Title

Neuroevolutionary Learning of Particles and Protocols for Self-Assembly

Journal

Physical Review Letters, 127(1)

ISSN

0031-9007

Authors

Whitelam, Stephen

Tamblyn, Isaac

Publication Date

2021-07-02

DOI

10.1103/physrevlett.127.018003

Peer reviewed

Neuroevolutionary Learning of Particles and Protocols for Self-Assembly

Stephen Whitelam*
Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, Califronia 94720, USA

Isaac Tamblyn†
National Research Council of Canada Ottawa, Ontario K1N 5A2, Canada
Vector Institute for Artificial Intelligence Toronto, Ontario M5G 1M1, Canada

Within simulations of molecules deposited on a surface we show that neuroevolutionary learning can design particles and time-dependent protocols to promote self-assembly, without input from physical concepts such as thermal equilibrium or mechanical stability and without prior knowledge of candidate or competing structures. The learning algorithm is capable of both directed and exploratory design: it can assemble a material with a user-defined property, or search for novelty in the space of specified order parameters. In the latter mode it explores the space of what can be made, rather than the space of structures that are low in energy but not necessarily kinetically accessible.

Introduction.—How do we make a material with specified properties? In pursuit of "synthesis by design" [1,2] the materials science community has developed and adapted algorithms of inverse design and machine learning. These approaches can identify interparticle potentials able to stabilize target structures or promote their self-assembly from solution [3–23], and can identify protocols or reaction conditions that optimize the self-assembly of specified particles [23–27].

Here we present an approach based on evolutionary learning [28] that simultaneously designs particles *and* protocols in order to self-assemble materials to order. We study a coarse-grained computational model of molecular self-assembly at a surface [29–31]. Coarse-grained models are simple by design [32–45] but can exhibit key features of real systems, including the formation of complex structures and kinetic traps that impair assembly [46–51]. The particular class of model we use here has been shown to reproduce the thermodynamic and dynamic behavior of a range of molecular and nanoscale assemblies at surfaces [52]. Such models provide a rigorous test of algorithmic control of self-assembly.

In order to allow thorough exploration of the self-assembly behavior accessible to this class of model we express the interparticle potential and time-dependent assembly protocol as arbitrary functions, encoded by neural networks. In evolutionary language, which reflects the method of learning used and provides a mnemonic for the role of each component of the algorithm, this encoding is the instruction code or "genome" for self-assembling a material. Molecular simulation carried out using the particle and protocols specified by the genome results in the "phenome," a material whose properties can be measured and compared to a design goal.

Evolutionary learning on the parameters of the neural networks, called neuroevolution [28,53–57], can produce materials whose properties satisfy user-defined goals, which can be directed or exploratory in nature. Current approaches to inverse design involve specifying a desired structure; the present algorithm needs no information about possible candidate or competing structures, nor prior knowledge of what constitutes a good self-assembly protocol or particle design. For directed design we specify materials with certain pore geometries and isolated clusters of certain sizes. The solutions identified by the learning algorithm include interparticle interactions whose symmetries can be realized by known molecules or complexes [29–31]. The approach is simple to implement and can identify sophisticated design strategies, realizing complex structures via hierarchical self-assembly pathways. It can also fail, if the design goal is too challenging, and we discuss how to modify the goal in such cases.

Our approach is similar to that of Refs. [8,9] in that we use an iterative learning method to promote self-assembly, but differs in that we do not specify or build the target structure in advance. In that respect it is similar to the approach of Ref. [23], and complementary to that work in that we express the design problem differently (in the form of neural networks) and optimize differently (via evolutionary methods). Our approach differs from other approaches to inverse design in that we consider the design of particles and protocols simultaneously. Physical laws are built into the molecular simulation protocol in a standard way [58]. The learning algorithm must operate within these laws, but does not directly appeal to concepts such thermal equilibrium or mechanical stability that are inputs to other design protocols; instead, it attempts to control only what is

kinetically accessible. We also go beyond traditional forms of inverse design and borrow from the machine-learning literature to specify the design goal of *novelty* [59]. Within a space specified by certain order parameters we instruct the evolutionary learning algorithm to produce materials not seen previously, rather than materials with particular properties. If we specify materials with three- and four-membered pores then novelty search identifies structures dual to regular and semiregular tilings of the plane and motifs that comprise 2D quasicrystals. In this mode the algorithm explores the space of what can be made, rather than the space of structures that are low in energy but not necessarily kinetically accessible.

