

recent years, and the study by Narasimhan *et al.* is the largest contribution to date, with more than 500 ancient human genomes sequenced (see the figure). This scale of genomic data enabled the authors to compare genomes across a large number of locations and time points and to home in on increasingly specific questions that would have been unanswerable even a few years ago.

The human story revealed by Narasimhan *et al.* in South Asia is similar to those



Narasimhan *et al.* studied the DNA from more than 100 ancient humans unearthed from northernmost South Asia.

from elsewhere in Eurasia: Successive waves of migration altered the genetic makeup of, but did not completely replace, preexisting groups (4–6). Modern South Asians appear to be a mixture of Iranian-like hunter-gatherers, a population ancestral to the Andaman Islands, and Eurasian steppe herders who first settled in Europe. Some South Asian populations later received immigrants from other outside groups. As more genomes become available from previously unexamined historical cultures around the world, stories like this one will fill other middle chapters in the book of human history. ■

REFERENCES AND NOTES

1. V. M. Narasimhan *et al.*, *Science* **365**, eaat7487 (2019).
2. M. Slatkin, F. Racimo, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6380 (2016).
3. I. Lazaridis *et al.*, *Nature* **513**, 409 (2014).
4. W. Haak *et al.*, *Nature* **522**, 207 (2015).
5. M. E. Allentoft *et al.*, *Nature* **522**, 167 (2015).
6. P. de Barros Damgaard *et al.*, *Science* **360**, eaar7711 (2018).
7. D. Reich *et al.*, *Nature* **461**, 489 (2009).
8. A. Basu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1594 (2016).
9. M. Raghavan *et al.*, *Science* **345**, 1255832 (2014).
10. I. Lazaridis *et al.*, *Nature* **536**, 419 (2016).
11. M. Lipson *et al.*, *Nature* **551**, 368 (2017).
12. I. Olalde *et al.*, *Nature* **555**, 190 (2018).

ACKNOWLEDGMENTS

We thank R. E. Green for helpful comments on and discussion about this manuscript.

10.1126/science.aay3550

COMPUTER SCIENCE

Machine learning transforms how microstates are sampled

A deep neural network is trained to optimally explore rugged potential energy landscapes in simulations

By Mark E. Tuckerman

Atomistic simulations of complex molecular systems can provide key microscopic insights not easily accessible to experiments, such as folding of proteins (see the first figure), binding of small-molecule drug candidates or peptide therapeutics to selected targets, and formation of different polymorphs of molecular crystals, provided that two major hurdles are overcome. First, an accurate description of the interatomic interactions is needed, which is captured in a single potential energy function $U(\mathbf{x})$, where \mathbf{x} denotes the full set of atomic coordinates. Second, given $U(\mathbf{x})$, predicting thermodynamic and other equilibrium properties of the system or estimating kinetics requires sampling a statistically sufficient number of realizations of \mathbf{x} from the so-called Boltzmann probability distribution $P(\mathbf{x})$, which is proportional to $\exp[-U(\mathbf{x})/k_B T]$, where T is the system temperature and k_B is Boltzmann's constant. On page 1001 of this issue, Noé *et al.* (1) introduce a machine learning-based approach to address the latter of these two challenges.

The primary difficulty in sampling physical realizations \mathbf{x} of the system (also called “microstates”) from the Boltzmann distribution lies in the nature of $U(\mathbf{x})$ itself. In large, complex systems, \mathbf{x} holds the positions of hundreds of thousands to millions of atoms. Thus, $U(\mathbf{x})$ should be viewed as a vast, rug-

ged “landscape” in this high-dimensional space characterized by an exponentially large number of low-energy regions or minima, all separated by ridges whose energies are typically several to many times $k_B T$ above these minima at room temperature.

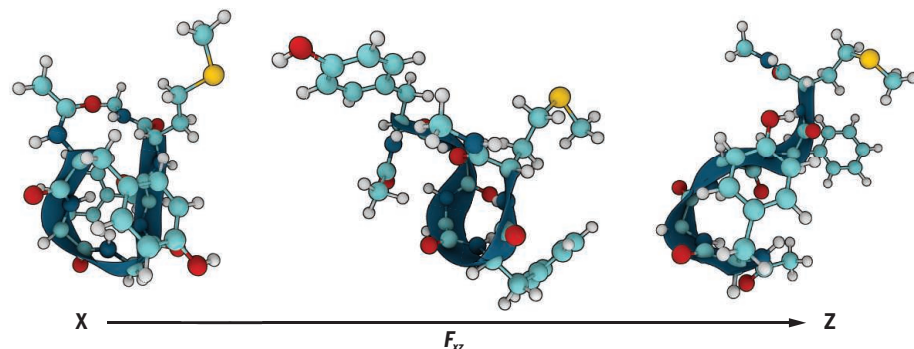
In a Boltzmann distribution, the minima and ridges represent microstates of high and low probability, respectively. It is generally not possible to sample the distribution directly, hence the challenge is to devise an efficient algorithm to explore the $U(\mathbf{x})$ landscape in order to locate these low-energy regions and identify the ridges that constitute the most likely transition pathways between the minima, much as a hiker tries to identify the easiest mountain pass to cross from one valley to another. Physically, minima largely determine thermodynamics, whereas kinetics depend mostly on the ridges.

