Void distributions reveal structural link between jammed packings and protein cores

John D. Treado,[1,2] Zhe Mei,[2,3] Lynne Regan,[4,3,2,*] and Corey S. O'Hern[1,2,5,6,7,†]

[1]*Department of Mechanical Engineering & Materials Science, Yale University, New Haven, Connecticut 06520, USA*
[2]*Integrated Graduate Program in Physical & Engineering Biology, Yale University, New Haven, Connecticut 06520, USA*
[3]*Department of Chemistry, Yale University, New Haven, Connecticut 06520, USA*
[4]*Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, Connecticut 06520, USA*
[5]*Department of Physics, Yale University, New Haven, Connecticut 06520, USA*
[6]*Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA*
[7]*Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA*

(Received 31 October 2018; published 21 February 2019)

Dense packing of hydrophobic residues in the cores of globular proteins determines their stability. Recently, we have shown that protein cores possess packing fraction $\phi \approx 0.56$, which is the same as dense, random packing of amino-acid-shaped particles. In this article, we compare the structural properties of protein cores and jammed packings of amino-acid-shaped particles in much greater depth by measuring their local and connected void regions. We find that the distributions of surface Voronoi cell volumes and local porosities obey similar statistics in both systems. We also measure the probability that accessible, connected void regions percolate as a function of the size of a spherical probe particle and show that both systems possess the same critical probe size. We measure the critical exponent τ that characterizes the size distribution of connected void clusters at the onset of percolation. We find that the cluster size statistics are similar for void percolation in packings of amino-acid-shaped particles and randomly placed spheres, but different from that for void percolation in jammed sphere packings. We propose that the connected void regions are a defining structural feature of proteins and can be used to differentiate experimentally observed proteins from decoy structures that are generated using computational protein design software. This work emphasizes that jammed packings of amino-acid-shaped particles can serve as structural and mechanical analogs of protein cores, and could therefore be useful in modeling the response of protein cores to cavity-expanding and -reducing mutations.

DOI: 10.1103/PhysRevE.99.022416

I. INTRODUCTION

A significant driving force in protein folding is the sequestration of hydrophobic amino acids from solvent. Moreover, these buried amino acids are densely packed in the protein core [1]. In fact, the packing of core residues has been linked directly to protein stability [2]. For example, large-to-small amino acid mutations, which can increase interior protein cavities, or voids, are known to destabilize proteins when they are subjected to hydrostatic pressure [3–5] and chemical denaturants [6,7]. Determining the connection between dense core packing and voids is therefore crucial to understanding the physical origins of protein stability and reliably designing new protein structures that are stable [8]. However, no such quantitative understanding yet exists, and it is currently difficult to distinguish computational protein designs that are not stable in experiments from experimentally observed structures [9].

In previous studies [10–12], we found, using collective side chain repacking, that the side chain conformations of residues in protein cores (from a collection of high-resolution protein crystal structures) are uniquely specified by hard-sphere, steric interactions. Moreover, we have shown that, when considering hard-sphere optimized atomic radii, the core regions in proteins possess the same packing fraction $\phi \approx 0.56$ as that found in simulations of dense, random packings of purely repulsive, amino-acid-shaped particles. This result suggests that the packing fraction of protein cores is determined by the bumpy and nonsymmetric shape of amino acids, and not by the backbone or local secondary structure.

However, materials that share the same packing fraction do not necessarily possess the same internal structure. In this article, we characterize the void space in experimentally obtained and computationally generated protein cores to further test the geometric similarities between these two systems. We show below that dense random packings of amino-acid-shaped particles have the same local packing fraction, void distribution, and percolation of connected void space as protein cores, which indicates structural equivalence.

Our results suggest that the computationally generated packings can be used as mechanical analogs of protein cores to predict their collective mechanical response. Further, our results emphasize the connection between structurally arrested, yet thermally fluctuating, protein cores and the jamming transition of highly nonspherical particles [13]. Although the similarity between structural glasses and proteins

*Present address: Centre for Synthetic and Systems Biology, Institute for Quantitative Biology, Biochemistry, and Biotechnology, School of Biological Sciences, Edinburgh University, Scotland, UK.
†corey.ohern@yale.edu

at low temperatures has been known for several decades [14–18], prior computational studies have mainly focused on the transition from harmonic to anharmonic conformational fluctuations on length scales spanning the full protein. In contrast, our studies identify key structural similarities between jammed packings of amino-acid-shaped particles and the cores of protein crystal structures.

This article is organized into four sections and three appendices. In Sec. II, we describe the database of high-resolution protein crystal structures that we use for our structural analyses and the computational methods we use to generate jammed packings of amino-acid-shaped particles. We also outline two methods to measure the void distribution in the two systems: a local measure of void space using surface Voronoi tessellation, and a nonlocal or "connected" measure of void space similar to that used by Kertèsz [19] and Cuff and Martin [20]. In Sec. III, we compare the results of both the local and connected void measurements for jammed packings of amino-acid-shaped particles and protein cores and find that both void measurements are the same for both systems. In Sec. III A, we show that the Voronoi cell volume distributions in both systems are described by a k-gamma distribution with similar shape factors k. In addition, we find that the distribution of the local porosity ($\eta = 1 - \phi$) is the same for protein cores and jammed packings of amino-acid-shaped particles. In Sec. III B, we identify the void percolation transition as a function of the probe particle accessibility for the connected voids, and find that protein cores and jammed packings of amino-acid-shaped particles share the same critical probe size that separates the percolating and nonpercolating regimes. In Sec III C, we investigate the critical properties of this percolation transition, and show that they are similar to void percolation of systems of randomly placed spheres, but distinct from void percolation in jammed sphere packings. In Sec. IV, we summarize our results, discuss their importance, and identify future research directions. We include three appendices with additional details of our computational methods. In Appendix A, we provide details for the computational method we use to generate jammed packings of amino-acid-shaped particles. In Appendix B, we discuss the differences between protein cores in the Dunbrack 1.0 database, and the core replicas we generate from jammed packings of amino-acid-shaped particles. In Appendix C, we discuss the differences between the connected void cluster size distributions in the systems considered in Sec. III C.

II. METHODS

To benchmark our studies of local and connected void regions, we use a subset of the Dunbrack PISCES Protein Database (PDB) culling server [21,22] of high-resolution protein crystal structures. This dataset, which we will refer to as "Dunbrack 1.0," contains 221 proteins with $< 50\%$ sequence identity, resolution $\leqslant 1.0$ Å, side chain B factors per residue $\leqslant 30$ Å^2 and R factor $\leqslant 0.2$. We add hydrogen atoms to each protein crystal structure using the REDUCE software [23]. To determine core amino acids, we calculate the solvent accessible surface area (SASA) for each residue using the NACCESS software [24] with a 1.4 Å water molecule-sized probe [25]. To compare the SASA for residues with different sizes, we calculate the relative SASA (rSASA), which is the ratio of the SASA of the residue in the protein context to that of the residue *outside* the protein context, along with the C_α, C, and O atoms of the previous amino acid in the sequence and the N, H, and C_α atoms of the next amino acid in the sequence. We define core residues as those with rSASA $\leqslant 10^{-3}$, and we define a protein core as a set of core residues that each share at least one Voronoi cell face (defined below) with each other. We find similar results if the threshold for defining a core residue is smaller, although there will be fewer "core" residues. We showed in previous work that the local packing fraction decreases significantly for residues with rSASA >0.05 [25]. See Fig. 2(a) for an example core region in a protein from the Dunbrack 1.0 database, and Fig. 9 for the size distribution of protein cores in the Dunbrack 1.0 database.

We will compare the structural properties of the cores of protein crystal structures and jammed packings [26] of amino-acid-shaped particles. In previous studies, we found that the packing fraction of core regions in proteins is $\phi \approx 0.56$, which is the same as that of jammed packings of purely repulsive amino-acid-shaped particles *without* backbone constraints [10,27]. Here, we will focus exclusively on packing the hydrophobic residues: Ala, Leu, Ile, Met, Phe, and Val. The amino-acid-shaped particles will include the backbone atoms N, C_α, C, and O, as well as all of the side chain atoms, with the atomic radii given in Ref. [10], which recapitulate the side chain dihedral angles of residues in protein cores. The packings of amino-acid-shaped particles contain mixtures of Ala, Leu, Ile, Met, Phe, and Val residues, with each residue treated as a purely repulsive, rigid body composed of a union of spherical atoms with fixed bond lengths, bond angles, and side-chain and backbone dihedral angles taken from instances in the Dunbrack 1.0 database.

We choose which residues are included in each packing using two methods. For method 1 (M1), we generate $\mathcal{C} = 20$ jammed packings of the exact residues found in each distinct protein core in the Dunbrack 1.0 database. For example, if protein X has a core with R residues, we produce $\mathcal{C}$ jammed packings of those exact R residues. If r of these R residues are not one of the hydrophobic residues we consider, these residues are removed and a jammed packing is generated with the remaining $R - r$ residues. This method seeks to mimic the core size and amino acid frequency distribution found in the Dunbrack 1.0 database. Note that we do not remove r residues from the corresponding protein core in the Dunbrack 1.0 database; nonhydrophobic residues are only excluded from the initial conditions used to generate packings of amino-acid-shaped particles. In method 2 (M2), we randomly select hydrophobic residues from the Dunbrack 1.0 database with frequencies set by the fraction of each amino acid type found in the Dunbrack 1.0 database. The frequencies are 0.29 (Ala), 0.19 (Leu), 0.17 (Ile), 0.05 (Met), 0.07 (Phe), and 0.23 (Val). In method 2, the identities of the residues in the jammed packings only match those in protein cores on average.