Model and learning algorithm.—We consider a class of coarse-grained model able to capture the essential physics of molecular and nanoscale self-assembly at surfaces [52]. It comprises circular particles of hard-core diameter a on a two-dimensional square substrate of side $50a$. The substrate has periodic boundary conditions in both directions. Particles evolve under a stochastic dynamical protocol consisting of a grand-canonical Monte Carlo algorithm with chemical potential μ, which allows particles to exchange with a notional solution [58], and the virtual-move Monte Carlo algorithm [60,61], which allows particles to move on the surface according to an approximation of Brownian motion [62]. Grand-canonical moves are proposed with probability $1/(1+P)$, where P is the instantaneous number of particles on the surface [52]. All trajectories start from distinct disordered configurations consisting of 500 particles randomly deposited on the surface, and are run for $t_0 = 10^9$ Monte Carlo steps. The angular component of the interparticle attraction $U_{\boldsymbol{x}}(\theta)$ and the time-dependent protocol $(\epsilon_{\boldsymbol{y}}(t), \mu_{\boldsymbol{y}}(t))$ are encoded as neural networks, specified in Sec. S1 of the Supplemental Material [63]. Simulation potentials with angular dependence are often called "patchy" [64]. The interaction potential reflects the idea that particles interact in a complementary way, such as through DNA hybridization, hydrogen bonding, or other directional donor-acceptor mechanisms [29–31,65]. (Intermolecular potentials depending on additional parameters, such as spherical polar coordinates in three dimensions, can be expressed using neural networks with additional input nodes.) The control parameters μ and ϵ influence the substrate density and the strength of interparticle attractions. We denote the parameters of the neural networks by $\boldsymbol{x}$ and $\boldsymbol{y}$. Using evolutionary language we call this encoding the material's *genome*, an idea sketched in Fig. 1. The result of molecular simulation, using the potential and protocol defined by the genome, is the *phenome*.

To evolve genomes whose phenomes possess a desired property we use an iterative genetic algorithm consisting of a population dynamics combined with neuroevolution [26,28,53–56], specified in Sec. S2 of the Supplemental Material [63]. Neuroevolution, stochastic mutation of the neural-network parameters, is equivalent in the limit of small mutations to gradient descent in the presence of Gaussian white noise [57]. The learning algorithm starts in Generation 0 with a population of 100 randomly chosen genomes, and "expresses" their phenomes via t_0 steps of the molecular simulation protocol described above. The algorithm identifies the 10 phenomes possessing the largest values of an objective function ϕ. The 10 corresponding genomes are cloned and mutated in order to produce a new population (or generation) of 100 genomes, whose phenomes are then expressed via molecular simulation. This iterative procedure continues, generation by generation, until terminated by the user.

The objective function or evolutionary pressure is a user-specified order parameter ϕ, evaluated at the final time point of each simulation. In this work we consider order parameters built from two quantities. One is C_k, the number of clusters of interacting particles of size k (called k-mers).

