Random close packing in protein cores

Jennifer C. Gaines,[1,2] W. Wendell Smith,[3] Lynne Regan,[1,2,4,5] and Corey S. O'Hern[1,2,3,6,7]

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

(Received 14 October 2015; revised manuscript received 11 February 2016; published 28 March 2016)

Shortly after the determination of the first protein x-ray crystal structures, researchers analyzed their cores and reported packing fractions $\phi \approx 0.75$, a value that is similar to close packing of equal-sized spheres. A limitation of these analyses was the use of extended atom models, rather than the more physically accurate explicit hydrogen model. The validity of the explicit hydrogen model was proved in our previous studies by its ability to predict the side chain dihedral angle distributions observed in proteins. In contrast, the extended atom model is not able to recapitulate the side chain dihedral angle distributions, and gives rise to large atomic clashes at side chain dihedral angle combinations that are highly probable in protein crystal structures. Here, we employ the explicit hydrogen model to calculate the packing fraction of the cores of over 200 high-resolution protein structures. We find that these protein cores have $\phi \approx 0.56$, which is similar to results obtained from simulations of random packings of individual amino acids. This result provides a deeper understanding of the physical basis of protein structure that will enable predictions of the effects of amino acid mutations to protein cores and interfaces of known structure.

DOI: 10.1103/PhysRevE.93.032415

I. INTRODUCTION

It is generally accepted that hydrophobic cores of proteins are tightly packed. In fact, many biology textbooks state that the packing fraction of protein cores is similar to that of densely packed equal-sized spheres with $\phi = 0.74$ [1]. Using a more accurate stereochemical representation, we show that the packing fraction of protein hydrophobic cores is $\phi \approx 0.56$ [Fig. 1(a), top left], which is similar to values for random close packing of nonspherical particles [2,3], not close packing of equal-sized spheres [Fig. 1(a), bottom right].

The most influential study of packing in protein cores was performed by Richards in 1974 [4]. He used Voronoi tessellation to calculate the packing fraction in the hydrophobic cores of two of the few proteins whose crystal structures had been determined at that time—lysozyme and ribonuclease S. He reported that the mean packing fraction of the two protein cores is $\phi_0 \approx 0.75$. More recent studies have obtained similar values for the packing fraction using larger data sets of protein cores [5–8]. We believe that the reason these prior studies have calculated such high values for the packing fraction of protein cores is that they use an extended atom representation of the heavy atoms. In this representation, hydrogen atoms are not included explicitly, rather the atomic radius of each heavy atom is increased by an amount proportional to the number of hydrogens that are bonded to it. An extended atom representation is often employed in computational studies of proteins because it significantly decreases the calculational complexity. In Fig. 1(b), we compare the extended atom representation of a Leu residue to one that includes hydrogen atoms explicitly. It is clear that the extended atom and explicit hydrogen representations of Leu possess different sizes and shapes.

In a 1987 paper on protein core repacking, Ponder and Richards [8] stated that "... the use of extended atoms was not satisfactory. In order for the packing criteria to be used effectively, hydrogen atoms had to be explicitly included ... ." Ponder and Richards argued that the extended atom model did not provide a sufficiently accurate representation of the stereochemistry of amino acids. In this paper, we examine the packing fraction of the hydrophobic cores of a large number of proteins using the explicit hydrogen representation, as Ponder and Richards [8] and other researchers [9] advocate.

We present several important results. First, we find that the average packing fraction of protein cores is $\phi \approx 0.56$. We show that the average packing fraction of each amino acid type is similar to the average packing fraction in protein cores, suggesting fairly uniform packing throughout the core. We obtain similar results from packing simulations of mixtures of residues that are isotropically compressed to jamming onset. We confirm the accuracy of our simulations by comparing the pair distribution functions of interatomic separations in protein cores and in the simulations. Both indicate only short-range positional order and the similarity of the two distributions confirms that the packing simulations mimic the atomic structure of protein cores.

The remainder of the paper is organized into three sections. In Sec. II, we describe the data set of protein crystal structures that we investigated in this study and the methods that we employed to calculate the packing fraction of the protein cores. In this section, we also provide strong support for the validity of the explicit hydrogen hard-sphere model for describing protein cores by showing that this model is able to reproduce the side chain dihedral angle distributions observed in proteins, whereas the extended atom model for proteins is not. In Sec. III, we show the results for the calculation of the