Numerous approaches have been proposed to perform an efficient exploration of $U(\mathbf{x})$. Some methods bias the search (2); others target a few preselected functions of \mathbf{x} , known as “reaction coordinates,” that are assumed to capture the most relevant features of the Boltzmann distribution (3–9). Such techniques are often quite sensitive to how these biases or reaction coordinates are chosen, and poor choices can waste many hours of computation searching irrelevant regions of $U(\mathbf{x})$.

The ruggedness of $U(\mathbf{x})$ that makes sampling the Boltzmann distribution so challenging originates in the way that physical microstates (coordinates \mathbf{x}) are represented.

Peptide folding as a search problem

Hypothetical peptide configurations that might arise in energy or example training. Images adapted from (13).



However, because physics dictates that observable phenomena cannot depend on the choice of atomic coordinates, an advantage could be gained if an alternative coordinate choice, here denoted as \mathbf{z} , could be identified, together with a transformation $\mathbf{x} \rightarrow \mathbf{z}$, such that the “image” of $U(\mathbf{x})$ in the \mathbf{z} -representation gives a Boltzmann distribution that is easier to sample. An example of such a transformation would be one that brings the minima of $U(\mathbf{x})$ closer together and simultaneously reduces the energy barriers between them.

Previously, my group and collaborators devised such a transformation tailored to linear polymers (which we termed a “spatial warping” transformation) that effectively eliminated barriers in the space of the backbone dihedral angles while simultaneously compressing the distances between conformational minima (10, 11). We reported gains in sampling efficiency of the resulting Boltzmann distribution that were orders of magnitude over those of standard sampling algorithms. The mathematical intricacies of this approach, however, made it difficult to adapt the transformation to other classes of complex systems.

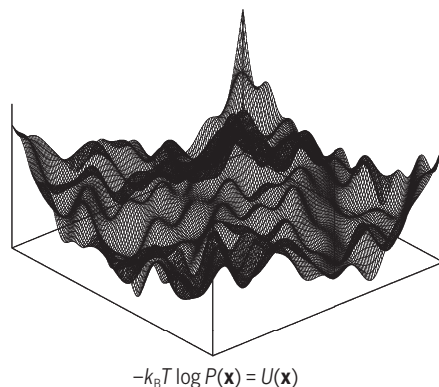
Noé *et al.* now show that the idea of using coordinate transformations can be made fully general and applicable to broad classes of systems by leveraging the power of machine learning (a subset of artificial intelligence). Specifically, they trained a deep neural network to learn a transformation $\mathbf{x} \rightarrow \mathbf{z}$ such that when \mathbf{z} is sampled from a simple probability distribution $[p(\mathbf{z})]$, for example, a Gaussian normal distribution, it maps back, through the inverse transformation $\mathbf{z} \rightarrow \mathbf{x}$, onto a high-probability region of $P(\mathbf{x})$ (see the second figure). They refer to this approach as “Boltzmann generators”

The training phase requires that the neural network be presented with both high- and low-probability states as “good” and “bad” examples, respectively, so that the network can distinguish between the two. Noé *et al.* accomplish this by randomly sampling \mathbf{z} from the desired $p(\mathbf{z})$ and using the resulting wide range of $U(\mathbf{x}(\mathbf{z}))$ values as part of the training set. They call this “training by energy.” If some of the interesting microstates \mathbf{x} are already known for the system, such as the folded state of a protein or some other experimentally determined structures, these microstates can be mapped onto the \mathbf{z} -space and used to “train by example” without having to know the associated probabilities of these example states a priori.

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA, and NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China. Email: mark.tuckerman@nyu.edu

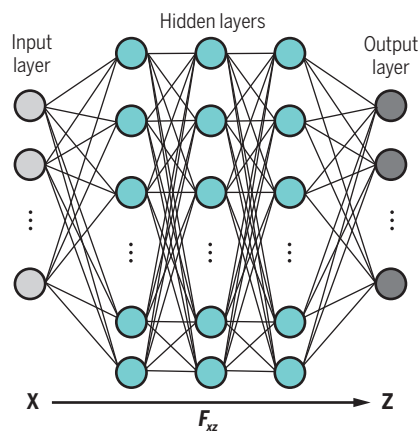
Transforming a rugged landscape

Algorithms that perform searches along a rugged potential energy landscape $U(\mathbf{x})$ can get trapped in local minima. Noé *et al.* used machine learning to identify transformations to a new coordinate \mathbf{z} that create a smoother search surface.