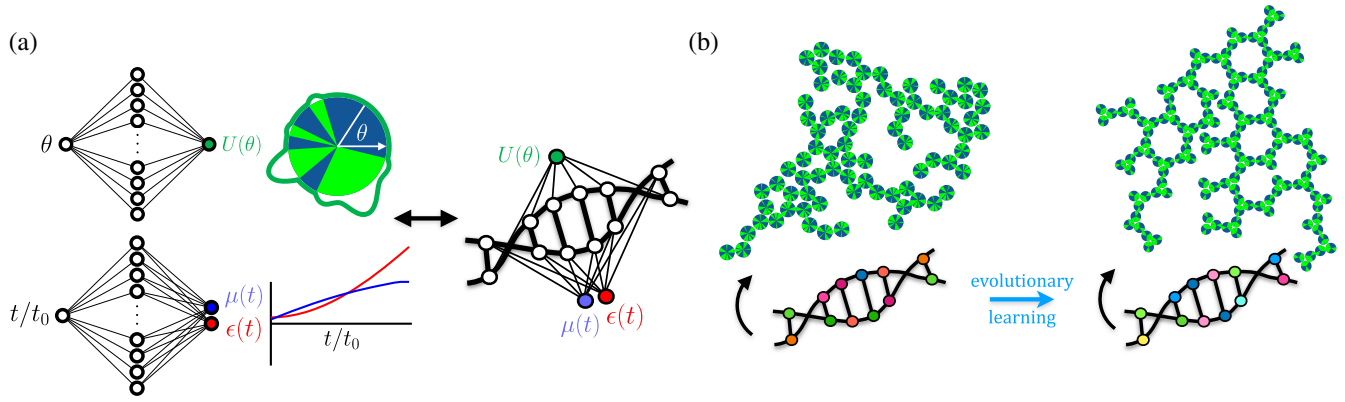

FIG. 1. (a) We express the angular interaction potential and the time-dependent self-assembly protocol for a set of model molecules in the form of two neural networks, which together comprise the "genome" for making a material. In images, attractive portions of particles are shown green. (b) Upon specifying a design goal, neuroevolutionary learning can produce a genome whose "phenome"—the result of molecular simulations carried out using the potential and protocol specified by the genome—is a material satisfying that goal.

The other is N_k, the number of convex loops of size k (called k-gons) that can be drawn by joining the centers of interacting particles. k-gons are pores: designing materials with specified pore sizes is useful, e.g., for gas separation [66]. In general, the method can work with any order parameter the user wishes to specify, including structural or dynamic measures, and can work with measures of novelty that do not require the specification of a particular outcome. Given a choice of order parameter, success is not guaranteed: if an overly challenging design goal is chosen, the random particle- and protocol search that constitutes the first stage of learning may fail to identify any solutions consistent with the goal. Such difficulties can be overcome using *curriculum learning*, choosing design goals of progressively increasing difficulty [67]. We illustrate this strategy in two cases, directing the algorithm to make a material with 12-sided pores (the intermediate goal being smaller pores) and isolated pentagonal clusters (which require a design goal maximizing 5-mers *and* 5-gons).

In our approach the learning algorithm attempts to identify a molecular symmetry [the potential $U(\theta)$] and a time-dependent protocol $(\epsilon(t), \mu(t))$ in order to achieve a particular outcome. We are not modeling experimental potentials that change with time: the potential $U(\theta)$ is fixed in time, during molecular simulations, and changes only *between* evolutionary generations. The idea behind this search strategy is the following: once a particular design is identified, a user could in principle consult a list of known molecules or complexes, and select those whose interactions display the required symmetries [29–31]. Note that key features of the self-assembly of a range of molecular and nanoscale systems can be reproduced by capturing the basic symmetries and energy scales of the building blocks in question. For instance, the three-patch particle identified in Fig. 3(c) can describe in a coarse way the equilibrium and nonequilibrium behavior of structures made by carbon atoms, a silica network, DNA nanostructures, and organic molecules possessing covalent interactions or hydrogen bonds [52,68–72].

The second element of our search scheme involves a time-dependent protocol. Such protocols can be achieved in a wide range of laboratory experiments. The parameter ϵ represents the strength (the basic energy scale) of interparticle binding. In laboratory systems this quantity can be controlled in a variety of ways, including by variation of temperature, magnetic field strength, pH, or the concentration of salt, depletion agents, or other complexes [65,73–76]. For brevity, we use the term "cooling" to refer to the process of increasing ϵ. The parameter μ controls the concentration of the notional vapor phase in contact with the two-dimensional substrate, which can be straightforwardly changed in, e.g., vapor-deposition experiments [77]. More generally, however, the protocol-learning algorithm can work with any quantity that can be varied as a function of time. It could be applied directly (independent of the molecular simulations) to any experimental system in which (1) one or more variables can be controlled as a function of time, and (2) the outcome of the experiment can be characterized. Based on the number of trajectories (independent experiments) and evolutionary cycles required to identify solutions, it is likely that a reasonable rate of learning would require at least (3) a few tens of experiments to be carried out per week. The conditions (1)–(3) admit a wide variety of experimental setups.