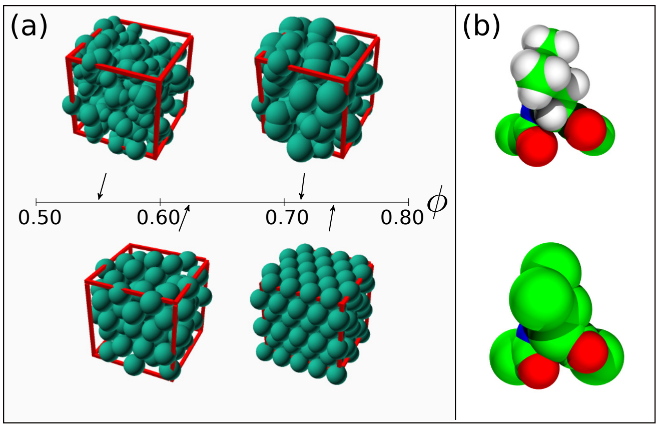

FIG. 1. (a) Visualization of core residues for a typical protein (carboxyl proteinase) in the Dunbrack database of crystal structures using explicit hydrogen (top left, $\phi \approx 0.56$) and extended atom (top right, $\phi \approx 0.72$) models compared to random close (bottom left, $\phi_{\rm RCP} \approx 0.64$) and face centered cubic packed (bottom right, $\phi_{\rm FCC} \approx 0.74$) systems with equal-sized spheres. (b) Leu residue with each atom represented as a sphere using the explicit hydrogen (top) and extended atom (bottom) models. The atom types are shaded green (carbon), red (oxygen), blue (nitrogen), and gray (hydrogen).

packing fraction in protein cores using the explicit hydrogen model and compare these results to those obtained using the extended atom representation. We then describe results from numerical simulations that compress collections of individual amino acids into jammed packings and compare the packing fraction and radial distribution function obtained from the simulations to those observed in protein cores. In Sec. IV, we summarize our results and propose future research directions.

II. METHODS

To calculate the packing fraction of protein cores, we use the Dunbrack database of high-resolution protein crystal structures, which is composed of 221 proteins with resolution $\leqslant 1.0$ Å, side chain B factors per residue $\leqslant 30$ Å^2, and R factor $\leqslant 0.2$ [10,11]. In prior studies, we showed that hard-sphere models of dipeptide mimetics with explicit hydrogens can recapitulate the side chain dihedral angle distributions observed in protein crystal structures [12–16].

As described in previous work [13], the hard-sphere model treats each atom i in a dipeptide mimetic as a sphere that interacts pairwise with all other nonbonded atoms j via

$$U_{\rm RLJ}(r_{ij}) = \frac{\epsilon}{72}\left[1-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right]^2\Theta(\sigma_{ij}-r_{ij}), \tag{1}$$

where r_{ij} is the center-to-center separation between atoms i and j, $\Theta(\sigma_{ij}-r_{ij})$ is the Heaviside step function, ϵ is the energy scale of the repulsive interactions, $\sigma_{ij}=(\sigma_i+\sigma_j)/2$, and $\sigma_i/2$ is the radius of atom i. A dipeptide mimetic is a single amino acid plus the C_α, C, and O of the prior amino acid and the N, H, and C_α of the next amino acid. Bond lengths and angles are set to the average values obtained from the Dunbrack database. Hydrogen atoms were added using the REDUCE software program [9], which sets the bond lengths for C-H, N-H, and S-H to 1.1, 1.0, and 1.3 Å, respectively, and the bond angles to 120° and 109.5° for angles involving $C_{\rm sp^2}$ and $C_{\rm sp^3}$ atoms. Additional dihedral angle degrees of freedom involving hydrogens are chosen to minimize steric clashes [9].

Predictions for the side chain dihedral angle distributions of a given dipeptide mimetic are obtained by rotating each of the side chain dihedral angles $\chi_1,\ldots,\chi_n$ and evaluating the total potential energy $U(\chi_1,\ldots,\chi_n)=\sum_{i<j}U_{\rm RLJ}(r_{ij})$ and Boltzmann weight

$$P(\chi_1,\ldots,\chi_n) \propto e^{-U(\chi_1,\ldots,\chi_n)/k_BT}. \tag{2}$$

We then average the Boltzmann weight over all dipeptide mimetics and normalize such that $\int P(\chi_1,\ldots,\chi_n)d\chi_1,\ldots,d\chi_n = 1$. We set the temperature $k_BT < 10^{-2}$ to be sufficiently small that we are in the hard-sphere limit and $P(\chi_1,\ldots,\chi_n)$ no longer depends on temperature. The values for the six atomic radii (C_{sp^3}, $C_{\rm aromatic}$: 1.5 Å; C_O: 1.3 Å; O: 1.4 Å; N: 1.3 Å; H: 1.10 Å; and S: 1.75 Å) were obtained in prior work [13] by minimizing the difference between the side chain dihedral angle distributions predicted by the hard-sphere dipeptide mimetic model and those observed in protein crystal structures for a small subset of amino acid types. The atomic radii are similar to values of van der Waals radii reported in earlier studies [4,15,17–27] (see Fig. 7 in the Appendix).

The packing fraction of each residue was calculated using

$$\phi = \frac{\sum V_i}{\sum V_i^v}, \tag{3}$$

where V_i is the nonoverlapping volume of atom i, V_i^v is the Voronoi volume of atom i, and the summation is over all atoms of a particular residue. The nonoverlapping volume of each atom is obtained by dividing overlapping atoms i and k by the plane of intersection between the two spheres. V_i^v for each atom was found using a variation of the VORO++ software library [28]. Voronoi cells were obtained for each atom using Laguerre tessellation, where the placement of the Voronoi cell walls is based on the relative radii of neighboring atoms (which is the same as the location of the plane that separates overlapping atoms).

