crystal. Getting electrons to interact with more-complicated laser-pulse sequences than in the current experiment, and with multiple colours of light, might facilitate entirely new forms of electron spectroscopy. Combined with methods for the light-induced temporal structuring of electron beams[9–11], Madan and colleagues' holographic approaches could enable the behaviour of materials to be studied on shorter timescales than that of a single wave cycle of light (the attosecond scale), and with the spatial resolution of an electron microscope.

It remains to be seen whether more-ambitious applications of the new findings will materialize, in which electron beams are used as part of quantum communication systems, or even in quantum computation. Such technologies would probably require the controlled coupling or quantum correlation of multiple free electrons with each other, neither of which has been achieved so far. In the meantime, Madan and colleagues' work represents exciting progress in the manipulation of electrons by light. ■

Claus Ropers *is at the IV. Physical Institute, University of Göttingen, 37077 Göttingen, Germany.*
e-mail: cropers@gwdg.de

1. Smalley, D. E. *et al. Nature* **553**, 486–490 (2018).
2. Madan, I. *et al. Sci. Adv.* **5**, eaav8358 (2019).
3. Lichte, H. & Lehmann, M. *Rep. Prog. Phys.* **71**, 016102 (2008).
4. García de Abajo, F. J., Asenjo-Garcia, A. & Kociak, M. *Nano Lett.* **10**, 1859–1863 (2010).
5. Park, S. T., Lin, M. & Zewail, A. H. *New J. Phys.* **12**, 123028 (2010).
6. Barwick, B., Flannigan, D. J. & Zewail, A. H. *Nature* **462**, 902–906 (2009).
7. Echternkamp, K. E., Feist, A., Schäfer, S. & Ropers, C. *Nature Phys.* **12**, 1000–1004 (2016).
8. Spektor, G. *et al. Science* **355**, 1187–1191 (2017).
9. Priebe, K. E. *et al. Nature Photon.* **11**, 793–797 (2017).
10. Kozák, M., Schönenberger, N. & Hommelhoff, P. *Phys. Rev. Lett.* **120**, 103203 (2018).
11. Morimoto, Y. & Baum, P. *Nature Phys.* **14**, 252–256 (2018).

This article was published online on 3 July 2019.

COMPUTATIONAL CHEMISTRY

Holistic models of reaction selectivity

Computational models that predict the selectivity of reactions are typically accurate for only a specific reaction type and a narrow range of reaction components. A more general model has now been reported. SEE ARTICLE P.343

PER-OLA NORRBY

Selectivity is a linchpin of chemical synthesis — if a synthetic reaction is not selective, it cannot give a good yield of the desired product, and will require tedious purification processes. Chemists have therefore long sought ways of predicting the selectivity of chemical reactions. Computational models can be constructed, but their development is laborious, and they are usually specific to a particular reaction type. On page 343, Reid and Sigman[1] now show that a selectivity model can be built in a semi-automated way and generalized over a range of reactions.

Chemical selectivity comes in many flavours, but it is especially difficult to achieve enantioselectivity, which depends on a property called chirality. Molecules are said to be chiral if they come as two mirror-image forms — enantiomers — that have many identical properties, but can differ in certain important aspects. A good analogy is with hands: a person's right and left hands have the same length, colour and mass, but only one fits into a right-handed glove.

Many biological targets for pharmaceuticals look like right-handed gloves to molecules — only one enantiomer of a molecule will fit into them. For this reason, pharmaceuticals should be synthesized as one enantiomer only; the other form might even be toxic. Asymmetrical catalysts are used to influence synthetic chemical reactions to form only one enantiomer of the product. Nature's asymmetrical catalysts are enzymes, which produce single enantiomers of biomolecules efficiently and with exquisite selectivity. Enzymes can also be used as catalysts for synthetic chemistry, but they generally have a limited range of substrates and can produce only one of the two possible enantiomers of a product.

Modern synthetic catalysts challenge the efficiency of enzymes, and can often be made as mirror-image forms that each produce a different enantiomer of a desired molecule. To support the development of new catalysts, chemists use models to understand and predict the enantioselectivity of catalytic reactions[2,3]. These range in complexity from simple models of the catalyst drawn on paper, onto which a molecular model of the substrate is superimposed to estimate the best fit, to quantum-mechanical calculations that describe an entire reaction path.