Directed search.—We direct the algorithm to evolve a material containing convex pores of size 12, and set $\phi = N_{12}$. This case provides an example of an objective that is too complex to achieve without additional guidance: 12-gons are sufficiently complex that they do not form spontaneously under the random particle- and protocol design that comprises the initial stage (Generation 0) of the learning algorithm. All phenomes score zero, and the learning algorithm has nothing to work with. In this case a simple modification of the objective is sufficient to overcome the problem. We set $\phi = N_{\min(x,12)}$, where x is the size of the largest convex pore seen across all 100 phenomes of a given generation. Thus if x is 12 or larger then the learning algorithm selects genomes that produce 12-gons; if x is smaller than 12 then it selects genomes that produce x-gons.

The results of several generations of learning using this objective are shown in Fig. 2. The largest pore sizes seen in the first four generations are 10, 9, 11, and 11, and thereafter the first 12-gons are produced. The learning algorithm improves its design and the yield of 12-gons over evolutionary time, and eventually achieves the self-assembly of a structure dual to the 3.12.12 Archimedean tiling, which has one 3-gon and two 12-gons around each vertex [78–80]. To do so requires a sophisticated design. The particle must present sticky patches whose bisectors are separated (approximately) by angles $\pi/3$ and $5\pi/6$. In addition, the patches must be inequivalent: if all patches possess equal binding energy then kinetic traps impair the formation of the structure [80]. The solution identified by the learning algorithm is to make one patch weaker than the other two, and to steadily cool the substrate. The result is a hierarchical dynamics that starts with many isolated 3-gons forming from the engagement of the strong patches. Eventually the weak patches engage and cause the 3-gons to form a network, which subsequently forms 12-gons (Fig. S2). The learning algorithm also evacuates the substrate, removing steric impediments to closure of the network. The resulting strategy produces a yield of 12-gons superior to that achieved by a human-designed particle and protocol [Fig. S3(a)].

In Fig. 3 we show the results of evolutionary learning instructed to make 4-mers ($\phi = C_4$), 5-mers and 5-gons [$\phi = \min(C_5, N_5)$], and 6-gons ($\phi = N_6$). In each case the strategy learned is efficient: to produce 4-mers the algorithm evolves particles with two patches separated by an angle $\pi/2$, leading to compact square clusters; to make