We define core residues as those that are neither on the protein surface nor on the surface of an interior void. We identify surface and void atoms as those with empty space next to them. Points were found that were greater than 1.4 Å (approximately the radius of a water molecule) from the surface of all atoms in the protein using Monte Carlo sampling. The closest atom to each of these points was designated as a surface atom. For a residue to be considered a core residue, it must not contain any surface atoms. According to this definition and using the explicit hydrogen representation, proteins in the Dunbrack database had an average of 15 core residues. Ala, Cys, Gly, Ile, Leu, Met, Phe, and Val residues make up over 80% of the protein cores (see Table I). However, in our calculations of the packing fraction of protein crystal structures we included all amino acid types.

We also performed similar studies of the side chain dihedral angle distributions and packing analyses using the extended atom representation with the same atom types and radii used by Richards (N: 1.7 Å, O: 1.4 Å, O(H): 1.6 Å, C: 2.0 Å, and S: 1.8 Å) with the exception of C for the ring systems

TABLE I. The second and third columns give the number of times and frequency that each amino acid occurs in the cores of proteins in the Dunbrack database.

Amino Acid	No. in Core	% of Core
Ala	537	16.9
Arg	6	0.19
Asn	50	1.57
Asp	78	2.45
Cys	143	4.50
Gln	17	0.53
Glu	31	1.01
Gly	457	14.38
His	24	0.76
Ile	306	9.63
Leu	357	11.23
Lys	3	0.09
Met	90	2.8
Phe	141	4.44
Pro	63	1.98
Ser	194	6.10
Thr	136	4.28
Trp	28	0.88
Tyr	70	2.20
Val	446	14.03

(Phe, Tyr, Trp, Arg, and His), which was set to 1.7 Å [4]. For both explicit hydrogen and extended atom representations, we calculated ϕ for the core of a given protein using Eq. (3) with the summation over all atoms of all residues in the core. We also calculated the packing fraction for each residue in the core with the summation limited to all atoms in a given residue.

In Fig. 2, we compare the observed side chain dihedral angle distributions for Ile residues in the Dunbrack database and the predicted distributions from the hard-sphere dipeptide mimetic model using both the explicit hydrogen and extended atom representations. The observed distribution for Ile [Fig. 2 (left)] possesses one strong peak at $\chi_1 = 300^\circ$, $\chi_2 = 180^\circ$ and three minor peaks at $\chi_1 = 300^\circ$, $\chi_2 = 300^\circ$, $\chi_1 = 60^\circ$, $\chi_2 = 180^\circ$, and $\chi_1 = 180^\circ$, $\chi_2 = 180^\circ$. The side chain dihedral angle distribution for Ile predicted using the hard-sphere dipeptide mimetic model with the explicit hydrogen representation reproduces each of these features [Fig. 2 (center)]. In contrast, the high probability regions of χ_1-χ_2 space for the extended atom representation of the Ile dipeptide mimetic occur near $\chi_1 = 60^\circ$, $\chi_2 = 120^\circ$ and $\chi_1 = 300^\circ$, $\chi_2 = 120^\circ$, which have extremely low probability in the observed distributions. We find similar results for all other nonpolar residues. These results show that the extended atom model of a dipeptide mimetic does not reproduce the observed side chain dihedral angle distributions, whereas the explicit hydrogen model of a dipeptide mimetic does.

III. RESULTS

The results for the packing fraction analyses on core residues in all proteins in the Dunbrack database are shown in Fig. 3. For the explicit hydrogen representation, we find that the average packing fraction in protein cores is $\langle\phi\rangle_{\rm EH} \approx 0.56 \pm 0.02$ (blue circles), with fluctuations that are larger in proteins with small cores. This value is significantly lower than that obtained using the extended atom representation, $\langle\phi\rangle_{\rm EA} \approx 0.71 \pm 0.05$ (red squares), which is similar to $\phi_0 \approx 0.75$ reported in Ref. [4]. [The slight difference between $\langle\phi\rangle_{\rm EA}$ and ϕ_0 is due to the higher resolution of the Dunbrack database and that Richards averaged the local atomic packing fractions rather than taking the ratio of the total volumes as in Eq. (3).]

We also performed packing simulations of residues confined within a cubic box (with periodic boundary conditions) to determine whether $\langle\phi\rangle_{\rm EH} \approx 0.56$ can be explained by jamming of nonspherical objects [29]. We studied mixtures of N residues with the number of Ala, Ile, Leu, Met, Phe, and Val residues chosen from a weighted distribution that matched the percentages found in protein cores. (We focused on nonpolar residues, but because Gly has no side chain and Cys can form disulfide bonds, these were not included in the simulations.) We initialized the system to a small packing fraction ($\phi_i = 10^{-3}$), set the bond lengths, bond angles, backbone and side chain dihedral angles of each residue with values from randomly chosen instances of the amino acid in the Dunbrack database, and placed each of the individual residues in the simulation box with random initial positions and orientations.

