **15**, 7206–7213.

[44] C.-H. Li and D. P. Tabor, *Chemical Science*, 2023, **14**, 11045–11055.

[45] T. Weiss, E. Mayo Yanes, S. Chakraborty, L. Cosmo, A. M. Bronstein and R. Gershoni-Poranne, *Nature Computational Science*, 2023, **3**, 873–882.

[46] A. S. Khazaal, M. Sprinborg, C. Fan and K. Huwig, *Computational Condensed Matter*, 2020, **25**, e00503.

[47] B. L. Greenstein, D. C. Hiener and G. R. Hutchison, *The Journal of Chemical Physics*, 2022, **156**, 174107.

[48] K. Kim, S. Kang, J. Yoo, Y. Kwon, Y. Nam, D. Lee, I. Kim, Y.-S. Choi, Y. Jung, S. Kim *et al.*, *npj Computational Materials*, 2018, **4**, 67.

[49] C. Kunkel, J. T. Margraf, K. Chen, H. Oberhofer and K. Reuter, *Nature Communications*, 2021, **12**, 2422.

[50] H. S. Kwak, Y. An, D. J. Giesen, T. F. Hughes, C. T. Brown, K. Leswing, H. Abroshan and M. D. Halls, *Frontiers in Chemistry*, 2022, **9**, 800370.