We now briefly describe the computational method for generating jammed packings of amino-acid-shaped particles. We use a pairwise, purely repulsive linear spring potential to model inter-residue interactions. Because the residues are rigid particles with each composed of a union of spheres, we test for overlaps between residues μ and ν by checking for

overlaps between all atoms i on residue μ and all atoms j on residue ν, respectively. Note that this potential is isotropic and depends only on the distances between atoms on different residues. [See Eq. (A1) in Appendix A.]

We place N residues with random initial positions and orientations at packing fraction $\phi_0 = 0.40$ in a cubic simulation box with periodic boundary conditions and then increase the packing fraction in small steps $\Delta\phi$ to isotropically compress the system. After each compression step, we relax the total potential energy using FIRE energy minimization [28]. This method is similar to a "fast" thermal quench that finds the nearest local potential energy minimum. We use quaternions to track the particle orientations for each residue, as described in Ref. [29]. If the total potential energy per residue is zero after energy minimization, i.e., $U/N\epsilon < 10^{-8}$, where ϵ is the energy scale of the atomic interactions, we continue to increase the packing fraction. If the total potential energy per residue is nonzero, i.e., $U/N\epsilon \geqslant 10^{-8}$ and residues have small overlaps, we decrease the packing fraction. The packing fraction increment $\Delta\phi$ is halved each time the algorithm switches from compression to decompression and vice versa. We terminate the packing-generation protocol when the residue packings satisfy $10^{-8} < U/N\epsilon < 2 \times 10^{-8}$ and possess a vanishing kinetic energy per residue (i.e., $K/N\epsilon < 10^{-20}$) [13]. [An example jammed packing of amino-acid-shaped particles is shown in Fig. 2(b) and further computational details are included in Appendix A.]

To measure the distribution of local voids in packings of amino-acid-shaped particles and protein cores, we use a Voronoi tessellation, which ascribes to each particle the region of space that is closer to that particle than all other particles in the system. For residues, which are highly nonspherical particles, we use a generalization of the standard Voronoi tessellation known as the *surface-* or *set*-Voronoi (SV) tessellation [30]. This tessellation partitions the empty space in the system using a bounding surface for each residue. An efficient algorithm to generate this tessellation is outlined in Ref. [30] and implemented using POMELO [31]. To construct the SV tessellation, consider a set of N particles with bounding surfaces $\{\partial K_\mu\}$ for $\mu = 1, ..., N$. The software approximates ∂K_μ by triangulating points on the particle surfaces, and uses the standard Voronoi tessellation of the surface points to construct the SV cell for each residue μ. We find that using 400 surface points per atom, or ≈6400 surface points per residue, gives an accurate representation of the SV cell, which does not change significantly as more surface points are added. An example SV cell from a packing of amino-acid-shaped particles is shown in Fig. 1(a). For an SV cell with volume V_μ^v surrounding residue μ with volume v_μ, the local porosity is given by

$$\eta_\mu = \frac{V_\mu^v - v_\mu}{V_\mu^v} = 1 - \phi_\mu, \tag{1}$$

where $\phi_\mu = v_\mu/V_\mu^v$ is the local packing fraction. This quantity measures the local *void* space associated with each residue.

We also quantify the "connected" void space shared between residues in packings of amino-acid-shaped particles and protein cores. To do this, we implement a grid-based method similar to that described by Kertèsz [19] and Cuff and Martin [20], where the "void space" is defined as the region of a system accessible to a spherical probe particle with radius a. The geometry and distribution of void space in a system is thus a function of a, the residue positions $\vec{r}_\mu$, and bounding surfaces ∂K_μ. We define a cubic lattice with G points in each direction within the simulation domain, which gives a lattice spacing $g = L/G$. For all lattice points $\mathbf{p}$, we define the set of void points $\mathcal{V}$ to be all points that can accommodate a spherical probe particle with radius a without causing overlaps with any atoms. We label all void points with a 1, and all other points with a 0. After all grid points are labeled, we use the Newman-Ziff algorithm [32] to cluster adjacent, similarly-labeled grid points. We consider all adjacent points on the nearest face, edge, and vertex of a cube

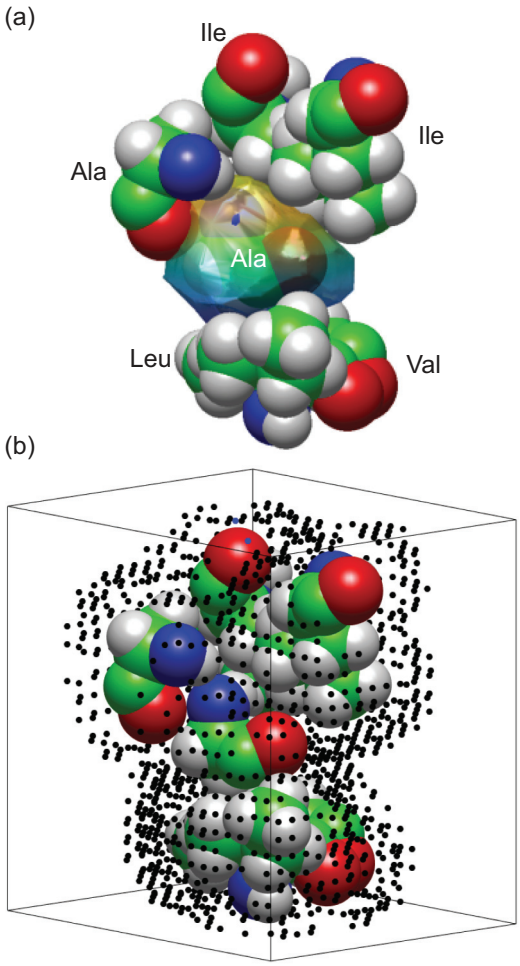

FIG. 1. Visualization of (a) local and (b) connected voids from the same computationally generated packing of $N = 64$ amino-acid-shaped particles. Only the central Alanine (Ala) with the neighboring Alanine, 2 Isoleucines (Ile), Leucine (Leu), and Valine (Val) are shown for clarity. The neighboring amino acids share at least one common surface Voronoi cell face with the central Ala. In (a), the central Ala is enclosed by its surface Voronoi cell. In (b), the connected void space is visualized using points on a grid. For clarity only 75% of the points are shown, and the grid spacing ($g = 0.7$Å) is large compared to values used in the text. In both (a) and (b), the atoms are colored as follows: C (green), O (red), N (blue), and H (white). See Fig. 8 for visualizations of the connected void space throughout the entire simulation domain.

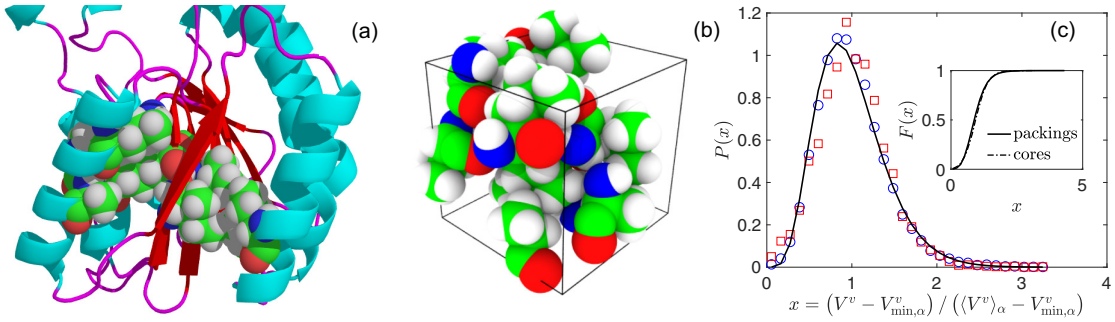

FIG. 2. (a) Core residues in an example globular protein (PDB code: 3F1L). Noncore regions are drawn using the ribbon representation, and the 11 core amino acids are drawn in all-atom representation. (b) Jammed packing of the same 11 core residues in (a). (c) The surface Voronoi cell volume V^v distribution plotted as a function of $x = (V^v - V^v_{\min,\alpha})/(\langle V^v \rangle_\alpha - V^v_{\min,\alpha})$ and fit to a k-gamma distribution (black line) with $k = 6.06 \pm 0.08$ and 5.29 ± 0.27 for packings of amino-acid-shaped particles (circles) and protein cores (squares), respectively. $\langle V^v \rangle_\alpha$ is the average and $V^v_{\min,\alpha}$ is the minimum SV cell volume of residue type α. The inset of (c) is the cumulative distribution function $F(x)$ for the data in the main panel.

of points surrounding each lattice point (i.e., next-to-next-to-nearest neighbor counting with 26 possible adjacencies for each point) when merging void clusters and implement periodic boundary conditions. A sketch of connected void lattice points in a subset of a packing of amino-acid-shaped particles is shown in Fig. 1(b).

When measuring void space in protein structures, we implement a similar procedure, but we only consider voids in core residues. We construct a box of dimension $L_x \times L_y \times L_z$ that circumscribes each protein core, with the box just outside the radii of core residues near the box edges. We pick a spherical probe particle of radius a, and label the void space as all points that are (a) not contained inside an atom, and (b) contained only within the union of the SV cells of core residues. With these constraints, we only consider connected void space specific to the core of the protein. We then use the Newman-Ziff algorithm to merge void clusters, and repeat the procedure for 100 different random protein orientations.