We then compressed the system while keeping the overlaps between nonbonded atoms at approximately 10^{-6} of the atomic radii by minimizing the enthalpy $U + PV$ of the system,

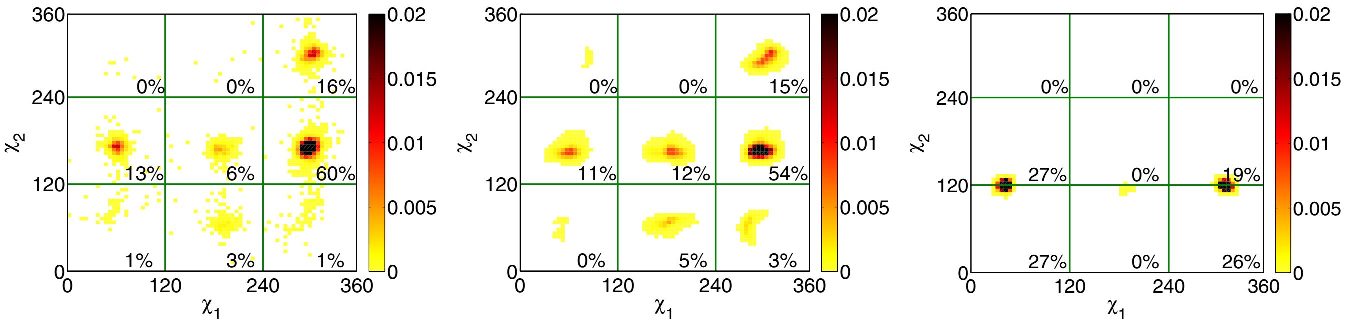

FIG. 2. (Left) The observed side chain dihedral angle probability distribution $P(\chi_1,\chi_2)$ for Ile residues in the Dunbrack database of protein crystal structures. We also show $P(\chi_1,\chi_2)$ predicted by the hard-sphere dipeptide mimetic model for Ile using the (center) explicit hydrogen and (right) extended atom representations. For the extended atom model, we used the atomic radii in the original work by Richards [4]. The probabilities increase from light to dark. The percentages give the fractional probabilities that occur in each of the nine square bins.

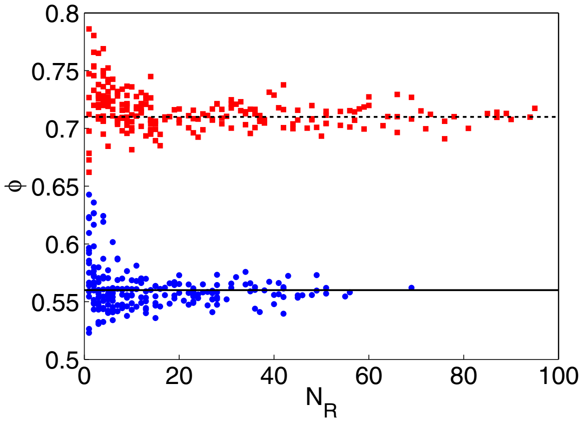

FIG. 3. A comparison of the packing fraction ϕ of the cores of proteins in the Dunbrack database as a function of the number of core residues N_R using the explicit hydrogen (blue circles) and extended atom (red squares) representations. More residues are designated as core using the extend atom model (25 on average) than using the explicit hydrogen model (15 on average). The solid and dashed horizontal lines indicate $\langle\phi\rangle_{\rm EH} \approx 0.56$ and $\langle\phi\rangle_{\rm EA} \approx 0.71$.

where U is the total repulsive Lennard-Jones potential energy between nonbonded atoms, $P = 10^{-6}\epsilon/\text{Å}^3$ is the pressure of the system, and V is the volume of the simulation box. The algorithm minimizes the enthalpy, using the conjugate gradient method, with respect to the variables $\vec{s}_i = \vec{r}_i/V^{1/3}$ and logarithm of the box volume $\eta \propto \ln(V/V_0)$, where V_0 is the initial volume. Residue conformations were strictly maintained using rigid body dynamics. We stopped the minimization algorithm when the system was in force balance, with the total force on each atom below the threshold value, $\max_i \sum_j |\vec{F}_{ij}| < 10^{-12}\epsilon/\text{Å}$ and final packing fraction ϕ_J.

Figure 4 shows that the distribution of packing fractions $P(\phi_J)$ from the packing simulations is similar to the distribution of packing fractions of protein cores from high resolution protein crystal structures. Both distributions possess a peak near 0.56 and have similar widths. Figure 4 includes results for $N = 24$ (~500 atoms), but we found similar results for $N = 8$ and 16. These results indicate that the connectivity of the protein backbone does not provide significant constraints on the free volume in protein cores.

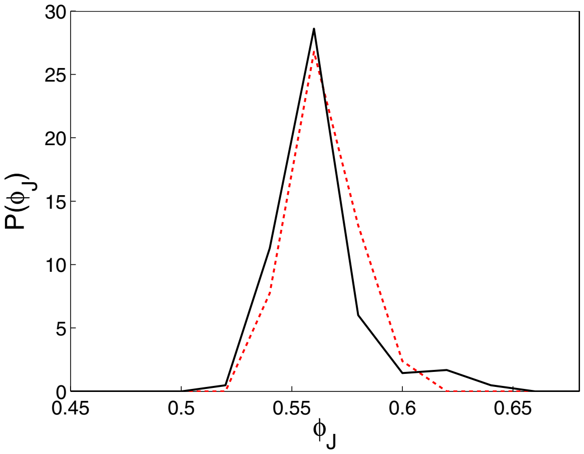

FIG. 4. The probability distribution (red dotted line) of packing fractions $P(\phi_J)$ from packing simulations of mixtures of residues found in protein cores. $P(\phi_J)$ from simulations was obtained from 100 jammed packings of $N = 24$ residues. The probability distribution of packing fractions from the cores of proteins in the Dunbrack database is shown by the solid black line.

