# WHAT IF MATERIALS PREDICTED ON A COMPUTER COULD BE ROUTINELY REALISED IN A LABORATORY?

Kim E. Jelfs

*Department of Chemistry, Molecular Sciences Research Hub, White City Campus, Imperial College London, 28 Wood Lane, London, W12 OBZ*

## Materials discovery

Chemistry is often a search for rare events, such as chemical bond breaking or formation, preferential binding, or phase transitions. But beyond this, chemical discovery itself centres on rare events - usually a large number of chemical systems or materials are synthesised and tested for desired properties before finding a novel system, or class of systems, for a target application. Often the discovery of a new material class and the discovery of the material's utility are two separate events, sometimes, perhaps even typically, separated by decades.[1] Cheetham *et al.* summarise that while some materials breakthroughs result from design principles based on scientists' knowledge, and some are serendipitous, for example buckyball and Teflon, the majority of discoveries result from repurposing of materials previously either made out of curiosity or with an alternative application in mind.[1] Examples of the latter include conducting polymers, topological insulators, and electrodes for lithium-ion batteries.

Chemical synthesis is generally time-consuming, resource intensive, requires field-specific expertise, and prone to plenty of failures. Indeed, even in reproducing previously reported syntheses, failure is worryingly common.[2] Set against this context, we must recognise that material discovery is frustratingly slow. This at a time where we face enormous challenges such as climate change and resource scarcity, and the discovery of new materials, for example for renewable energy generation and storage, molecular separations and catalysis, is critical. Therefore, we must accelerate discovery. High-throughput synthesis and chemical automation as well as artificial intelligence (AI) are highly topical, but is it realistic to expect that they can accelerate discovery, or will they face the same hurdles as earlier innovations? In an ideal world, we would be able to *a priori* screen, or

predict, on a computer which materials have the properties required for a target application. Computational chemistry has made enormous contributions over the last decades, especially in unravelling structure-property relationships through atomistic insight, helping to facilitate both understanding and material design. However, this research is typically post-rationalisation, not prediction, although we are beginning to see some notable examples of the latter, as recently reviewed by Barbatti.[3]

## Hurdles and hopes for the realisation of computational predictions

A central barrier to the realisation of computational predictions is, in my opinion, the challenge of including consideration of whether a material predicted to have promising properties on a computer is thermodynamically and kinetically stable and can be synthesised (see Figure 1). Therefore, materials predicted to be promising on a computer, more often than not, remain "hypothetical" and stuck *in silico*. Indeed, this was a problem termed by Jansen and Schön as "putting the cart before the horse", where materials are "designed" or screened on a computer, but without ever giving consideration of whether the material is "capable of existence".[4] This problem will remain in the era of AI unless discovery approaches augmented by AI take into consideration this "synthesisability" consideration. One aside however, is that the simplest way to factor in consideration would be to limit oneself to very small regions of chemical space - for example just combining a few compounds already physically in hand. This would limit oneself to very small regions of chemical space relative to the near limitless possibilities and greatly reduce the possibility of truly novel material compositions and properties. Thus, instead, we must focus on the development of tools that can assist in synthesisability consideration for the full possibility of all chemical space.

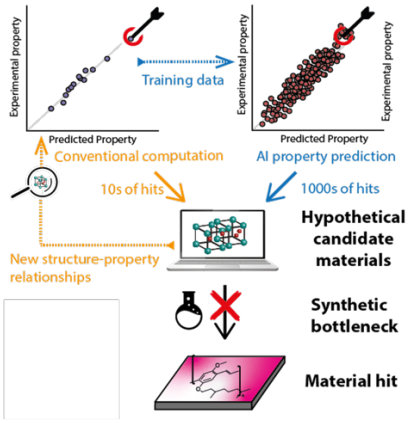


Fig. 1. The synthesis bottleneck preventing experimental realisation of computational predictions.