III. RESULTS

A. Local void analysis

We begin with an analysis of local voids associated with each amino acid in jammed packings of amino-acid-shaped particles and protein cores. We measure the distribution of the SV cell volumes and show that the distributions in both systems can be fit to a k-gamma distribution, which also describes Voronoi cell distributions in jammed packings of spheres [33,34], ellipsoids [35], attractive emulsion droplets [36], wet granular materials [37], and model cell monolayers [38]. The k-gamma distribution for the SV cell volume V^v_μ for each residue has the form

$$P(x) = \frac{k^k}{\Gamma(k)} x^{k-1} \exp(-kx), \tag{2}$$

where $x = (V^v_\mu - V^v_{\min,\alpha})/(\langle V^v_\mu \rangle_\alpha - V^v_{\min,\alpha})$, which sets the scale factor of the distribution to 1. Here,

$$\langle V^v_\mu \rangle_\alpha = \frac{1}{N_\alpha} \sum_{\mu=1}^{N_\alpha} V^v_\mu \tag{3}$$

is the average SV cell volume of residue type α. The sum involving μ is over all N_α residues of type α in all packings, and $V^v_{\min,\alpha}$ is the minimum SV cell volume of residue type α. We consider minima and averages for each residue type separately to account for the large differences in residue volumes; that is, each residue type α, when considered individually, has a SV cell volume distribution described by Eq. (2).

We measure the shape factor k_α for each residue type α either by fitting the SV cell volume distribution to Eq. (2) using Maximum Likelihood Estimation (MLE), or by calculating

$$k_\alpha = \frac{\left(\langle V^v_\mu \rangle_\alpha - V^v_{\min,\alpha}\right)^2}{\langle \left(V^v_\mu\right)^2 \rangle_\alpha - \langle V^v_\mu \rangle^2_\alpha}. \tag{4}$$

We obtain similar k values using both methods. Although the values of k_α depend on the type of amino acid α, when we average the values of k_α we recover the value of k obtained from fitting the combined distribution. We focus on the distributions of SV cell volumes averaged over all hydrophobic residues.

In Fig. 2(c), we show the SV cell volume distributions $P(x)$ for packings of core amino-acid-shaped particles modeled after specific protein cores (method M1) and for all core residues in the Dunbrack 1.0 database. We find that the distributions for these two systems are similar; both obey a k-gamma distribution [Eq. (2)] with similar shape parameters, $k = 6.06 \pm 0.08$ and $k = 5.29 \pm 0.27$, for core residues in the Dunbrack 1.0 database and packings of amino-acid-shaped particles, respectively. As expected, the cumulative distributions $F(x)$ of the SV cell volumes for residues in protein cores and packings of amino-acid-shaped particles are also nearly indistinguishable.

The strong similarity between the SV cell volume distributions indicates that jammed packings of amino-acid-shaped particles (at $\phi_J \approx 0.56$) and protein cores possess the same underlying structure. To better understand this result, in Fig. 3 we plot the shape parameter k that describes the form of the Voronoi cell volume distributions for packings of $N = 10^3$ monodisperse spheres (with $\phi_J \approx 0.64$) and of $N = 64$ amino-acid-shaped particles versus ϕ. When $\phi \ll \phi_J$, and the systems are sufficiently dilute, the Voronoi cell

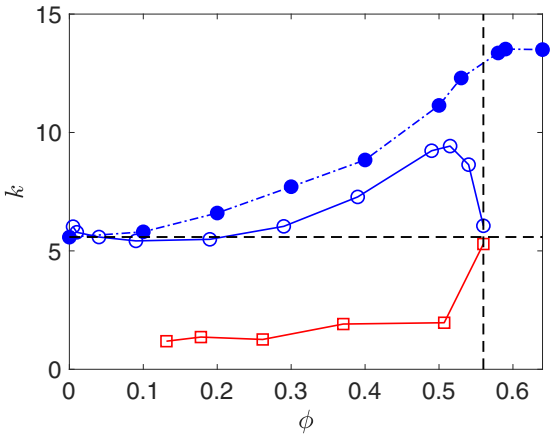

FIG. 3. The shape parameter k for fits of the k-gamma distribution [Eq. (2)] to the SV cell volume distributions $P(x)$ for packings of amino-acid-shaped particles (open circles), monodisperse spheres (filled circles), and both core and surface residues in the Dunbrack 1.0 database (open squares) as a function of packing fraction ϕ. The dashed horizontal line at $k = 5.59$ is the analytical value of the shape factor for the Voronoi cell volume distribution of a random Poisson point process [39], and the dashed vertical line at $\phi_J = 0.56$ is the packing fraction for protein cores and jammed packings of amino-acid-shaped particles.

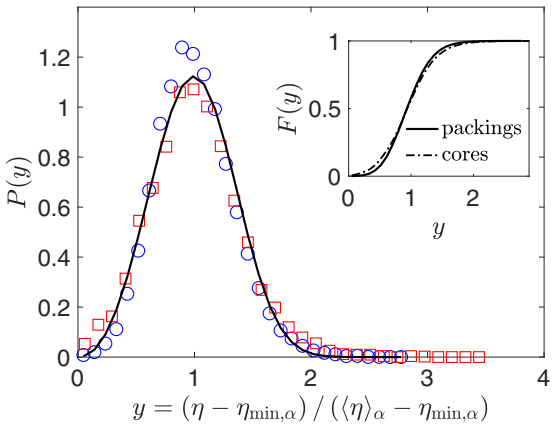

FIG. 4. Distribution of the scaled local porosity $y = (\eta - \eta_{\min,\alpha})/(\langle\eta\rangle_\alpha - \eta_{\min,\alpha})$, where $\langle\eta\rangle_\alpha$ is the average and $\eta_{\min,\alpha}$ is the minimum porosity of residue type α, for packings of amino-acid-shaped particles (circles) and residues in protein cores in the Dunbrack 1.0 database (squares). The solid line is a Weibull distribution with shape parameter $b \approx 3.2$ [Eq. (7)]. The inset is the cumulative distribution function $F(y)$ of the data in the main panel.

volume distributions of the packings of monodisperse spheres and amino-acid-shaped particles resemble that for a random Poisson point process [39] with $k \approx 5.6$. In this regime, free volume is assigned randomly to each particle since the particle positions are uncorrelated. However, as ϕ increases, the k values for packings of monodisperse spheres and amino-acid-shaped particles begin to grow, but at different rates, since the particle geometry becomes important in determining the local free volume. Near $\phi \simeq \phi_J$, the shape parameter plateaus at $k \approx 13$ for packings of monodisperse spheres, but the shape parameter decreases strongly to $k \approx 6$ for packings of amino-acid-shaped particles. This decrease in k indicates a transition from having the shape of the Voronoi cell volume distribution determined by spherical particles (for $\phi \lesssim \phi_J$) to that determined by bumpy, asymmetric amino-acid-shaped particles (for $\phi \simeq \phi_J$). Note, however, that the SV cell volume distribution of *jammed* packings of amino-acid-shaped particles is similar (in terms of k value) to that of *randomly* placed Poisson points. This suggests that the void distribution of jammed packings of amino-acid-shaped particles and protein cores share structural properties with randomly placed points. We will expand on this similarity in Sec. III C.

In addition, we calculate k for the SV cell volume distributions for residues in the Dunbrack 1.0 database as a function of packing fraction. In previous studies, we have found a one-to-one correlation between solvent accessibility and packing fraction [25]; residues with lower values of ϕ in Fig. 3 are therefore more solvent-exposed, i.e., closer to the protein surface. For most of the range in ϕ, $k \approx 2$, whereas $k \gtrsim 5.6$ for packings of monodisperse spheres and amino-acid-shaped particles. In particular, k does not equal the value for a random Poisson point process ($k = 5.6$) in the limit $\phi \ll \phi_J$ for residues in protein cores. In protein cores, the backbone constraint gives rise to correlations in the residue positions. However, as $\phi \to \phi_J$, k increases, reaching $k \approx 6$ when $\phi = \phi_J$. This result shows that there is a fundamental change in the SV cell distribution near the onset of jamming in protein cores. For $\phi \lesssim \phi_J$, the backbone determines the shape of the SV cell volume distribution, whereas for $\phi \to \phi_J$, the shapes of the amino acids determine the SV cell volume distribution.

We also compare the local porosity distributions for protein cores and packings of amino-acid-shaped particles in Fig. 4. We scale the porosity [as in Eq. (2)] by defining

$$y = \frac{\eta_\mu - \eta_{\min,\alpha}}{\langle\eta_\mu\rangle_\alpha - \eta_{\min,\alpha}}, \tag{5}$$

where

$$\langle\eta_\mu\rangle_\alpha = \frac{1}{N_\alpha}\sum_{\mu=1}^{N_\alpha}\eta_\mu, \tag{6}$$

and $\eta_{\min,\alpha}$ is the minimum porosity over all N_α core residues of type α. Again, the porosity distributions $P(y)$ [and cumulative distributions $F(y)$] for residues in protein cores and packings of amino-acid-shaped particles are similar, but here $P(y)$ has the shape of a Weibull distribution with scale factor $\lambda = 1$,

$$P(y) = by^{b-1}\exp(-y^b), \tag{7}$$

where b is the shape parameter of the Weibull distribution.