Our simulations of packings of individual amino acids do not give rise to large packing fractions above 0.70 as found for the extended atom model for several reasons. First, the compression protocol that we implement represents a fast packing process, which gives rise to random close-packed structures. In contrast, slow packing protocols give rise to crystal close-packed structures with significant positional order [30]. For example, when we apply our compression protocol to a system composed of monodisperse spheres, we obtain random close packed structures with $\phi_J \approx 0.64$, not face centered cubic structures with $\phi_J = 0.74$ (see Fig. 9 in the Appendix).

We also employed our compression protocol to mixtures of atoms (without bond constraints) with four different radii and concentrations similar to those found in protein cores. Specifically, we generated packings of 400 atoms with radii 1.5 Å(C_{sp^3}, $C_{\rm aromatic}$), 1.3 Å(C_O, N), 1.4 Å(O), and 1.1 Å(H) and number concentrations 26, 13, 6.4, and 54.6%, respectively. The packing simulations of unequal-sized spheres give an average packing fraction of $\phi_J \approx 0.64$, as shown in Fig. 9 in the Appendix, which is similar to random close packing of monodisperse spheres. This packing fraction is not larger than random close packing for monodisperse spheres because the ratio of the radii of the largest and smallest atom types is close to 1 (i.e., only 1.4). Similar results have been found in our prior work on binary mixtures of hard spheres [31].

In contrast, the packing simulations for collections of amino acids yield $\phi_J \approx 0.56$. Atoms in proteins are bonded together and possess particular bond length and bond angle constraints. The fundamental packing units in our simulations are nonspherical amino acids (as shown in Fig. 8 in the Appendix), not individual spheres. The lower value $\phi_J \approx 0.56$ (compared to that for individual spheres) is due to the bulkiness of the amino acids, and matches the value found in protein cores.

We also investigated the presence of positional order in the cores of protein crystal structures and in the simulated packings by calculating the pair distribution function $g(r_{ij})$ of interatomic separations r_{ij}. In crystalline systems with long-range positional order, $g(r_{ij})$ possesses strong peaks corresponding to separations between lattice sites that do not decay with increasing r_{ij}. In contrast, $g(r_{ij})$ for protein cores only possesses strong peaks below 2 Å that correspond to bonded atoms and a weak next-nearest-neighbor peak, which indicates only short-range positional order (Fig. 5). $g(r_{ij})$ for the packings of amino acids generated from the simulations is very similar to that observed in protein cores, which further confirms that the packing simulations effectively mimic the atomic structure of protein cores.

To further analyze the packing efficiency in protein cores, we also calculated the distribution of the local packing fractions (i.e., ϕ for each residue type) in protein cores for

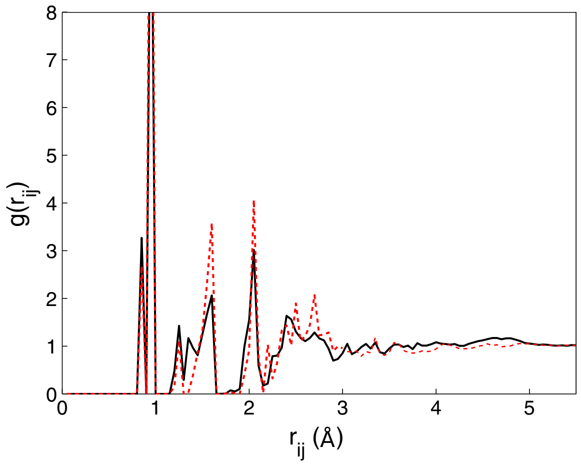

FIG. 5. The pair distribution function $g(r_{ij})$ of interatomic separations r_{ij} in protein cores from the Dunbrack data base (black solid line) and packings of individual amino acids generated from the packing simulations (red dotted line).

both the protein structures in the Dunbrack database and the packings from the simulations (Fig. 6). In Table II, we summarize the results for the average and standard deviation of the packing fraction for each core residue. We find that the distributions of the local packing fractions for each residue have similar average values, differing by only ≈5%. In addition, the average values for the local packing fractions are similar to the global average in the core with standard deviations that are slightly larger, which reflects the fact that the local packing fraction is obtained by averaging over fewer atoms than the global packing fraction.