The synthesis bottleneck for computational predictions is present either when using computational chemistry simulations or AI for materials discovery. The specific challenges to synthesisability consideration vary by system, although can be divided broadly into considering component availability, the material synthesis step, and the material formulation or processing step (Figure 2). Focusing on the area of my group's research, organic materials, then for component availability, both the ease and cost with which the precursors for the material can be purchased or synthesised needs to be considered. There is no simple one-size-fits-all answer as to how cheap the components need to be, or how many synthesis steps are acceptable, or whether only certain reaction classes are viable, sufficiently high yielding or safe. Indeed, I find that for different experimental groups focused on different types of materials, what they consider to be "easily synthesisable" varies considerably. For the material synthesis step, there are challenges such as finding the (potentially) small region of material formation phase space where the material will successfully form, and the need for the material to be stable. Finally, the requirements for the material to be processable into the end device form vary widely by application, such as membrane processing, but may require the material to be soluble (rarely the case for polymers) and will require the material to be stable for a device's operating conditions (potentially high temperature/pressure such as in a battery) over the necessary device lifetime.

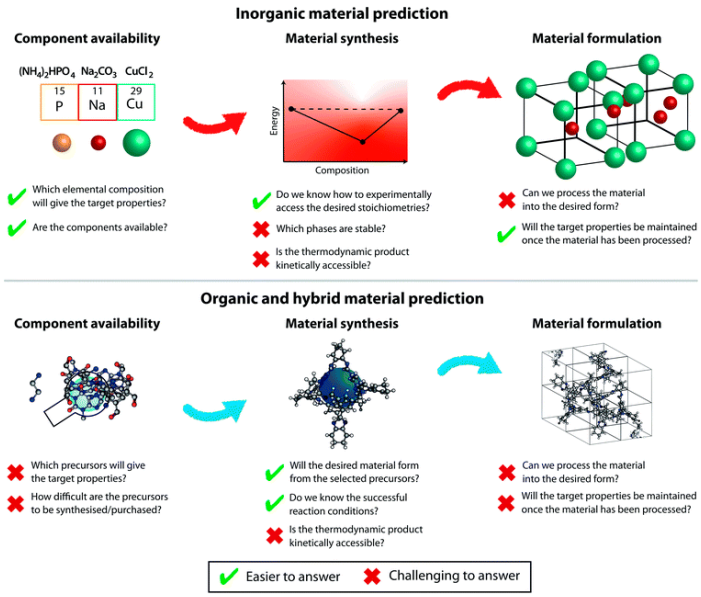


Fig. 2. The hurdles to predicting synthesisable materials on a computer. Reproduced from reference 5.

While there is significant progress in the development of approaches that can assist in consideration of synthesisability, these are not routinely used in computational (organic) materials discovery processes. An exception is the use of algorithms for the ranking of organic molecules based on their structural complexity or predicted number of reaction synthesis steps, although these are not originally developed for application in materials discovery and therefore we have shown such metrics do not necessarily correlate well with the synthesisability expectations from synthetic materials chemists.[6] To develop a algorithmic scoring method that mimics a synthetic material chemist's brain, we trained a supervised machine learning model to score organic molecule's synthesisability.[6] The training data for this model was collected by producing an "app" where our experimental collaborators had to rank molecules when considering the question "*Can you make 1 g of this compound in under 5 steps?"*, a question our experimental collaborators rationalised that they were effectively asking themselves when considering the synthesis of hypothetical computational predictions that we presented them with. The trained model, the *Materials Precursor Score* (*MPS*), was highly effective at simplifying the materials proposed for synthesis that still had promising properties.[6] To further consider subsequent steps in the material synthesis, we are testing automation methods, which have

demonstrated already that only 15% of the material synthesis reactions are successful for the porous organic materials we study, even when the chemical precursors are available and expected to react successfully by system experts.[7] The multiple considerations such as solvent and catalyst choice are very difficult to predict using computational chemistry approaches alone, and thus the benefits of automated approaches are that we are beginning to build larger quantities of consistently collected experimental data to include in data-driven predictions in the future.[7,8]