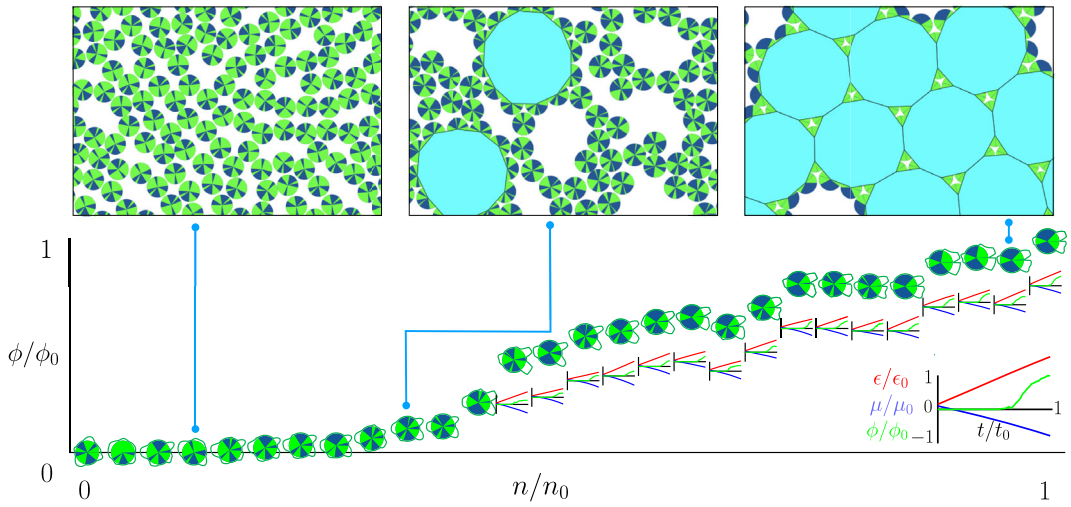

FIG. 2. Evolutionary learning directed to maximize the number of 12-gons, convex pores of 12 sides (shown light blue in images). The main panel shows the yield ϕ, as a function of generation n, produced by the most successful genome (potential and protocol). The particle symbols are the data points, which also show the form of the learned potential. Below the data points are plots of the learned protocol (red and blue lines) and resulting yield (green lines) as a function of time t. The format of those plots is shown bottom right. Large positive values of ϵ indicate strong particle attractions, and large positive and negative values of μ promote dense and sparse substrates, respectively. The snapshots at the top show portions of a simulation box from three different generations, indicated by the blue lines. Parameters: $\phi_0 = 90$ 12-gons, $\epsilon_0 = \mu_0 = 20k_BT$, $n_0 = 27$ generations, $t_0 = 10^9$ Monte Carlo steps.

pentagonal 5-gons it evolves particles with patches separated by an angle $3\pi/5$; and to make 6-gons the algorithm evolves particles with approximate threefold rotational symmetry, which self-assemble under the learned protocol into the honeycomb lattice. Similar motifs are seen in a range of real systems that realize the honeycomb lattice [29–31,52]. The evolutionary pressure to achieve geometrical perfection of the interaction is relatively weak:

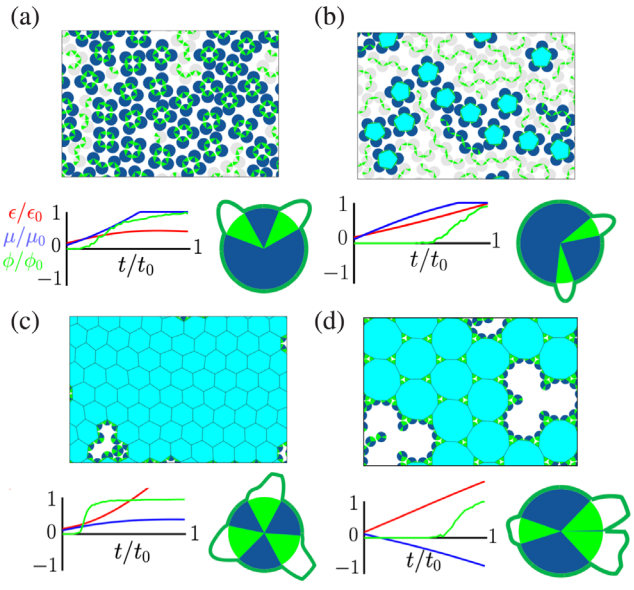

FIG. 3. The results of evolutionary learning (genome bottom and phenome top) instructed to produce (a) 4-mers, (b) 5-mers and 5-gons, (c) 6-gons, and (d) 12-gons.

self-assembly of a particle with perfect threefold rotational symmetry and a learned time-dependent protocol results in a comparable yield of 6-gons [Fig. S3(b)]. The evolutionary trajectories showing the emergence of these designs are shown in Figs. S4, S5, and S6.

Exploratory search.—We end by showing that evolutionary learning can be used in an exploratory mode, searching for novelty rather than to achieve a desired property [59]. To search for novelty within the space of 4-gons and 3-gons we impose the objective function

$$\phi = \sum_j \sqrt{\frac{1}{4}(N_3 - N_3^{(j)})^2 + (N_4 - N_4^{(j)})^2}, \qquad (1)$$

where j runs over all phenomes produced in all previous and current generations. Maximizing Eq. (1) leads to an evolutionary pressure favoring materials most unlike those produced to date, rather than materials with specified values of N_3 and N_4. Over the course of ten generations of novelty search the learning algorithm produces the coverage of (N_3, N_4) space shown in Fig. 4. Some of the polygon structures dual to the particle structures found during that exploration are shown in the figure. These include size-limited motifs; the square and triangle regular tilings of the plane; the 3.6.3.6 and 3.3.3.4.4 Archimedean tilings [78–80]; and the σ, H, and Z binding motifs prominent in dodecagonal quasicrystals [81]. The particle that gives rise to those motifs is shown in the figure: it has irregular sixfold symmetry, different to the fivefold and sevenfold coordination known to form similar motifs [81].