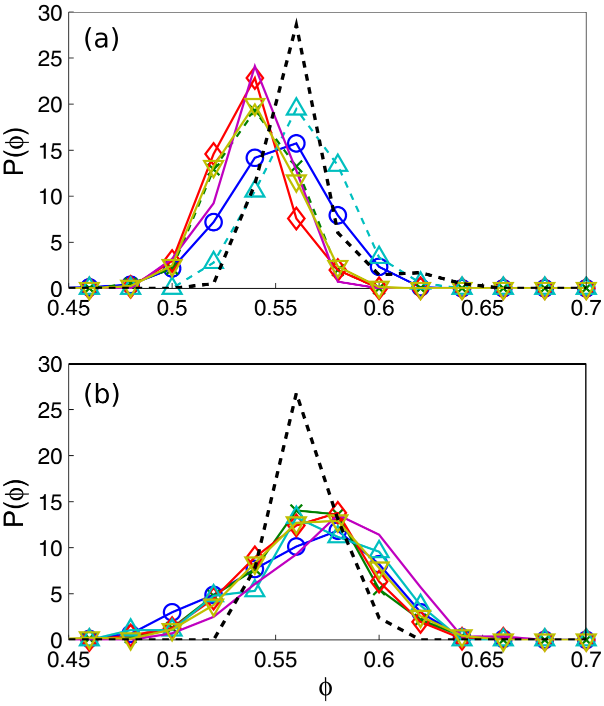

FIG. 6. The probability distribution of local packing fractions for Ala (blue circles), Ile (green crosses), Leu (red diamonds), Met (teal upward triangles), Phe (purple solid line), and Val (yellow downward triangles) residues (a) in the cores of proteins from the Dunbrack database and (b) from the packing simulations. We also show the probability distributions of the packing fractions for each protein core as black dotted lines for the observed and simulated structures in (a) and (b).

TABLE II. The mean and standard deviation $\Delta\phi$ of the packing fraction for each residue in protein cores (labeled c) and from simulations of mixtures of individual residues (labeled s). The last row gives the average packing fraction over all protein cores or over all 100 simulations.

Residue	ϕ_c	$\Delta\phi_c$	ϕ_s	$\Delta\phi_s$
Ala	0.55	0.02	0.56	0.03
Ile	0.54	0.02	0.56	0.03
Leu	0.54	0.02	0.56	0.03
Met	0.56	0.02	0.57	0.03
Phe	0.54	0.02	0.58	0.03
Val	0.54	0.02	0.57	0.03
Total	0.56	0.02	0.56	0.01

We also find that the average packing fraction of each amino acid type is similar to the average packing fraction in protein cores. (In the Dunbrack database, Ala and Met residues have a slightly larger average packing fraction than the rest of the amino acids, which is not reflected in the simulations.) The similarity of the average packing fraction for individual residues and the average packing fraction in protein cores suggests that there are only small variations of the packing fraction within each protein core (after removing large interior voids). Since we explicitly do not consider interior voids, the packing fraction of protein cores is determined roughly by the volume fraction that each amino acid occupies in the Voronoi cell formed by neighboring residues.

IV. CONCLUSIONS

In this paper, we showed that the explicit hydrogen hard-sphere model, which reproduces the side chain dihedral angle distributions observed in protein crystal structures, gives a packing fraction of $\langle\phi\rangle_{\mathrm{EH}} \approx 0.56$ for protein cores, not $\phi_0 \approx 0.75$ [4] found previously using the extended atom model. However, this result does not imply that protein cores are loosely packed. By comparing the packing fraction in protein cores to that found in simulations of collections of individual amino acids, we show that protein cores achieve dense *random* packing of amino acid packing subunits. The relatively low value for the packing fraction arises from the bulkiness of amino acids and their inability to pack efficiently in disordered configurations. Our results thus revise the prior picture of protein cores as dense packings of nearly equal-sized spheres.

Our results provide new insights into the atomic-scale structure of protein cores that can be applied to studies of amino acid mutations in protein cores and at protein-protein interfaces. Recent studies have shown that packing efficiency can be used as a metric for assessing the stability of mutations in proteins [9,32]. However, most of the current work on assessing the packing efficiency of mutated structures employs the extended atom model and does not implement the Voronoi tessellation methods presented here [32,33]. In future studies, we will build on this work and determine whether mutations lower or raise the packing fraction outside of the range found in protein cores. Developing a method to calculate accurately the packing fraction in protein cores and at protein-protein interfaces is a significant step forward in enabling researchers to critically assess mutations and new designs.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences, National Library of Medicine Training Grant No. T15LM00705628 (J.C.G.), and National Science Foundation Grant No. DMR-1307712 (L.R.). This work also benefited from the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center and the National Science Foundation (Grant No. CNS-0821132), which in part funded acquisition of the computational facilities.

APPENDIX

In this section, we present additional details about the explicit hydrogen hard-sphere model for describing protein structure. In Fig. 7, we display the values for the six atomic radii that we used in the current study and show that they are similar to values of van der Waals radii reported in earlier studies [4,15,17–27]. In Fig. 8, we include illustrations of six amino acids (Ala, Ile, Leu, Met, Phe, and Val) that are commonly found in protein cores using the explicit hydrogen representation and atomic radii given in Fig. 7. Figure 8 emphasizes the bulkiness of amino acids, which limits their ability to pack efficiently in disordered arrangements in protein cores.