The small differences in $P(x)$ and $P(y)$ between core residues in protein crystal structures and packings of amino-acid-shaped particles can be explained by the small differences between the volumes of core residues in crystal structures and in packings. The atoms on neighboring amino acids interact differently for free amino acids in packings versus backbone atoms in protein cores, which form covalent and hydrogen bonds. Thus, we find that the volumes of residues in protein cores have larger variances and smaller means than

those in packings of amino-acid-shaped particles. Also, the overlaps between covalently bonded backbone atoms that link adjacent residues slightly decreases the mean SV cell volume, which gives rise to a larger population of small SV cells and a small deviation between $P(x)$ for residues in protein cores and in packings for small x in Fig. 2(c).

B. Connected void analysis of protein cores

We next quantify the distribution of "connected" void space that is shared between residues. Using a grid-based method, we calculate the volume of regions of connected void space as a function of the radius a of a spherical probe particle. As we increase a, the connected void space transitions from highly connected throughout the system to compact and localized with distinct void regions. We measure the probability $\rho(a)$ of finding a percolating void region, where we define percolation as the appearance of a cluster that spans one of the system dimensions when the boundary is closed, and a cluster that both spans, wraps around the boundary, and self-intersects when the boundaries are periodic. We identify the critical probe radius a_c by setting $\rho(a_c) = 0.5$. Because the definition of connected void regions depends on the boundary condition, the value of a_c, especially in systems as small as protein cores, is affected by the boundary conditions. Thus, to calculate $\rho(a)$, we create packings of amino-acid-shaped particles with similar boundary conditions as those in protein cores. From a packing of amino-acid-shaped particles with periodic boundary conditions ($N = 64$, method M2), we extract a representative protein core of $R - r$ residues that all share at least one SV cell face. We sample $R - r$ from the distribution of core sizes $P(R)$ found in the Dunbrack 1.0 database. (See Fig. 9 in Appendix A.) The resulting packings have boundary conditions similar to protein cores in the Dunbrack 1.0 database. We then determine the connected void regions as a function of a and identify the critical probe size a_c as shown in Fig. 5(a). We find the same critical probe size $a_c = 0.48 \pm 0.01$ Å for both protein cores and packings of amino-acid-shaped particles with similar boundary conditions. Note that this value of the critical probe radius is smaller than that of a water molecule, which is ≈1.4 Å, and thus the voids we consider here are not accessible by aqueous solvents. However, as we discuss below, this value of the probe radius corresponds to a critical point; we will exploit the behavior of the voids near this critical point to understand the geometric properties of the connected voids, and to differentiate between the voids in various systems.

Thus, determining the connected void regions in protein cores is a type of percolation problem. In lattice site percolation, sites on a lattice in d spatial dimensions are either occupied randomly with probability p or not occupied with probability $1 - p$. At the percolation threshold p_c, adjacent occupied sites form a percolating cluster that spans the system and becomes infinite in the large-system limit. Continuum percolation occurs in systems that are not confined to a lattice. Both particle contact and void percolation have been studied in randomly placed overlapping spheres [19,40,41] and percolation of particle contacts [42,43] has been studied in packings of repulsive [44] and adhesive particles [45].

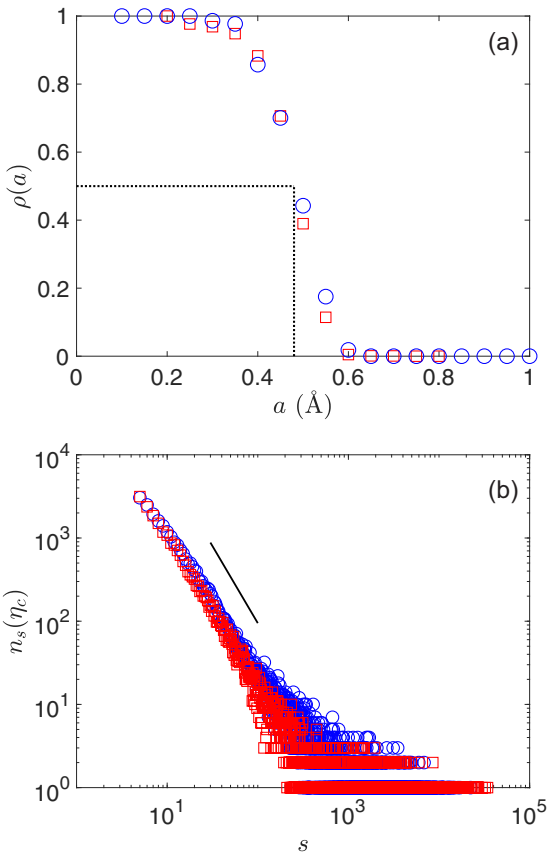

FIG. 5. (a) Percolation probability $\rho(a)$ plotted versus the probe radius a for protein cores from the Dunbrack 1.0 database (crosses) and clusters of core residues extracted from static packings of $N = 64$ amino-acid-shaped particles (circles). The horizontal and vertical dashed lines indicate the critical probe radius $a_c = 0.48$ Å that satisfies $\rho(a_c) = 0.5$. (b) Cluster size distribution n_s with size s at the critical porosity η_c, which scales as $n_s(\eta_c) \sim s^{-\tau}$. The Fisher power-law exponents $\tau = 1.95 \pm 0.06$ and 1.85 ± 0.05 for protein cores from the Dunbrack 1.0 database (crosses) and representative clusters of core residues in packings of amino-acid-shaped particles (circles), respectively. The solid line has slope equal to -1.85.

In this article, we consider percolation of the void space accessible to a spherical probe particle with radius a in packings of spheres and amino-acid-shaped particles, as well as systems composed of randomly placed spheres [40,41]. As the probe particle radius is increased, the amount of space available to the probe is restricted and the number of void lattice sites decreases. We define an effective porosity η as the ratio of the number of void lattice sites to the total number of lattice sites G^d. We determine the percolation threshold using a bisection method, where we begin with two initial guesses for the percolation transition, a_H and a_L with $a_H > a_L$, and iteratively check for percolation of void sites at the probe radius $a = (a_H + a_L)/2$. We set $a_H = a$ if we find a percolated cluster of void sites, and $a_L = a$ if we do not find a percolated cluster. We terminate the algorithm when the difference between successive values for a_c is within a small tolerance $\delta a = 10^{-8}$ Å. Note that our use of a lattice of points to measure the connected void region does not imply that our model is a lattice model. The lattice is simply a tool to

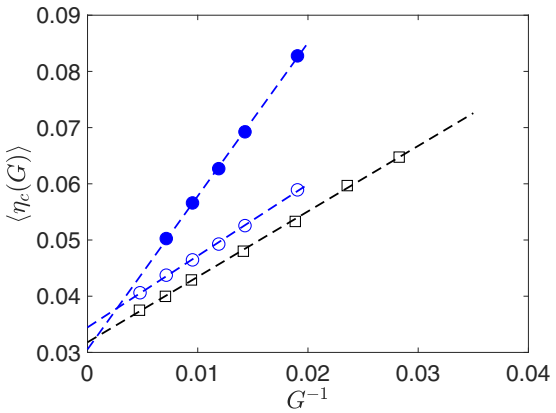

FIG. 6. Critical porosity $\langle \eta_c(G) \rangle$ using a lattice with G points along each dimension plotted versus G^{-1} for jammed packings of $N = 64$ amino-acid-shaped particles with $N_a = 1024$ atoms (open circles), $N = 10^3$ randomly placed spheres (open squares), and $N = 10^3$ monodisperse spheres (filled circles). The dashed lines have vertical intercepts that indicate $\eta_c(G \to \infty) \approx 0.0345$, 0.0318, and 0.0305 for packings of amino-acid-shaped particles, randomly placed spheres, and monodisperse sphere packings, respectively.

calculate the connected void space volume [19]. Furthermore, in the continuum limit (i.e., $G \to \infty$), we recover the critical porosity $\eta_c \approx 0.03$ measured using Kerstein's method [40,41] on systems of randomly placed spheres [46]. (See Fig. 6.) Since there is a one-to-one mapping between a and η, we will use η as the order parameter for continuum void percolation.

C. Connected void analysis of packings

We now focus on the statistical properties of the connected void regions in packings of spheres and amino-acid-shaped particles prepared in systems with cubic, periodic boundaries. In Fig. 7 and Table I, we summarize the results of this analysis, and we show visualizations of the percolating connected voids in jammed packings of spheres and amino-acid-shaped particles in Fig. 8. We first measure the correlation length exponent ν, where the correlation length ξ is defined as the average distance between two points in the largest connected void cluster. Near η_c, ξ diverges as $|\eta - \eta_c|^{-\nu}$. Using finite-size scaling [47], we can write

$$\eta_c(N) - \eta_c(\infty) \sim N^{-1/d\nu}, \tag{8}$$

where $\eta_c(\infty)$ is the percolation threshold in the large-system limit and $N \sim L^d$. $\eta_c(N)$ is a random variable with standard deviation $\Delta\eta_c(N)$, which will approach $\eta_c(\infty)$ as $N \to \infty$. Thus, we make the ansatz that

$$\Delta\eta_c(N) \sim N^{-1/d\nu}, \tag{9}$$

which can be used to measure ν. [See Fig. 7(a).] We also measure the Fisher exponent τ, defined by

$$n_s(\eta_c) \sim s^{-\tau}, \tag{10}$$

where n_s is the number of void clusters containing s sites. While we measure this exponent for protein cores and random packings with representative boundary conditions in Fig. 5(b), in Fig. 7(b) we measure this exponent in systems with cubic, periodic boundary conditions.

