and exhibit residual binding to CD25.

Silva *et al.* therefore began afresh, and set out to design a protein structure from scratch that would provide a stable scaffold onto which they could add the structural elements required to produce the specific protein surfaces that bind to IL-2Rβ and IL-2Rγ. Crucially, these binding surfaces must be correctly positioned relative to each other in space to ensure that the designer cytokine engages the IL-2Rβγ$_c$ heterodimer and triggers signalling.

The authors obtained information about the structural and spatial requirements of their designer cytokine by analysing the crystal structures of naturally occurring cytokine–receptor complexes[7,8]. IL-2 is one of a large family of cytokines that have at their core a bundle of four structural elements termed α-helices (Fig. 1a). These four α-helices are linked, in a defined order, by a series of short or long connecting loops. Instead of keeping this particular arrangement of α-helices and re-engineering the binding surfaces, Silva *et al.* reversed the process. They started by defining the positions of the all-important binding surfaces, and then used computational methods to design an arrangement of α-helices that not only links these surfaces but is also predicted to be stable.

The proof of the pudding is in the eating, however. When the authors prepared and characterized the best candidates from the first round of design, the proteins showed promise in terms of IL-2Rβγ$_c$ binding, but had fairly poor thermal stability. Clearly, the recipe required some improvement. Silva *et al.* went back to the drawing board, taking the best arrangement of α-helices from the first round and substantially extending the computational search for optimal loops to link them together. This second round of design-generated candidates had improved stability and exhibited excellent binding to IL-2Rβγ$_c$.

Silva *et al.* then carried out an additional, experimentally driven round of mutagenesis — a fine-tuning process in which single amino-acid residues are changed — to enhance the binding properties of the best candidate proteins, and then fully characterized the cytokine that had the highest overall binding affinity for IL-2Rβγ$_c$. The results are impressive. The final designer cytokine is highly stable and binds strongly to IL-2Rβγ$_c$, but not at all to CD25. Excitingly, this new protein is effective as a therapy in mouse models of skin and colon cancer, delivering the immunotherapeutic effects characteristic of natural IL-2, but with lower toxicity. The authors named their designer protein Neoleukin-2/15 (Neo-2/15), because it is a new cytokine that mimics natural interleukins 2 and 15.

The researchers then determined the crystal structure of Neo-2/15 in complex with IL-2Rβγ$_c$. Gratifyingly, the binding surfaces are positioned as designed, and the four-helix bundle matches the computational blueprint with almost pinpoint accuracy. The redesign of the interleukin's four-helix bundle achieved by Silva *et al.* is remarkably radical: the order in which the α-helices are linked has been rearranged (Fig. 1b), and the amino-acid sequence of the resulting 100-residue protein is very different from that of either mouse or human IL-2.

It remains to be seen whether Neo-2/15 will deliver on its initial promise in the clinic. Moreover, perhaps the four-helix bundle is a particularly favourable case for re-engineering — other cytokine families that have more-complex architectures might be harder to redesign. Nevertheless, Neo-2/15 excitingly demonstrates that bold *de novo* design, when combined with a deep knowledge of the structural determinants of receptor binding, can deliver designer cytokines that have bespoke binding properties. More broadly, Silva and colleagues' approach to protein design has the potential for re-engineering any of the myriad biological systems that involve interactions between multiple proteins. In the meantime, the authors have opened up uncharted territory for therapeutics based on four-helix bundles, and there is plenty still to explore. ■

E. Yvonne Jones *is in the Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.*
e-mail: yvonne@strubi.ox.ac.uk

1. Silva, D.-A. *et al. Nature* **565**, 186–191 (2019).
2. Rosenberg, S. A. *J. Immunol.* **192**, 5451–5458 (2014).
3. Siegel, J. P. & Puri, R. K. *J. Clin. Oncol.* **9**, 694–704 (1991).
4. Sim, G. C. *et al. J. Clin. Invest.* **124**, 99–110 (2014).
5. Boyman, O. & Sprent, J. *Nature Rev. Immunol.* **12**, 180–190 (2012).
6. Kureshi, R., Bahri, M. & Spangler, J. B. *Curr. Opin. Chem. Eng.* **19**, 27–34 (2018).
7. Wang, X., Rickert, M. & Garcia, K. C. *Science* **310**, 1159–1163 (2005).
8. Stauber, D. J., Debler, E. W., Horton, P. A., Smith, K. A. & Wilson, I. A. *Proc. Natl Acad. Sci. USA* **103**, 2788–2793 (2006).