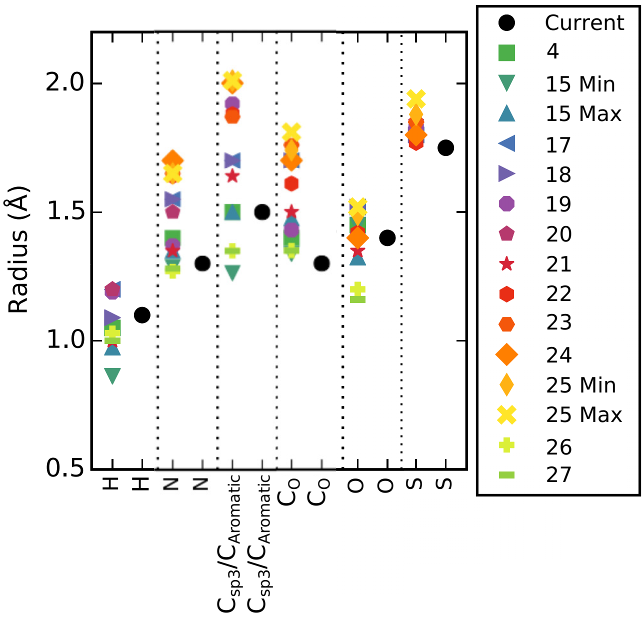

FIG. 7. The atomic radii for the six atom types (H, N, $C_{sp^3}/C_{\text{aromatic}}$, C_O, O, and S) used in the explicit hydrogen hard-sphere dipeptide model (black circles) compared to definitions used in other studies [4,15,17–27]. The atom sizes for the explicit hydrogen hard-sphere model were chosen so that the side chain dihedral angle distributions predicted by the model match the observed distributions for Leu and Val. Using these atom sizes, we also confirmed that the side chain dihedral angle distributions predicted from the hard-sphere dipeptide model for Ile, Phe, Tyr, Thr, Ser, and Cys also agree with the observed side chain dihedral angle distributions.

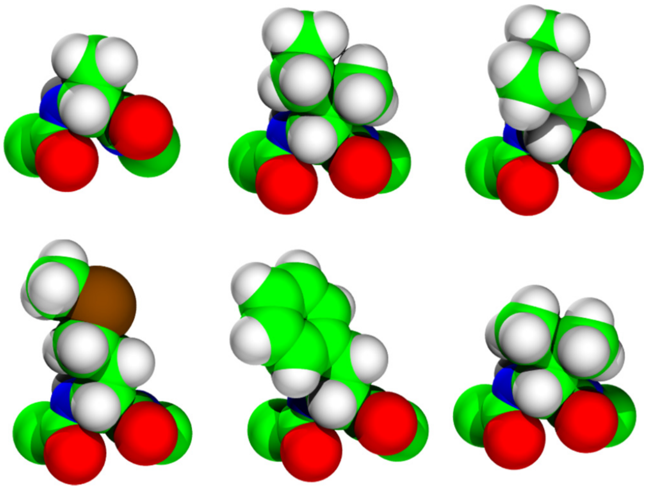

FIG. 8. Illustrations of Ala, Ile, Leu, Met, Phe, and Val (from left to right and top to bottom) dipeptide mimetics in the explicit hydrogen representation using the atomic radii in Fig. 7: C (green), O (red), N (blue), H (gray), and S (brown).

In Fig. 9, we compare the probability distribution of jammed packing fractions $P(\phi_J)$ for packings of amino acids and for packings of individual spheres obtained from the packing simulations described in Sec. II. $P(\phi_J)$ from packings of amino acids shows a peak near 0.56. In contrast, $P(\phi_J)$ for individual

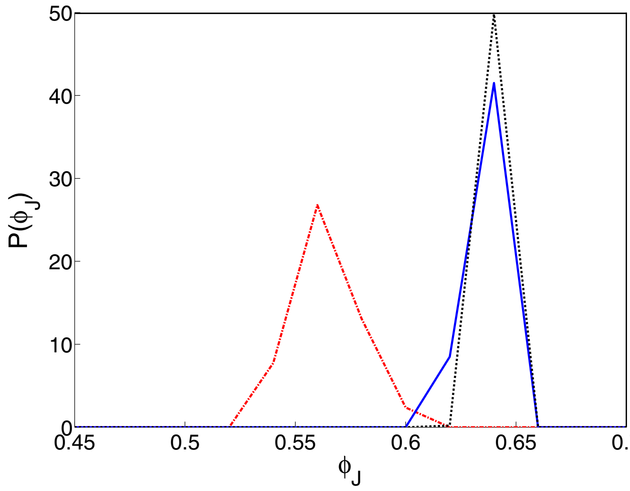

FIG. 9. The probability distribution of packing fractions $P(\phi_J)$ from the packing simulations. The distribution of packing fractions for packings of amino acids (red dash-dotted line) was obtained from 100 packings each containing $N = 24$ residues. We also show $P(\phi_J)$ for packings of unequal sized spheres with the same atomic radii and number fractions as found in protein cores (black dotted line) and $P(\phi_J)$ for monodisperse spheres (blue solid line).

spheres with different atomic sizes and number concentrations that match those in protein cores possesses a peak near 0.64, which is similar to random close packing for monodisperse spheres.

[1] J. Kyte, *Structure in Protein Chemistry*, 2nd ed. (Garland Science, New York, 2007).

[2] G. T. Nolan and P. E. Kavanagh, Random packing of nonspherical particles, Powder Technol. **84**, 199 (1995).