We also measure the fractal dimension and percolation strength of the percolating void clusters. The fractal dimension is defined by

$$s_{\max}(\eta_c, N) \sim N^{D/d}, \tag{11}$$

where $s_{\max}(\eta_c, N)$ is the number of sites contained in the largest void cluster in the system at percolation onset. If $D = d$, the largest void cluster is a compact, nonfractal object. However, if $D < d$, the void cluster is fractal [48]. [See Fig. 7(c).] The percolation strength is the probability $\mathcal{P}(\eta)$ that a given lattice site is part of the percolating void cluster at a given porosity. Near η_c, the probability scales as $\mathcal{P}(\eta) \sim |\eta - \eta_c|^{\beta}$. The probability obeys finite size scaling,

$$\mathcal{P}(\eta_c, N) \sim N^{-\beta/d\nu}. \tag{12}$$

Once we determine ν using Eq. (9), we can determine β from Eq. (12). [See Fig. 7(d).] We also expect β, ν, and D to satisfy the hyperscaling relation,

$$D = d - \frac{\beta}{\nu}. \tag{13}$$

In Table I, we report our measurements for the critical exponents ν, τ, D, and β for void percolation (using a spherical probe particle), as well as for $d = 3$ lattice site percolation on a cubic lattice and void percolation in systems of randomly placed spheres using two methods: the connected void method described previously and the Voronoi vertex method introduced by Kerstein [40] and implemented by Rintoul [41]. Note that protein cores and representative subsets of jammed packings of amino-acid-shaped particles (denoted "rep.") are small systems with $N < 30$, and thus we cannot use finite-size scaling to measure the critical exponents. We can, however, measure the critical exponents for full packings of amino-acid-shaped particles (denoted "full"), which mimic the geometric properties of void clusters in protein cores.

We observe that across all models and methods studied, the correlation length exponent $\nu \approx 0.9$–1.0 for void percolation. In particular, $\nu \approx 0.90$ for packings of amino-acid-shaped particles is similar to that (0.90) for randomly placed spheres [41], as well as for standard site percolation [47]. In addition, the fractal dimension $D \approx 2.4$–2.6 is similar for all models and methods for calculating void percolation. We find that the percolation strength exponent $\beta < 0.5$ for randomly placed spheres and packings of amino-acid-shaped particles when using the connected void method, but $\beta > 0.5$ for packings of monodisperse and bidisperse spheres. (The bidisperse systems include $N/2$ large and $N/2$ small spheres with diameter ratio $d = 1.4$.)

However, because of the limited range of system sizes studied here, it is difficult to determine the critical exponents with high precision. Thus, given the results for the ν, D, and β exponents alone, it is difficult to distinguish the statistical properties of the void content of packings of jammed spheres, randomly placed spheres, and jammed amino-acid-shaped particles from each other, or from void percolation on a cubic lattice, for that matter.

We do see a strong distinction in the Fisher exponent τ [Eq. (10)] between void percolation on a cubic lattice and in packings of amino-acid-shaped particles. For these two systems, $\tau = 2.07 \pm 0.01$ and $\tau = 1.29 \pm 0.06$, respectively. For

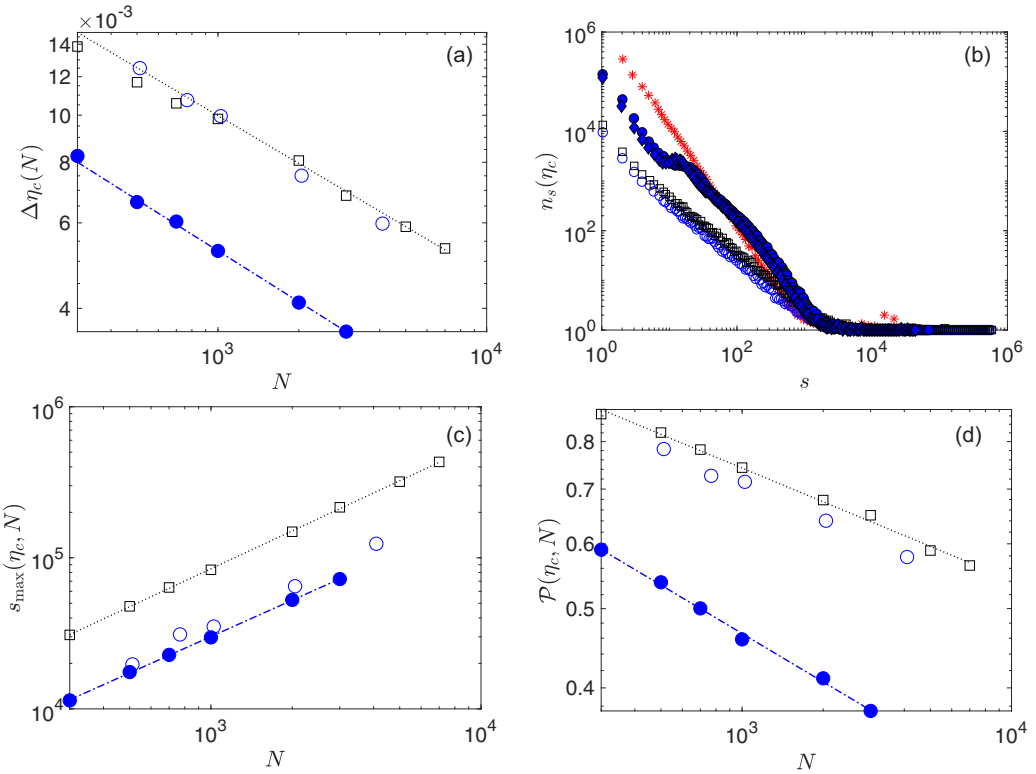

FIG. 7. (a–d) Scaling behavior for jammed packings of amino-acid-shaped particles (open circles), bidisperse spheres (filled circles), and randomly placed spheres (open squares). In (a), we show that the standard deviation in the critical porosity scales as $\Delta\eta_c(N) \sim N^{-1/d\nu}$, where ν is the correlation length exponent. The lines have slopes -0.33 (dotted line) and -0.31 (dot-dashed line). In (b), we show that for randomly placed spheres and jammed packings of amino acids, the distribution of connected void clusters of size s at the percolation threshold scales as $n_s(\eta_c) \sim s^{-\tau}$, with a Fisher exponent $\tau \approx 1.25$. In addition, we plot n_s at criticality for standard site percolation on a cubic lattice (red stars), which has an apparent $\tau \approx 2.07$. We also plot the same distribution for the connected voids in jammed packings of bidisperse and monodisperse spheres (filled diamonds), which do *not* display power-law scaling. In (c), we show that the maximum cluster size near the percolation onset scales as $s_{\max}(\eta_c, N) \sim N^{D/d}$, were D is the fractal dimension. The lines have slopes 0.83 (dotted line) and 0.82 (dot-dashed line). In (d), we show that the probability for a given site to be in the percolating void cluster at η_c scales as $\mathcal{P}(\eta_c, N) \sim N^{-\beta/d\nu}$, where β is the percolation strength scaling exponent. The lines have slopes -0.14 (dotted line) and -0.19 (dot-dashed line).

packings of randomly placed spheres, we find that the Fisher exponent is $\tau = 1.22 \pm 0.05$, which is similar to our result for packings of amino-acid-shaped particles. (See Table I.) These values for τ were obtained from a cubic lattice with $G = 100$ sites per box length, packings of amino-acid-shaped packings with $N = 128$ particles and, on average, $N_a = 2048$ atoms, and systems of $N = 2000$ randomly placed spheres. These results suggest that the properties of connected voids are similar in packings of amino-acid-shaped particles and randomly placed spheres, and in general that connected voids in these systems are distinct from those for void regions in cubic lattices near percolation onset.

We do not report values of τ for jammed packings of monodisperse and bidisperse spheres, since we observe non-power-law behavior in the cluster size distributions for these systems. As discussed in Appendix C, this behavior is most likely due to a residual finite length scale at the percolation threshold. We also observe non-power-law behavior in the cluster size distribution for void percolation in randomly placed spheres using Kerstein's method, and do not report a value for τ in Table I. As shown in Appendix C, this non-power-law behavior is most likely due to the sparsity of the Voronoi vertex network, which truncates the cluster size distribution.