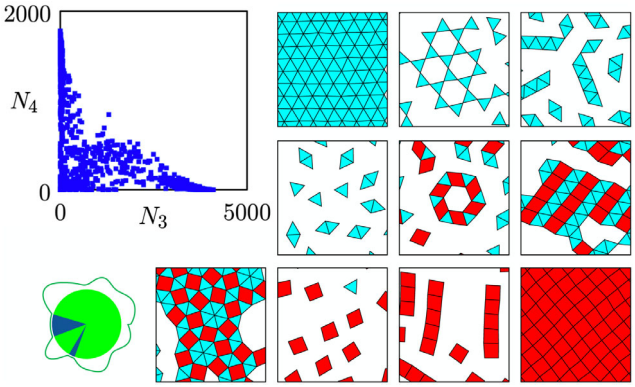

FIG. 4. The results of evolutionary learning instructed to produce novelty in the space of 3-gons and 4-gons. The scatterplot shows the coverage obtained in this space after ten generations of learning; the images show examples of the structures produced (the particles underlying the polygons are not shown).

A larger section of the material made by this particle is shown in Fig. S7.

Conclusions.—We have shown that a neuroevolutionary learning algorithm can identify particles and protocols for the self-assembly of materials with desired properties, without input from physical principles and with no prior knowledge of self-assembly. The learning algorithm is capable of both directed and exploratory design. It can assemble a material with a user-defined property, or search for novelty in the space of specified order parameters. In the latter mode it explores the space of what can be made rather than the space of structures that are low in energy but not necessarily kinetically accessible. Moreover, the approach described here can be used to address design problems of considerable complexity. The neural-network encoding of potential and protocol extends to an arbitrary number of inputs and outputs—one could use state-dependent information to inform the protocol [26], or express the intermolecular potential as a function of additional variables, such as the angles required to define a potential in 3D—and evolutionary learning works with large numbers of parameters [54,56].

This work was performed as part of a user project at the Molecular Foundry, Lawrence Berkeley National Laboratory, supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02–05CH11231. I. T. acknowledges NSERC and performed work at the NRC under the auspices of the AI4D and MCF Programs.

*swhitelam@lbl.gov
†isaac.tamblyn@nrc.ca

[1] J. De Yoreo, D. Mandrus, L. Soderholm, T. Forbes, M. Kanatzidis, J. Erlebacher, J. Laskin, U. Wiesner, T. Xu, S. Billinge *et al.*, Basic research needs workshop on synthesis science for energy relevant technology, Technical Report, USDOE Office of Science (SC), United States, 2016.
[2] C. Broholm, I. Fisher, J. Moore, M. Murnane, A. Moreo, J. Tranquada, D. Basov, J. Freericks, M. Aronson, A. MacDonald *et al.*, Basic research needs workshop on quantum materials for energy relevant technology, Technical Report, USDOE Office of Science (SC), United States, 2016.
[3] H. Cohn and A. Kumar, Proc. Natl. Acad. Sci. U.S.A. **106**, 9570 (2009).
[4] S. Torquato, Soft Matter **5**, 1157 (2009).
[5] E. Bianchi, G. Doppelbauer, L. Filion, M. Dijkstra, and G. Kahl, J. Chem. Phys. **136**, 214102 (2012).
[6] M. C. Rechtsman, F. H. Stillinger, and S. Torquato, Phys. Rev. E **74**, 021404 (2006).
[7] I. G. Johnston, S. E. Ahnert, J. P. K. Doye, and A. A. Louis, Phys. Rev. E **83**, 066105 (2011).
[8] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, J. Chem. Phys. **145**, 111101 (2016).
[9] R. Jadrich, B. Lindquist, and T. Truskett, J. Chem. Phys. **146**, 184103 (2017).
[10] A. W. Long and A. L. Ferguson, Mol. Syst. Des. Eng. **3**, 49 (2018).
[11] K. R. Gadelrab, A. F. Hannon, C. A. Ross, and A. Alexander-Katz, Mol. Syst. Des. Eng. **2**, 539 (2017).
[12] A. L. Ferguson, J. Phys. Condens. Matter **30**, 043002 (2018).
[13] W. D. Piñeros, B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, J. Chem. Phys. **148**, 104509 (2018).
[14] G. van Anders, D. Klotsa, A. S. Karas, P. M. Dodd, and S. C. Glotzer, ACS Nano **9**, 9542 (2015).
[15] C. S. Adorf, J. Antonaglia, J. Dshemuchadse, and S. C. Glotzer, J. Chem. Phys. **149**, 204102 (2018).
[16] J. Madge and M. A. Miller, Soft Matter **13**, 7780 (2017).
[17] X. Jiang, J. Li, V. Lee, H. M. Jaeger, O. G. Heinonen, and J. J. de Pablo, J. Chem. Phys. **148**, 234302 (2018).
[18] R. Kumar, G. M. Coli, M. Dijkstra, and S. Sastry, J. Chem. Phys. **151**, 084109 (2019).
[19] P. Zhou, J. C. Proctor, G. van Anders, and S. C. Glotzer, Mol. Phys. **117**, 3968 (2019).
[20] Z. M. Sherman, M. P. Howard, B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, J. Chem. Phys. **152**, 140902 (2020).
[21] A. Reinhardt and D. Frenkel, Phys. Rev. Lett. **112**, 238103 (2014).
[22] F. Romano, J. Russo, L. Kroc, and P. Šulc, Phys. Rev. Lett. **125**, 118003 (2020).
[23] M. Z. Miskin, G. Khaira, J. J. de Pablo, and H. M. Jaeger, Proc. Natl. Acad. Sci. U.S.A. **113**, 34 (2016).
[24] D. Klotsa and R. L. Jack, J. Chem. Phys. **138**, 094502 (2013).
[25] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, Nature (London) **533**, 73 (2016).
[26] S. Whitelam and I. Tamblyn, Phys. Rev. E **101**, 052604 (2020).
[27] X. Tang, B. Rupp, Y. Yang, T. D. Edwards, M. A. Grover, and M. A. Bevan, ACS Nano **10**, 6791 (2016).