COMPUTER SCIENCE

Unprovability comes to machine learning

Scenarios have been discovered in which it is impossible to prove whether or not a machine-learning algorithm could solve a particular problem. This finding might have implications for both established and future learning algorithms.

LEV REYZIN

During the twentieth century, discoveries in mathematical logic revolutionized our understanding of the very foundations of mathematics. In 1931, the logician Kurt Gödel showed that, in any system of axioms that is expressive enough to model arithmetic, some true statements will be unprovable[1]. And in the following decades, it was demonstrated that the continuum hypothesis — which states that no set of distinct objects has a size larger than that of the integers but smaller than that of the real numbers — can be neither proved nor refuted using the standard axioms of mathematics[2–4]. Writing in *Nature Machine Intelligence*, Ben-David *et al.*[5] show that the field of machine learning, although seemingly distant from mathematical logic, shares this limitation. They identify a machine-learning problem whose fate depends on the continuum hypothesis, leaving its resolution forever beyond reach.

Machine learning is concerned with the design and analysis of algorithms that can learn and improve their performance as they are exposed to data. The power of this idea is illustrated by the following example: although it seems hopelessly difficult to explicitly program a computer to determine what objects are in a picture, the Viola–Jones machine-learning system can detect human faces in real time after being trained on a labelled sample of photographs[6]. Today, we regularly interact with machine-learning algorithms, from virtual assistants on our phones to spam filters for our e-mail. But these modern real-world applications trace their origins to a subfield of machine learning that is concerned with the careful formalization and mathematical analysis of various machine-learning settings.

The goal of learning a predictor (a mathematical function that can be used to make predictions) from a database of random examples was formalized in the aptly named probably approximately correct (PAC) learning model[7]. In this model, the aim is to train the predictor to match some true function that labels the data. A different model, called online learning, has the learner making immediate predictions as data arrive — for example, capturing a trading system's task of executing transactions in an ever-changing market. And another model known as multi-armed bandits can simulate clinical trials, in which the medical outcomes that an experimenter observes depend on his or her own choices.

These are only a few examples of the many

models used in machine learning. In each case, the basic goal is to perform as well, or nearly as well, as the best predictor in a family of functions, such as neural networks or decision trees. For a given model and function family, if this goal can be achieved under some reasonable constraints, the family is said to be learnable in the model.

Machine-learning theorists are typically able to transform questions about the learnability of a particular function family into problems that involve analysing various notions of dimension that measure some aspect of the family's complexity. For example, the appropriate notion for analysing PAC learning is known as the Vapnik–Chervonenkis (VC) dimension[8], and, in general, results relating learnability to complexity are sometimes referred to as Occam's-razor theorems[9]. These notions of dimension happen to be simple enough to leave no room for the spectre of unprovability to manifest itself. But Ben-David and colleagues show that machine learning cannot always escape this fate. They introduce a learning model called estimating the maximum (EMX), and go on to discover a family of functions whose learnability in EMX is unprovable in standard mathematics.

Ben-David *et al.* describe an example EMX problem: targeting advertisements at the most frequent visitors to a website when it is not known in advance which visitors will visit the site. The authors formalize EMX as a question about a learner's ability to find a function, from a given family, whose expected value over a target distribution is as large as possible. EMX is actually quite similar to the PAC model, but the slightly different learning criterion surprisingly connects it to the continuum hypothesis and brings unprovability into the picture.

The authors' proof involves a beautiful connection between machine learning and data compression that was first observed[10] in the 1980s. The intuition is that, if a training sample labelled by a function from some family can always be compressed, the family must in some sense have low complexity, and therefore be learnable. Moreover, certain learning algorithms can be used to compress data. The authors introduce monotone compression — a variant of compression that they show to be appropriate for characterizing the learnability of particular function families in EMX.

Ben-David and colleagues then prove that the ability to carry out a weak form of monotone compression is related to the size of certain infinite sets. The set that the authors ultimately use in their work is the unit interval, which is the set of real numbers between 0 and 1. Their results imply that the finite subsets of the unit interval have monotone-compression schemes, and therefore are learnable in EMX, if and only if the continuum hypothesis is true, which is known to be unprovable.

Because EMX is a new model in machine learning, we do not yet know its usefulness for developing real-world algorithms. So these results might not turn out to have practical importance. But we do now know that we should be careful when introducing new models of learning. Moreover, we might need to look again at the many subtleties that can come up, even in established learning models.