Our results suggest that the critical exponent τ is able to distinguish the geometries of connected void regions in different systems. In particular, the connected void regions in packings of amino-acid-shaped particles and systems of randomly placed spheres are similar, but distinct from that for jammed sphere packings. In Fig. 8, we show examples of the connected void surface in packings of (a) amino-acid-shaped particles, (b) randomly placed spheres, and (c) bidisperse spheres. Qualitatively, the connected void surfaces in systems of randomly placed spheres and amino-acid-shaped particles look similar, while the connected void surface in jammed packings of bidisperse spheres looks different, with a characteristic void size.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

In this article, we analyzed local and connected void regions in protein cores and in jammed packings of purely repulsive amino-acid-shaped particles and showed that these

TABLE I. Table of critical exponents ν, τ, D, and β for several models of void percolation. In the last column, we provide the value for the hyperscaling relation, $d - \frac{\beta}{\nu}$, which matches the fractal dimension D if hyperscaling is satisfied. In the first four rows, we report the critical exponents for packings of amino-acid-shaped particles with periodic boundary conditions (full) and boundary conditions representative of protein cores (rep.). We also report the critical exponents for void percolation in jammed packings of monondisperse (Mono.) and bidisperse (Bidis.) spheres. In the last four rows, we compare these results to those for void percolation in several systems that were studied previously. We report our measurements of the critical exponents for site percolation on a cubic lattice, where only nearest neighbors are counted as adjacent sites. We also report the critical exponents for void percolation and Voronoi vertex percolation in systems composed of randomly placed spheres. Previously reported values of the exponents are given in parentheses, and references are given in the footnotes. Throughout the table, error bars are obtained from the bootstrap method, where we randomly, independently sample 20% of the data over 200 trials, and fit the exponents over each individual trial. Values for the exponents are the average over trials, and the error bars are standard deviations.

System	ν	τ	D	β	$d - \frac{\beta}{\nu}$
residue packings (full)	0.90 ± 0.24	1.29 ± 0.06	2.58 ± 0.18	0.37 ± 0.19	2.58 ± 0.33
residue packings (rep.)	–	1.85 ± 0.05	–	–	–
Protein cores, Dunbrack 1.0	–	1.95 ± 0.06	–	–	–
Mono. Spheres (jammed)	1.05 ± 0.12	–	2.46 ± 0.09	0.60 ± 0.07	2.43 ± 0.15
Bidis. Spheres (jammed)	0.93 ± 0.10	–	2.40 ± 0.08	0.56 ± 0.07	2.40 ± 0.14
Cubic Lattice[a]	0.91 ± 0.04 (0.88)	2.07 ± 0.01 (2.18)	2.49 ± 0.03 (2.53)	0.48 ± 0.02 (0.42)	2.47 ± 0.04
Randomly Placed Spheres (connected void method)	1.10 ± 0.06	1.22 ± 0.05	2.51 ± 0.03	0.44 ± 0.03	2.60 ± 0.05
Randomly Placed Spheres (Voronoi vertex method)	0.99 ± 0.05 (0.902 ± 0.005[b])	–	2.44 ± 0.03	0.48 ± 0.02 (0.45 ± 0.2[c])	2.52 ± 0.05

[a]Ref. [47].
[b]Ref. [41].
[c]Ref. [40].

two systems share the same void structure. We first investigated the surface-Voronoi (SV) cell volume distributions and found that in both systems these distributions are well-described by a k-gamma distribution with $k \approx 6$. This k value is much smaller than that ($k \approx 13$) obtained for jammed sphere packings, which indicates that packings of amino-acid-shaped particles have a broader distribution of Voronoi volumes. We also studied the SV cell volume distribution as a function of the packing fraction, and found that only near the onset of jamming do the SV cell distributions in protein cores and packings of amino-acid-shaped particles match. In the dilute case $\phi \ll \phi_J$, the local packing environment in protein cores is determined by the backbone, whereas the local packing environment of packings of free residues resembles a Poisson point process. At jamming onset, the local packing environment is determined by the "bumpy," asymmetric shape of amino acids, not the backbone constraints.

Using a grid-based method, we also measured the distribution of nonlocal, connected voids in protein cores and jammed packings of amino-acid-shaped particles. We found that when we consider similar boundary conditions in protein cores and jammed packings of amino-acid-shaped particles, the two systems also have the same critical probe size a_c (at which the accessible, connected void region spans the system) and Fisher exponent τ (which characterizes the scaling of the size of the void clusters near percolation onset). We also compared the finite-size scaling results for void percolation in packings of amino-acid-shaped particles, in packings of monodisperse and bidisperse spheres, and systems of randomly placed spheres. We found that the void percolation

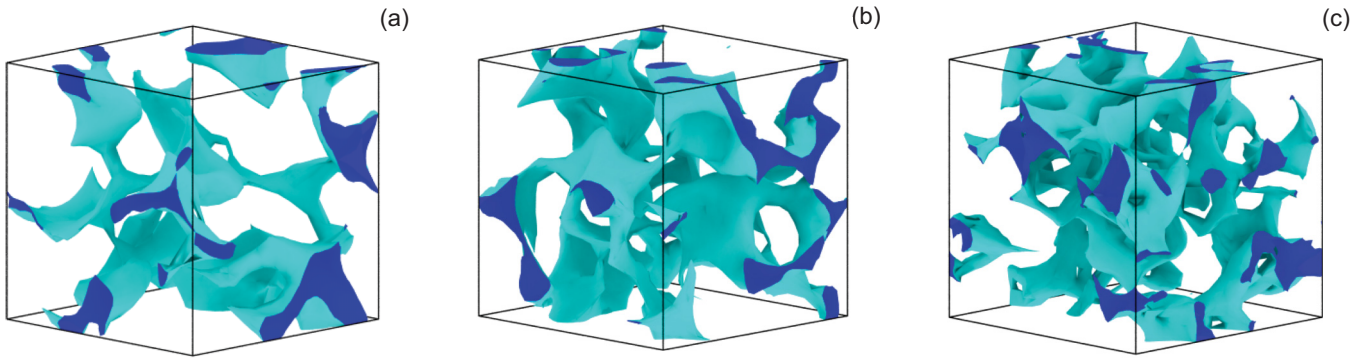

FIG. 8. Visualization of the surface of a connected void region (light domain) at the percolation threshold $\eta_c \approx 0.03$. The dark domains are the "inside" of the void region, which connects across the periodic boundaries. These three systems are (a) a jammed packing of $N = 16$ amino-acid-shaped particles, with 298 atoms in total, (b) $N = 300$ randomly placed spheres, and (c) a jammed packing of 300 bidisperse spheres.

critical exponents in packings of amino-acid-shaped particles are similar to those in randomly placed spheres. Specifically, the Fisher exponent τ takes a similar value for these two systems ($\approx$1.2 and $\approx$1.3, respectively), while this exponent is significantly different for void percolation in cubic lattices ($\approx$2.1). This result may explain why the distribution of SV cell volumes is similar for jammed packings of amino-acid-shaped particles and randomly placed Poisson points with $\phi = 0$, as seen in the horizontal line Fig. 3. Interestingly, these results echo similar observations by Liang and Dill, where the authors recognize the similarity between the void distribution of randomly-placed spheres and of protein crystal structures, although they did not connect packing in protein cores with random close packing of nonspherical particles [2].

In future work, we will use jammed packings of amino-acid-shaped particles to understand the structural and mechanical response of protein cores to amino acid mutations. We can assess the response in two ways. First, we can prepare jammed packings of amino-acid-shaped particles that represent wildtype protein cores, substitute one or more of the wildtype residues with other hydrophobic residues, relax the "mutant" packing using potential energy minimization, and measure the changes in void structure. We can also measure the vibrational density of states (VDOS) in jammed packings that represent the wildtype and mutant cores. The VDOS and the associated eigenmodes can provide detailed information on how the low-energy collective motions change in response to mutations. There are several advantages for calculating the VDOS in jammed packings of amino-acid-shaped particles. For example, in jammed packings, only hard-sphere-like steric interactions are included. In contrast, molecular dynamics force fields for proteins typically include many terms in addition to those that enforce protein stereochemistry, which makes it difficult to determine the interactions that control the collective motions. Studying jammed packings of amino-acid-shaped particles also decouples the motions of core versus surface residues.

Studies of the VDOS in jammed packings of amino-acid-shaped particles will also shed light on the protein "glass" transition, where the root-mean-square deviations in the atomic positions switch from harmonic to anharmonic behavior [17] in globular proteins near $T_g \approx 200$ K [18]. We will investigate the vibrational response of jammed packings of amino-acid-shaped particles to thermal fluctuations. In particular, we will measure the Fourier transform of the position fluctuations and determine the onset of anharmonic response. Additionally, in this work we did not include backbone connectivity between amino acids in our packings, nor did we treat the side chain dihedral angles as "soft" degrees of freedom with harmonic constraints. In future work, to more accurately model the geometrical and topological properties of dynamically fluctuating protein cores, we will incorporate harmonic bond length, bond angle, and dihedral angle interactions (for both backbone and side chain atoms), with stiffnesses taken from bond length, bond angle, and dihedral angle distributions observed in high-resolution protein crystal structures.

In addition, our analysis of void distributions in protein cores will provide new methods for identifying protein decoys, which are computationally generated protein structures that are not observed experimentally. However, it is currently difficult to distinguish between real structures and decoys. For example, in the most recent Critical Assessment of Protein Structure Prediction (CASP12), researchers were given a set of target sequences, and were tasked with predicting the structures of those sequences using a variety of methods [49]. Each group was allowed to submit five structures per target sequence; when assessing which of their submissions were the most accurate, only 3 groups out of 31 had >50% success at identifying the most accurate structure [50]. The average success rate was 30%, just slightly better than guessing at random. Thus, assessing the viability of computationally-designed structures is an incredibly difficult task.