Rough landscapes

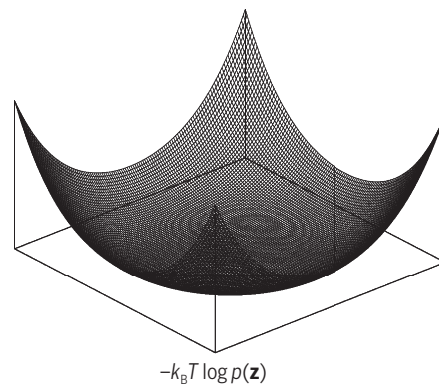
A typical rugged energy surface $U(\mathbf{x})$ can be represented in terms of microstates that have a Boltzmann distribution $P(\mathbf{x})$. The product of k_B (Boltzmann's constant) and T (temperature) is the thermal energy.



Training transformation

During the training phase, a deep neural network learns a function $F_{\mathbf{z}\mathbf{x}}$ mapping $P(\mathbf{x})$ to a smoother distribution $p(\mathbf{z})$ using sample configurations, such as those shown on the previous page, to achieve training by energy or by example.

○ Artificial neurons — Connections



Smoother surfaces

The transformation to \mathbf{z} creates a smoother surface expressed as a Gaussian distribution $p(\mathbf{z})$.

Throughout the training, the $\mathbf{x} \rightarrow \mathbf{z}$ and $\mathbf{z} \rightarrow \mathbf{x}$ maps are continually refined so that once fully trained, high-probability states can each be generated in “one shot” simply by sampling directly from the \mathbf{z} -distribution $p(\mathbf{z})$ and mapping back to \mathbf{x} through the neural network. The breakthrough of this approach is that these transformations can be learned with only minimal knowledge of the system, and no arbitrary reaction coordinates need to be introduced. As a benchmark demonstration of their approach, Noé *et al.* considered the challenging case of the bovine pancreatic trypsin inhibitor (BPTI) protein. Not only could the Boltzmann generators sample with impressive efficiency key metastable conformation states and pathways linking them, which were not included in the training, they also yielded accurate free energy differences between these states. Obtaining these states and their associated free energies using ordinary methods would have required extremely long and costly simulations.

With their new development, Noé *et al.* have opened a promising direction in the computational exploration of complex molecular systems. Nonetheless, additional challenges remain. In the reported applications, the dimension of \mathbf{x} ranged from two to several thousand; for BPTI, the dimensionality is kept low by simulating the protein in an implicit solvent rather than in a bath of explicit water molecules. If large systems (such as explicitly solvated biomolecules or defects in materials) are to be treated, scaling up to hundreds of thousands or millions of dimensions becomes necessary—which, as the authors admit, will require further development of the methodology. This could include exploring how Boltzmann generators can be combined with other widely used and powerful enhanced sampling methods (3–9, 12) to improve their performance. These challenges will likely be solved soon, whereupon the Boltzmann generator approach may well constitute one of the highest-performing techniques for sampling equilibrium distributions. ■

REFERENCES AND NOTES

1. F. Noé *et al.*, *Science* **365**, eaaw1147 (2019).
2. D. Hamelberg *et al.*, *J. Chem. Phys.* **120**, 11919 (2004).
3. G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
4. E. Darve, A. Pohorille, *J. Chem. Phys.* **115**, 9169 (2001).
5. A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
6. L. Rosso *et al.*, *J. Chem. Phys.* **116**, 4389 (2002).
7. L. Maragliano, E. Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006).
8. M. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3235 (2015).
9. J. B. Abrams, M. E. Tuckerman, *J. Phys. Chem. B* **112**, 15742 (2008).
10. Z. Zhu *et al.*, *Phys. Rev. Lett.* **88**, 100201 (2002).
11. P. Minari *et al.*, *SIAM J. Sci. Comput.* **30**, 2055 (2008).
12. Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
13. J. Zhang, M. Chen, *Phys. Rev. Lett.* **121**, 010601 (2018).

ACKNOWLEDGMENTS

Supported by NSF grant CHE-1565980.

10.1126/science.aay2568

Machine learning transforms how microstates are sampled

Mark E. Tuckerman

Science **365** (6457), 982-983.
DOI: 10.1126/science.aay2568

ARTICLE TOOLS

<http://science.sciencemag.org/content/365/6457/982>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/365/6457/eaaw1147.full>

REFERENCES

This article cites 13 articles, 3 of which you can access for free
<http://science.sciencemag.org/content/365/6457/982#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.