Machine learning has matured as a mathematical discipline and now joins the many subfields of mathematics that deal with the burden of unprovability and the unease that comes with it. Perhaps results such as this one will bring to the field of machine learning a healthy dose of humility, even as machine-learning algorithms continue to revolutionize the world around us. ■

Lev Reyzin *is in the Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607, USA.*
e-mail: lreyzin@uic.edu

1. Gödel, K. *Monatsh. Math.* **38**, 173–198 (1931).
2. Gödel, K. *The Consistency of the Continuum Hypothesis* (Princeton Univ. Press, 1940).
3. Cohen, P. J. *Proc. Natl Acad. Sci. USA* **50**, 1143–1148 (1963).
4. Cohen, P. J. *Proc. Natl Acad. Sci. USA* **51**, 105–110 (1964).
5. Ben-David, S., Hrubeš, P., Moran, S., Shpilka, A. & Yehudayoff, A. *Nature Mach. Intell.* https://doi.org/10.1038/s42256-018-0002-3 (2019).
6. Viola, P. & Jones, M. J. *Int. J. Comput. Vis.* **57**, 137–154 (2004).
7. Valiant, L. G. *Commun. ACM* **27**, 1134–1142 (1984).
8. Vapnik, V. N. & Chervonenkis, A. Y. *Theory Probab. Appl.* **16**, 264–280 (1971).
9. Blumer, A., Ehrenfeucht, A., Haussler, D. & Warmuth, M. K. *Inf. Process. Lett.* **24**, 377–380 (1987).
10. Littlestone, N. & Warmuth, M. K. *Relating Data Compression and Learnability*, tech. rep. (Univ. California Santa Cruz, 1986).

METABOLISM

Fat cells with a sweet tooth

Some fat cells convert energy into heat, so targeting them to induce weight loss is appealing. The discovery that a subset of the cells burns glucose, rather than both glucose and lipids, could improve our ability to do just that. SEE ARTICLE P.180

WENFEI SUN & CHRISTIAN WOLFRUM

Fat is often thought of as a means to store energy in the form of lipids. But this is just the role of white fat cells. The body also contains a second type of fat that burns the energy stored in nutrients to produce heat, enabling mammals to maintain their body temperature in a cold environment[1]. Activating this 'thermogenic' fat is thought to be an attractive way to combat obesity[2]. On page 180, Chen *et al.*[3] identify a previously uncharacterized type of thermogenic fat cell, which is derived from a hitherto unknown cell lineage

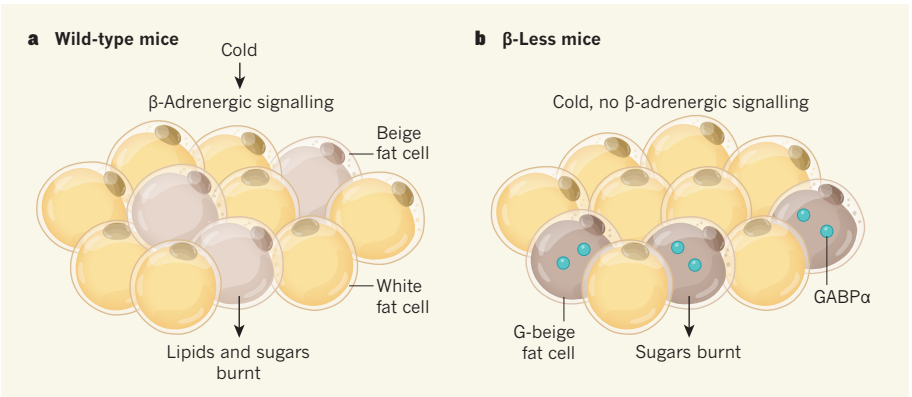

Figure 1 | Dual routes to heat-producing fat. **a**, White fat cells store energy. In wild-type mice under cold conditions, proteins called β-adrenergic receptors on fat-precursor cells (not shown) are activated by a β-adrenergic signalling pathway, which causes some precursors in white-fat tissues to differentiate into beige fat cells. This beige fat burns energy from both lipids and sugar to produce heat, maintaining body temperature. **b**, Chen *et al.*[3] analysed β-less mice, which lack β-adrenergic receptors, meaning that β-adrenergic signalling is blocked. The authors showed that a different type of beige fat, dubbed glycolytic beige fat (g-beige fat), arises in these animals under cold conditions. The differentiation of these cells from a group of precursors that express the muscle-precursor protein MyoD (not shown) is driven by the transcription factor GABPα. The cells burn mainly glucose.