Since the structure of void regions in the cores of protein crystal structures is the same as that found in packings of amino-acid-shaped particles, the properties of void regions can serve as a benchmark for ranking computationally designed protein structures. Recent studies have suggested that protein decoys [8] possess local packing fraction inhomogeneities that are not present in protein crystal structures. In addition, the void-based analyses presented here can be used to evaluate the conformational dynamics of proteins sampled in all-atom molecular dynamics simulations. An understanding of the expected void properties from high-resolution protein crystal structures can improve our ability to identify unphysical conformational fluctuations that occur during molecular dynamics trajectories. We propose that detailed characterizations of the void space, using the methods described here, will be a sensitive metric than can be used to assess a variety of protein designs.

ACKNOWLEDGMENTS

The authors acknowledge support from NIH training Grant No. T32EB019941 (J.D.T.), the Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences (Z.M.), and NSF Grant No. PHY-1522467 (C.S.O.). This work also benefited from the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center. We thank J. C. Gaines for providing the code to analyze cores in protein crystal structures and Z. Levine for helpful comments on this research.

APPENDIX A: PACKING-GENERATION PROTOCOL

As described in Sec. II, we generate jammed packings of amino-acid-shaped particles using successive small steps of isotropic compression or decompression with each step followed by potential energy minimization. Each residue was modeled as a rigid union of spheres with fixed bond lengths, bond angles, and dihedral angles. Forces between atoms i and j on distinct residues μ and ν were calculated using $\vec{F}_{ij}^{\mu\nu} = -\vec{\nabla} U(r_{ij}^{\mu\nu})$, with the pairwise, purely repulsive linear spring potential energy,

$$U\left(r_{ij}^{\mu\nu}\right) = \frac{\epsilon}{2}\left(1 - \frac{r_{ij}^{\mu\nu}}{\sigma_{ij}^{\mu\nu}}\right)^2 \Theta\left(1 - \frac{r_{ij}^{\mu\nu}}{\sigma_{ij}^{\mu\nu}}\right). \tag{A1}$$

In Eq. (A1), ϵ is the characteristic energy scale of the repulsive interactions and $\sigma_{ij}^{\mu\nu} = (\sigma_i^{\mu} + \sigma_j^{\nu})/2$, where σ_i^{μ} is the

diameter of atom i on residue μ. The quantity $r_{ij}^{\mu\nu} = |\vec{r}_j^{\,\mu} - \vec{r}_i^{\,\nu}|$ is the separation distance between atoms i and j on distinct residues μ and ν, and Θ is the Heaviside step function that sets the potential energy to zero when atoms i and j are not in contact. We consider each residue as a rigid body. Forces between pairs of atoms on contacting residues generate torques. The torque on residue μ arising from a force on atom i on residue μ from atom j on residue ν is $\vec{h}_i^{\,\mu} \times \vec{F}_{ij}^{\,\mu\nu}$, where $\vec{h}_i^{\,\mu}$ is the position of atom i relative to the center of mass of residue μ. Note that this pair potential reduces to a hard-sphere-like interaction in the limit of small atomic overlaps [27]. The total potential energy U is given by

$$U = \sum_{\nu<\mu} \sum_{i,j} U\left(r_{ij}^{\mu\nu}\right). \tag{A2}$$

We use the velocity-Verlet algorithm to integrate the translational equations of motion for each particle's center of mass, and a quaternion-based variant of the velocity-Verlet method described in Ref. [29] to integrate the rotational equations of motions for each residue.

To simulate isotropic compression, we scale all lengths in the system (except the box edges) at each iteration m by the scale factor

$$\alpha = \left(\frac{\phi_m + \Delta\phi_m}{\phi_m}\right)^{1/3}, \tag{A3}$$

where ϕ_m is the packing fraction and $\Delta\phi_m$ is the packing fraction increment at iteration m. This process uniformly grows or shrinks all atoms, and thus the packing fraction satisfies $\phi_{m+1} = \phi_m + \Delta\phi_m$. After each compression or decompression step, we use the FIRE algorithm [28] to minimize the potential energy in the packing. The packing fraction increment is halved each time the total poential energy switches from zero (i.e., $U/N\epsilon < 10^{-8}$) to nonzero or vice versa. We terminate the packing-generation algorithm when the total potential energy per residue satisfies $10^{-8} < U/N\epsilon < 2 \times 10^{-8}$ and the kinetic energy per residue is below a small threshold, $K/N\epsilon < 10^{-20}$. We set the initial values of the packing fraction and packing fraction increment to be $\phi_0 = 0.4$ and $\Delta\phi_0 = 10^{-3}$, but our results do not depend sensitively on these values.

APPENDIX B: PROTEIN CORE SIZE DISTRIBUTION

In this Appendix, we show the distributions of the number of core residues in protein crystal structures from the Dunbrack 1.0 database. (See Fig. 9.) As described in Sec. II, we define protein cores as clusters of residues that all share at least one SV cell face with other residues in the cores, and every atom in each residue has an rSASA $\leqslant 10^{-3}$. In Method M1 for generating jammed packings of amino-acid-shaped particles, we create $\mathcal{C}$ replicas of each protein core with the specific $R - r$ residues found in that core, where R is the number of core residues and $R - r$ is the number of Ala, Ile, Leu, Met, Phe, and Val core residues. Before pruning nonhydrophobic residues, the average core size is $\langle R\rangle \approx 16$ residues, and $\langle R - r\rangle \approx 12$ after pruning.

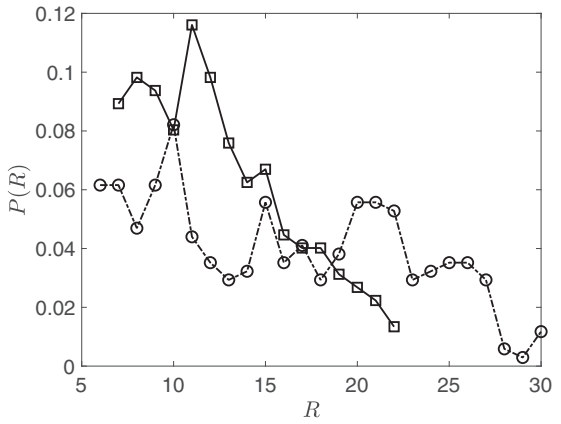

FIG. 9. Distribution of the number of core residues $P(R)$ in the Dunbrack 1.0 database, before (circles) and after (squares) pruning nonhydrophobic residues from the core replicas as described in Sec. III A. The mean number of residues before pruning is $\langle R\rangle \approx 16$, and after pruning is $\langle R - r\rangle \approx 12$.

APPENDIX C: MEASUREMENT OF THE FISHER EXPONENT τ

In this Appendix, we explain the differences we observe in the Fisher exponent τ for different systems. In void percolation in cubic lattices, systems of randomly placed spheres and jammed packings of amino-acid-shaped particles, the distribution of void cluster sizes at percolation onset $n_s(\eta_c)$ has a well defined power-law decay, as shown in Fig. 7(b). We demonstrate this difference further in Fig. 10, where we plot the average Fisher exponent $\langle\tau\rangle$ with error bars as a function of the number of void grid points G measured along the box length. Error bars are estimated by the bootstrap method (described above). We find that the value of τ for randomly placed spheres and jammed packings of amino-acid-shaped particles are similar, while markedly different from τ measured for void percolation on cubic lattices.

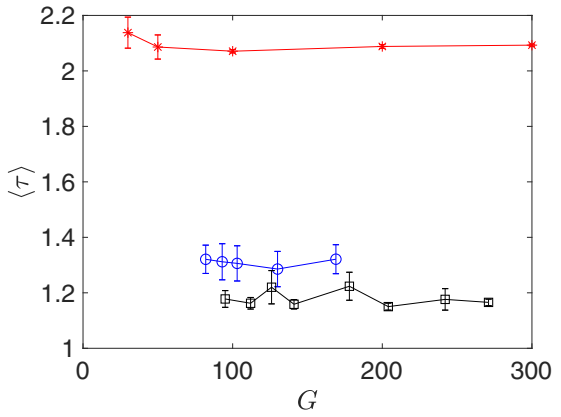

FIG. 10. Plot of the average Fisher exponent $\langle\tau\rangle$ as a function of the number of lattice points along a given direction G for void percolation on a cubic lattice (stars), in packings of jammed amino-acid-shaped particles (open circles), and packings of randomly placed spheres (open squares). Error bars are calculated by the bootstrap method, where τ is fit to 200 separate trials of independent, random subsets of the cluster size data. The average Fisher exponent $\langle\tau\rangle$ is the mean of these fits, and the error bars are the standard deviations.

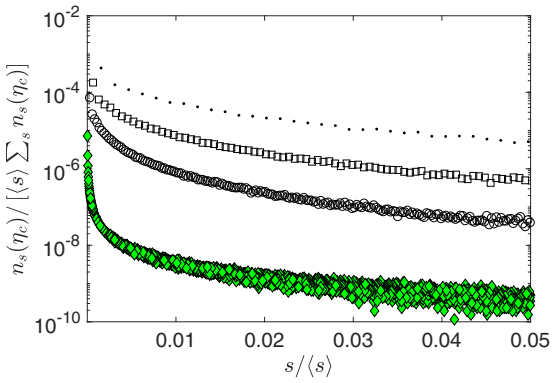

FIG. 11. Cluster size distribution at percolation $n_s(\eta_c)$ versus $s/\langle s\rangle$, normalized such that each curve has unit area, for void percolation through randomly placed spheres using Kerstein's method with $N = 300$ (dots), 10^3 (squares), and 10^4 (open circles). We also show the normalized $n_s(\eta_c)$ calculated using the connected void method for $N = 7 \times 10^3$ randomly placed spheres (filled diamonds).