[28] J. H. Holland, Sci. Am. **267**, 66 (1992).
[29] L. Bartels, Nat. Chem. **2**, 87 (2010).
[30] J. A. A. W. Elemans, S. Lei, and S. De Feyter, Angew. Chem. Int. Ed. **48**, 7298 (2009).
[31] G. F. Swiegers and T. J. Malefetse, Coord. Chem. Rev. **225**, 91 (2002).
[32] J. P. K. Doye, A. A. Louis, and M. Vendruscolo, Phys. Biol. **1**, P9 (2004).
[33] M. F. Hagan and D. Chandler, Biophys. J. **91**, 42 (2006).
[34] V. Molinero and E. B. Moore, J. Phys. Chem. B **113**, 4008 (2009).
[35] F. Romano and F. Sciortino, Nat. Mater. **10**, 171 (2011).
[36] S. Glotzer, M. Solomon, and N. A. Kotov, AIChE J. **50**, 2978 (2004).
[37] J. P. K. Doye, A. A. Louis, I. C. Lin, L. R. Allen, E. G. Noya, A. W. Wilber, H. C. Kok, and R. Lyus, Phys. Chem. Chem. Phys. **9**, 2197 (2007).
[38] D. C. Rapaport, Phys. Biol. **7**, 045001 (2010).
[39] A. Murugan, J. Zou, and M. P. Brenner, Nat. Commun. **6**, 1 (2015).
[40] S. Whitelam and R. L. Jack, Annu. Rev. Phys. Chem. **66**, 143 (2015).
[41] M. Grunwald and P. L. Geissler, ACS Nano **8**, 5891 (2014).
[42] M. Nguyen and S. Vaikuntanathan, Proc. Natl. Acad. Sci. U.S.A. **113**, 14231 (2016).
[43] J. F. Lutsko, Sci. Adv. **5**, eaav7399 (2019).
[44] Z. Fan and M. Grunwald, J. Am. Chem. Soc. **141**, 1980 (2019).
[45] J. E. Carpenter and M. Grunwald, J. Am. Chem. Soc. **142**, 10755 (2020).
[46] K. Thorkelsson, P. Bai, and T. Xu, Nano Today **10**, 48 (2015).
[47] P. L. Biancaniello, A. J. Kim, and J. C. Crocker, Phys. Rev. Lett. **94**, 058302 (2005).
[48] S. Y. Park, A. K. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, Nature (London) **451**, 553 (2008).
[49] D. Nykypanchuk, M. M. Maye, D. van der Lelie, and O. Gang, Nature (London) **451**, 549 (2008).
[50] W. Pfeifer and B. Saccà, Biol. Chem. **399**, 773 (2018).
[51] J. J. De Yoreo, P. U. Gilbert, N. A. Sommerdijk, R. L. Penn, S. Whitelam, D. Joester, H. Zhang, J. D. Rimer, A. Navrotsky, J. F. Banfield *et al.*, Science **349**, aaa6760 (2015).
[52] S. Whitelam, I. Tamblyn, T. K. Haxton, M. B. Wieland, N. R. Champness, J. P. Garrahan, and P. H. Beton, Phys. Rev. X **4**, 011044 (2014).
[53] D. B. Fogel and L. C. Stayton, BioSystems **32**, 171 (1994).
[54] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever, arXiv:1703.03864.
[55] D. J. Montana and L. Davis, in *IJCAI*, Vol. 89 (1989), pp. 762–767.
[56] F. P. Such, V. Madhavan, E. Conti, J. Lehman, K. O. Stanley, and J. Clune, arXiv:1712.06567.