"It is highly encouraging to see that holistic reaction models can be produced by using a wide training set."

A direct predecessor of Reid and Sigman's modelling work is a computational approach called quantitative structure–selectivity relationships (QSSR), in which a correlation is sought between the properties of reaction components and the observed selectivity. The relevant properties can be either determined experimentally or calculated, and can include such things as molecular-bond lengths, vibrational frequencies and atomic charges. Using a semi-automated statistical approach (multiple linear regression), these properties are used to construct a model that outputs one numeric value for each reaction system being studied[3]. A result of zero means that there is no selectivity — both enantiomers are produced in equal amounts. A high value indicates a very selective system, and the sign of the numerical output (positive or negative) indicates which enantiomer is mostly produced. Opposite enantiomers of a catalyst produce opposite enantiomers of the product, and this should also be reflected in QSSR models of synthetic catalysts; this requirement is not essential for models of enzymes, however, because only one enantiomeric form of any enzyme exists in nature.

QSSR models are normally limited to a narrow set of substrates and catalysts, because the assumptions built into the machine-learning procedures are invalidated by large deviations from the molecular structures used to train the model. Reid and Sigman have taken on the challenge of making a general QSSR model, starting from an earlier model reported by Reid and colleagues[4].

Inspired by enzyme models, Reid and Sigman ignored the sign conventions usually adhered to in models of synthetic catalysis — that is, they produced a model that predicts the magnitude of enantioselectivity for a group of catalytic reactions (Fig. 1), but only for one enantiomer of the catalyst. Switching the catalyst to its mirror image will therefore not switch the sign of the output in their model, and the model cannot predict which enantiomer is produced as the major isomer. However, the major enantiomer can be predicted from the preceding work[4]. Within this framework, the authors demonstrated that one of the components of the modelled reactions could be varied to an unprecedented degree, without affecting the high accuracy of the predictions.

How can one model achieve such a wide range of accurate predictions? Part of the explanation is probably that all the reactions share a similar mechanism: a planar substrate (an imine molecule; Fig. 1) is 'gripped' from

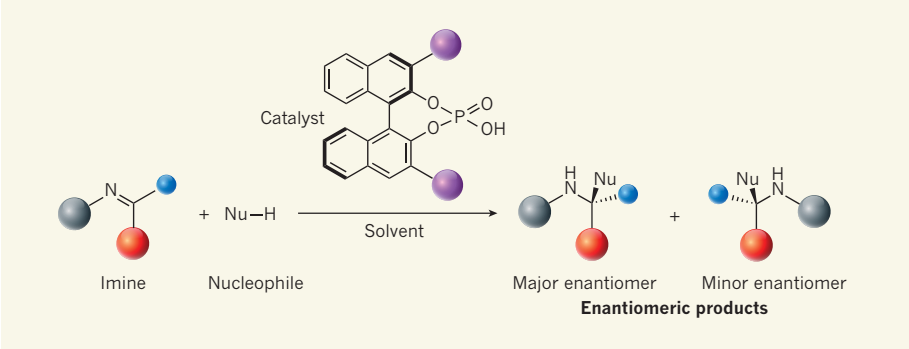

Figure 1 | Model reactions. Reid and Sigman[1] report a computational model that predicts the outcome of reactions when a wide range of nucleophilic molecules react with imines in the presence of a catalyst, accounting for factors such as molecular structure and solvent. More specifically, the model reports the magnitude of the enantioselectivity of the reactions — a measure of the ratio of the two mirror-image isomers (enantiomers) of the product formed in the reaction. Spheres represent a variety of chemical groups; bonds shown in bold or as solid wedges project above the plane of the page; broken wedges project below the plane of the page. Nu represents a range of groups or molecular structures.

one side by the chiral catalyst[4], so that any reaction has to occur on the other side. The third reaction component (a nucleophile), can therefore be varied substantially in the model. But the main reason is that the authors made a huge effort to produce a comprehensive training set of 367 individual reactions, each of which required multiple calculations to describe all the components, including the variability in shape (the conformations) of each component. It is highly encouraging to see that holistic reaction models can be produced by using such a wide training set.