Non-power-law decay in the void cluster size distribution, as displayed in Fig. 7(b) for jammed sphere packings, may be due to the existence of competing length scales in the system. The typical form of Eq. (10) at any porosity η is [47]

$$n_s = s^{-\tau} \exp(-s/s_\xi), \tag{C1}$$

where s_ξ is the number of sites in a cluster with correlation length ξ. In systems where ξ is the only length scale, $s_\xi \to \infty$ as $\eta \to \eta_c$ and Eq. (C1) reduces to Eq. (10). However, if there is another intrinsic length scale in the system that is still relevant at the void percolation transition, it is not necessarily true that $s_\xi \to \infty$. s_ξ can remain finite and add an exponential tail to $n_s(\eta_c)$. Indeed, this behavior is what we find for the connected void size distribution in jammed sphere packings. The "kink" in $n_s(\eta_c)$ in Fig. 7(b) indicates that $s_\xi \approx 20$. However, the end of the n_s distribution for void percolation in jammed packings of spheres seems to gain a power-law tail which matches that of void percolation on cubic lattices. Studies of larger systems would be required to confirm this similarity, but the data in Fig. 7(b) suggests that void percolation in jammed packings of spheres resembles that in systems on a cubic lattice.

This second length scale is most likely set by the nearest-neighbor distances between particles. Qualitatively, if the nearest-neighbor distance between particles is a δ function (or a set of δ functions, in the case of polydisperse spheres), there are a limited number of local cavities in the system. In particular, there can be small, particle-scale voids that persist even even at the percolation threshold. However, in packings of amino-acid-shaped particles and in systems of randomly placed spheres, there are a wide range of interparticle distances, and a continuous range of local cavity sizes that can form. In Fig. 8, we show that the void regions are well-connected for jammed packings of amino-acid-shaped particles and randomly placed spheres, while the void regions have a characteristic cavity size for jammed sphere packings at percolation onset. Thus, there is a well-defined Fisher exponent τ in jammed packings of amino-acid-shaped particles and randomly placed spheres, but not in jammed monodisperse and bidisperse sphere packings. Additionally, we do not find power law scaling for $n_s(\eta_c)$ in systems of randomly placed spheres when the void connectivity is analyzed using Kerstein's method [40]. We see in Fig. 11 that this is likely due to the sparsity of the Voronoi-based network, as smaller systems display an exponential decay in $n_s(\eta_c)$ rather than a power law decay. However, as the system size increases, the distribution of void cluster sizes approaches that for void clusters measured using the lattice-based connected-void method (described in Sec. II in the main text), which has many more sites per void cluster. Thus, if we were to study larger system sizes (e.g. $N \sim 10^5$), the void-cluster size distribution obtained using Kerstein's method would likely collapse onto the cluster size distribution we obtain using the connected-void method.

[1] K. A. Dill, Biochemistry **29**, 7133 (1990).
[2] J. Liang and K. A. Dill, Biophys. J. **81**, 751 (2001).
[3] J. Roche, J. A. Caro, D. R. Norberto, P. Barthe, C. Roumestand, J. L. Schlessman, A. E. Garcia, B. García-Moreno E., and C. A. Royer, Proc. Natl. Acad. Sci. USA **109**, 6945 (2012).
[4] N. V. Nucci, B. Fuglestad, E. A. Athanasoula, and A. J. Wand, Proc. Natl. Acad. Sci. USA **111**, 13846 (2014).
[5] M. T. Lerch, C. J. López, Z. Yang, M. J. Kreitman, J. Horwitz, and W. L. Hubbell, Proc. Natl. Acad. Sci. USA **112**, E2437 (2015).
[6] B. Borgo and J. J. Havranek, Proc. Natl. Acad. Sci. USA **109**, 1494 (2012).
[7] A. E. Eriksson, W. A. Baase, X.-J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthews, Science **255**, 178 (1992).
[8] W. Sheffler and D. Baker, Protein Sci. **18**, 229 (2009).
[9] S. J. Fleishman *et al.*, J. Mol. Biol. **414**, 289 (2011).
[10] J. C. Gaines, W. W. Smith, L. Regan, and C. S. O'Hern, Phys. Rev. E **93**, 032415 (2016).
[11] J. Gaines, A. Virrueta, D. Buch, S. Fleishman, C. S. O'Hern, and L. Regan, Protein Eng., Design Select. **30**, 387 (2017).
[12] D. Caballero, A. Virrueta, C. S. O'Hern, and L. Regan, Protein Eng., Design Select. **29**, 367 (2016).
[13] K. VanderWerf, W. Jin, M. D. Shattuck, and C. S. O'Hern, Phys. Rev. E **97**, 012909 (2018).
[14] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach, A. H. Xie, and R. D. Young, Phys. Rev. Lett. **62**, 1916 (1989).
[15] D. L. Stein, Proc. Natl. Acad. Sci. USA **82**, 3670 (1985).
[16] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
[17] R. J. Loncharich and B. R. Brooks, J. Mol. Biol. **215**, 439 (1990).
[18] D. Ringe and G. A. Petsko, Biophys. Chem. **105**, 667 (2003).
[19] J. Kertész, J. Phys. Lett. **42**, 393 (1981).
[20] A. L. Cuff and A. C. R. Martin, J. Mol. Biol. **344**, 1199 (2004).

[21] G. Wang and R. L. Dunbrack, Jr., Bioinformatics **19**, 1589 (2003).
[22] G. Wang and R. L. Dunbrack, Jr., Nucleic Acids Res **33**, W94 (2005).
[23] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson, J. Mol. Biol. **285**, 1735 (1999).
[24] S. J. Hubbard and J. M. Thornton, "Naccess" (1993), http://wolf.bms.umist.ac.uk/naccess/.
[25] J. C. Gaines, S. Acebes, A. Virrueta, M. Butler, L. Regan, and C. S. O'Hern, Proteins: Struct., Funct., Bioinf. **86**, 581 (2018).
[26] C. S. O'Hern, L. E. Silbert, A. J. Liu, and S. R. Nagel, Phys. Rev. E **68**, 011306 (2003).
[27] J. C. Gaines, A. H. Clark, L. Regan, and C. S. O'Hern, J. Phys.: Condens. Matter **29**, 293001 (2017).
[28] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Phys. Rev. Lett. **97**, 170201 (2006).
[29] D. Rozmanov and P. G. Kusalik, Phys. Rev. E **81**, 056706 (2010).
[30] F. M. Schaller, S. C. Kapfer, M. E. Evans, M. J. F. Hoffmann, T. Aste, M. Saadatfar, K. Mecke, G. W. Delaney, and G. Schröder-Turk, Philos. Mag. **93**, 3993 (2013).
[31] S. Weis, P. W. A. Schönhöfer, F. M. Schaller, M. Schröter, and G. E. Schröder-Turk, EPJ Web Conf. **140**, 06007 (2017).
[32] M. E. J. Newman and R. M. Ziff, Phys. Rev. E **64**, 016706 (2001).
[33] T. Aste and T. Di Matteo, Phys. Rev. E **77**, 021309 (2008).
[34] T. Aste, T. D. Matteo, M. Saadatfar, T. J. Senden, M. Schröter, and H. L. Swinney, Europhys. Lett. **79**, 24003 (2007).
[35] F. M. Schaller, R. F. B. Weigel, and S. C. Kapfer, Phys. Rev. X **6**, 041032 (2016).
[36] I. Jorjadze, L.-L. Pontani, K. A. Newhall, and J. Brujić, Proc. Natl. Acad. Sci. USA **108**, 4286 (2011).
[37] J. Li, Y. Cao, C. Xia, B. Kou, X. Xiao, K. Fezzaa, and Y. Wang, Nat. Commun. **5**, 5014 (2014).
[38] A. Boromand, A. Signoriello, F. Ye, C. S. O'Hern, and M. D. Shattuck, Phys. Rev. Lett. **121**, 248003 (2018).
[39] E. Pineda, P. Bruna, and D. Crespo, Phys. Rev. E **70**, 066119 (2004).
[40] A. R. Kerstein, J. Phys. A: Math. Gen. **16**, 3071 (1983).
[41] M. D. Rintoul, Phys. Rev. E **62**, 68 (2000).
[42] M. D. Rintoul and S. Torquato, J. Phys. A: Math. Gen. **30**, L585 (1997).
[43] S. Ostojic, E. Somfai, and B. Nienhuis, Nature **439**, 828 EP (2006).
[44] T. Shen, C. S. O'Hern, and M. D. Shattuck, Phys. Rev. E **85**, 011308 (2012).
[45] G. Lois, J. Blawzdziewicz, and C. S. O'Hern, Phys. Rev. Lett. **100**, 028001 (2008).
[46] Y. B. Yi, Phys. Rev. E **74**, 031112 (2006).
[47] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (CRC Press, Boca Raton, FL, 1994).
[48] G. Grimmet, *Percolation*, 2nd ed., Grundlehren der mathematischen Wissenschaften, Vol. 321 (Springer-Verlag, Berlin/Heidelberg, 1999).
[49] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede, and A. Tramontano, Proteins: Struct., Funct., Bioinfo. **86**, 7 (2018).
[50] L. Hovan, V. Oleinikovas, H. Yalinca, A. Kryshtafovych, G. Saladino, and F. L. Gervasio, Proteins: Struct., Funct., Bioinfo. **86**, 152 (2018).