[57] S. Whitelam, V. Selin, S.-W. Park, and I. Tamblyn, arXiv:2008.06643.
[58] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, Inc. Orlando, FL, USA, 1996).
[59] E. Conti, V. Madhavan, F. P. Such, J. Lehman, K. Stanley, and J. Clune, in *Adv. Neural Inf. Process. Syst.* (2018), pp. 5027–5038.
[60] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, Soft Matter **5**, 1251 (2009).
[61] L. O. Hedges, Libvmmc, http://vmmc.xyz.
[62] T. K. Haxton, L. O. Hedges, and S. Whitelam, Soft Matter **11**, 9307 (2015).
[63] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.018003 for additional figures and details of the simulation protocol.
[64] Z. Zhang and S. C. Glotzer, Nano Lett. **4**, 1407 (2004).
[65] W. Pfeifer and B. Saccà, Chembiochem **17**, 1063 (2016).
[66] G. Liu, W. Jin, and N. Xu, Angew. Chem. Int. Ed. **55**, 13384 (2016).
[67] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, in *Proceedings of the 26th Annual International Conference on Machine Learning* (2009), pp. 41–48.
[68] A. K. Geim and K. S. Novoselov, Nat. Mater. **6**, 183 (2007).
[69] Y. He, Y. Chen, H. Liu, A. E. Ribbe, and C. Mao, J. Am. Chem. Soc. **127**, 12202 (2005).
[70] L. Lichtenstein, M. Heyde, and H.-J. Freund, Phys. Rev. Lett. **109**, 106101 (2012).
[71] C.-A. Palma, P. Samorì, and M. Cecchini, J. Am. Chem. Soc. **132**, 17880 (2010).
[72] M. Bieri, M. Treier, J. Cai, K. Aït-Mansour, P. Ruffieux, O. Gröning, P. Gröning, M. Kastler, R. Rieger, X. Feng *et al.*, Chem. Commun., 6919 (2009).
[73] J. E. Martin and A. Snezhko, Rep. Prog. Phys. **76**, 126601 (2013).
[74] A. Mersmann, *Crystallization Technology Handbook* (CRC Press, New York, 2001).
[75] G. E. Fernandes, D. J. Beltran-Villegas, and M. A. Bevan, Langmuir **24**, 10776 (2008).
[76] P. Yin, T. Li, R. S. Forgan, C. Lydon, X. Zuo, Z. N. Zheng, B. Lee, D. Long, L. Cronin, and T. Liu, J. Am. Chem. Soc. **135**, 13425 (2013).
[77] D. Dobkin and M. K. Zuraw, *Principles of Chemical Vapor Deposition* (Springer Science & Business Media, New York, 2003).
[78] B. Grunbaum and G. C. Shephard, Math. Mag. **50**, 227 (1977).
[79] M. Antlanger, G. Doppelbauer, and G. Kahl, J. Phys. Condens. Matter **23**, 404206 (2011).
[80] S. Whitelam, Phys. Rev. Lett. **117**, 228003 (2016).
[81] M. N. van der Linden, J. P. Doye, and A. A. Louis, J. Chem. Phys. **136**, 054904 (2012).