Other exciting developments for synthesisability prediction include the development of retrosynthesis prediction algorithms,[9,10] although there remain challenges here in how to rank the large network of possible pathways that are produced by such algorithms, and for synthesisability consideration, how to therefore score material synthesisability from retrosynthesis predictions. Data-driven approaches in chemistry more generally can also help with cheminformatics-based predictions of factors such as solubility, especially as increasing amounts of open-source data is available for the more complex systems of consideration for materials, far beyond small molecules. Of course, one does not just want a predicted promising material to be ranked as "synthesisable", one also wants the synthesis recipe to be predicted *a priori*. While there are extensive efforts in the use of machine learning algorithms for literature data extraction to retrieve reported syntheses for known materials, the data is unlikely to be sufficiently complete and accurate to allow direct prediction of new material synthesis routes, especially where there is a greater degree of extrapolation beyond known systems if we want to find more novel new materials.[11,12]

## Future outlook

We need to move to a future where computational predictions are *synthesis-informed*, so that predictions can be more routinely realised, which holds great potential for accelerating the discovery of new materials. While there is significant excitement and early promise for the potential of AI and automation in accelerating chemical discovery, there is lots of work to truly consider synthesisability from the outset, and this will plague the use of AI, such as generative AI (GenAI), in this area just as it has for computational chemistry simulations before. Or else, restrict us to exploration of "safe", but limited regions of chemical space similar to those already explored. Even in scenarios where we

work in an augmented fashion with agentic AI, it is essential to develop synthesisability consideration tools, and these are likely to require the use of both AI and computational chemistry approaches.

**Acknowledgments**

The EPSRC is acknowledged for funding the AI for Chemistry: AIchemy (EP/Y028775/1 and EP/Y028759/1). Dr. Steven Bennett is acknowledged for figure production.

**References**

1. A. K. Cheetham, R. Seshadri, F. Wudl, *Nature Synthesis*, **1**, 514 (2022)
2. R. G. Bergman, R. L. Danheiser, *Angew. Chem. Int. Ed.*, **55**, 12548 (2016).
3. M. Barbatti, *Pure and Applied Chem*. (2025)
4. M. Jansen, J. C. Schön, *Nature Materials*, **3**, 838 (2004)
5. F. T. Szczypiński, S. Bennett and K. E. Jelfs, *Chem. Sci.* **12**, 830 (2021).
6. S. Bennett, F. T. Szczypiński, L. Turcani, M. E. Briggs, R. L. Greenaway, K. E. Jelfs, *J. Chem. Inf. Model.* **61**, 9, 4342 (2021).
7. A. R. Basford, S. Bennett, M. Xiao, L. Turcani, J. Allen, K. E. Jelfs, R. Greenaway*, Chem. Sci.,* **15**, 6331 (2024)
8. A. R. Basford, A. H. Bernardino, P. C. P. Teeuwen, B. D. Egleston, J. Humphreys, K. E. Jelfs, J. R. Nitschke, I. A. Riddell, R. L. Greenaway *Angew. Chem. Int. Ed.,* **64** (24), e202424270.
9. M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* **555**, 604–610 (2018).
10. F. Hastedt, R. M. Bailey, K. Hellgardt, S. N. Yaliraki, E. A. del Rio Chanona, D. Zhang*, Digital Discovery*, **3**, 1194 (2024).
11. Z. Wang, O. Kononova, K. Cruse, T. He, H. Huo, Y. Fei, Y. Zeng, Y. Sun, Z. Cai, W. Sun, G. Ceder, *Sci. Data*, **9**, 231 (2022).
12. W. Sun and N. David, *Faraday Discuss.*, **256**, 614 (2025).

**Checklist**

- Do check all figures for clarity. Indicate those that need to be in color.

[all to be in colour]

- Select words (~10) that can be included in the Index,

[AI, computational prediction; materials discovery; materials synthesis; materials modelling; computational chemistry; machine learning]

- Send figures as separated high resolution files