Where next? A dream for reactivity modellers is to build an ultimate tool that accurately predicts the products of any reaction from the reaction components, thereby allowing computational screening of new reactions. Modellers have a long way to go to achieve this, but Reid and Sigman have shown that they can accurately predict outcomes for groups of related reactions, rather than having to model one type of reaction at a time. Other machine-learning methods are being tested on even bigger data sets[5].

The broadening of reaction scope demonstrated in the current work will encourage the search for more-general models, and might eventually enable models that predict the outcomes of reactions very different from those used for training. For now, making such predictions is still the domain of humans, but synthetic chemists will increasingly rely on theoretical tools to guide their work. I, for one, look forward to a future in which the tedious trial and error of synthetic chemistry is removed, and in which chemists can cut to the chase by carrying out only successful reactions. ■

Per-Ola Norrby *is in Data Science & Modelling, Pharmaceutical Sciences, R&D, AstraZeneca Gothenburg, 43183 Mölndal, Sweden.*
e-mail: per-ola.norrby@astrazeneca.com

1. Reid, J. P. & Sigman, M. S. *Nature* **571**, 343–348 (2019).
2. Brown, J. M. & Deeth, R. J. *Angew. Chem. Int. Ed.* **48**, 4476–4479 (2009).
3. Reid, J. P. & Sigman, M. S. *Nature Rev. Chem.* **2**, 290–305 (2018).
4. Reid, J. P., Simón, L. & Goodman, J. M. *Acc. Chem. Res.* **49**, 1029–1041 (2016).
5. Segler, M. H. S., Preuss, M. & Waller, M. P. *Nature* **555**, 604–610 (2018).

EVOLUTION

A deep dive into sea-squirt development

An analysis of gene expression in sea-squirt embryos at different stages of development deepens our understanding of how the body plans of vertebrates might have evolved from those of less complex animals. SEE ARTICLE P.349

NORIYUKI SATOH

Sea squirts such as *Ciona intestinalis* are the closest living invertebrate relatives of vertebrates. Their tadpole-like larvae feature some of the same organs and tissues as those found in developing vertebrates. On page 349, Cao *et al.*[1] use gene-expression data to examine the embryonic development of *C. intestinalis* larvae and to compare its development with that of other chordate animals, including vertebrates and cephalochordates, to reveal fresh insights into the evolution of vertebrates.

Single-cell analyses of gene expression have revolutionized various biological subdisciplines[2]. Such analyses at different stages of embryonic development have revealed how cells give rise to the various cell types that perform distinct functions and make up specific parts of the embryo[3,4]. As examples, studies of frog and zebrafish embryos have demonstrated that the three layers of cells that form these embryos — the ectoderm, endoderm and mesoderm — contain at least 50 cell types that have similar gene-expression profiles[3,4]. Studies into how different species develop often unveil clues to their evolutionary origins.

There are several advantages to studying embryonic development in sea squirts — which are also known as ascidians. As the closest relatives of vertebrates, they provide a reference for understanding the evolution of vertebrate body plans (Fig. 1). In *C. intestinalis*, embryogenesis — that is, the period of development that begins when cells are initially reorganized into a multilayered body of cells called a gastrula, and ends with larval hatching — takes just a day to complete. A *Ciona* larva comprises only about 2,500 cells, which make up distinctly differentiated organs and systems, including bilateral muscle, the central nervous system (CNS) and the notochord — a rod-like structure that gives rise to the backbone in vertebrates, and which is a defining characteristic of all chordate animals.

The cell lineages that comprise ascidian embryos have long been described[5]; the developmental fate of cells is restricted early in embryogenesis, at around the 110-cell stage. The *C. intestinalis* genome has been sequenced[6], and a network of genes and regulatory molecules that provides the blueprint for the body plan of all chordate animals has been characterized in *C. intestinalis*[7].

Cao *et al.* profiled the gene expression of more than 90,000 single cells from *C. intestinalis* at 10 developmental stages, from gastrulae to swimming larvae. The authors used these gene-expression data — carefully considering the expression of molecular markers of different cell types and lineages — to construct developmental trajectories of individual cell types. Whereas the larvae of *C. intestinalis* were previously thought to have approximately 20 cell types[8], Cao and colleagues' analysis identified 60 distinct cell types. A similarly comprehensive profiling of larval and embryonic cell types in vertebrates