[3] X. Jia, R. Caulkin, R. A. Williams, Z. Y. Zhou, and A. B. Yu, The role of geometric constraints in random packing of non-spherical particles, Europhys. Lett. **92**, 68005 (2010).

[4] F. M. Richards, The interpretation of protein structures: Total volume, group volume distributions and packing density, J. Mol. Biol. **82**, 1 (1974).

[5] J. Liang and K. Dill, Are proteins well-packed?, Biophys. J. **81**, 751 (2001).

[6] P. J. Fleming and F. M. Richards, Protein packing: Dependence on protein size, secondary structure and amino acid composition, J. Mol. Biol. **299**, 487 (2000).

[7] K. Rother, R. Preissner, A. Goede, and C. Frommel, Inhomogeneous molecular density: Reference packing densities and distribution of cavities within proteins, Bioinformatics **19**, 2112 (2003).

[8] J. W. Ponder and F. M. Richards, Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes, J. Mol. Biol. **193**, 775 (1987).

[9] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. **285**, 1735 (1999).

[10] G. Wang and R. L. Dunbrack Jr., PISCES: A protein sequence culling server, Bioinformatics **19**, 1589 (2003).

[11] G. Wang and R. L. Dunbrack Jr., PISCES, Recent improvements to a PDB sequence culling server, Nucleic Acids Res. **33**, W94 (2005).

[12] A. Q. Zhou, D. Caballero, C. S. O'Hern, and L. Regan, New insights into the interdependence between amino acid stereochemistry and protein structure, Biophys. J. **105**, 2403 (2013).

[13] A. Q. Zhou, C. S. O'Hern, and L. Regan, Predicting the side-chain dihedral angle distributions of non-polar, aromatic, and polar amino acids using hard sphere models, Proteins Struct. Funct. Bioinf. **82**, 2574 (2014).

[14] A. Q. Zhou, C. S. O'Hern, and L. Regan, Revisiting the Ramachandran plot from a new angle, Protein Sci. **20**, 1166 (2011).

[15] A. Q. Zhou, C. S. O'Hern, and L. Regan, The power of hard-sphere models: Explaining side-chain dihedral angle distributions of Thr and Val, Biophys. J. **102**, 2345 (2012).

[16] D. Caballero, J. Määttä, A. Q. Zhou, M. Sammalkorpi, L. Regan, and C. S. O'Hern, Intrinsic α-helical and β-sheet conformational preferences: A computational case study of Alanine, Protein Sci. **23**, 970 (2014).

[17] C. Ramakrishnan and G. N. Ramachandran, Stereochemical criteria for polypeptide and protein chain conformations, Biophys. J. **5**, 909 (1965).

[18] A. Bondi, Van der Waals volumes and radii, J. Phys. Chem. **68**, 441 (1964).

[19] Element data and radii, Cambridge Crystallographic Data Centre. https://www.ccdc.cam.ac.uk/theccdcprofile [Online; accessed December 4, 2011].

[20] D. Seeliger and B. L. de Groot, Atomic contacts in protein structures. A detailed analysis of atomic radii, packing, and overlaps, Proteins Struct. Funct. Bioinf. **68**, 595 (2007).

[21] L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, 1948).

[22] L. L. Porter and G. D. Rose, Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints, Proc. Natl. Acad. Sci. USA **108**, 109 (2011).

[23] J. Tsai, R. Taylor, C. Chothia, and M. Gerstein, The packing density in proteins: Standard radii and volumes, J. Mol. Biol. **290**, 253 (1999).

[24] C. Chothia, Structural invariants in protein folding, Nature (London) **254**, 304 (1975).

[25] A. J. Li and R. Nussinov, A set of van der Waalsand coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking, Proteins Struct. Funct. Bioinf. **32**, 111 (1998).

[26] F. A. Mamony, L. M. Carruthers, and H. A. Scheraga, Intermolecular potentials from crystal data. iii. Determination of empirical potentials and application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines, and amides, J. Phys. Chem. **78**, 1595 (1974).

[27] N. L. Allinger and Y. H. Yuh, Quantum Chemistry Program Exchange **12**, 395 (1980).

[28] C. H. Rycroft, Voro++: A three-dimensional Voronoi cell library in C++, Chaos **19**, 041111 (2009).

[29] C. F. Schreck, M. Mailman, B. Chakraborty, and C. S. O'Hern, Constraints and vibrations in static packings of ellipsoidal particles, Phys. Rev. E **85**, 061305 (2012).

[30] C. F. Schreck, C. S. O'Hern, and L. E. Silbert, Tuning frictionless disk packings from isostatic to hyperstatic, Phys. Rev. E **84**, 011305 (2011).

[31] K. Zhang, W. W. Smith, M. Wang, Y. Liu, J. Schroers, M. D. Shattuck, and C. S. O'Hern, Connection between the packing efficiency of binary hard spheres and the glass-forming ability of bulk metallic glasses, Phys. Rev. E **90**, 032311 (2014).

[32] W. Sheffler and D. Baker, RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design and validation, Protein Science **18**, 229 (2009).

[33] W. Sheffler and D. Baker, RosettaHoles2: A volumetric packing measure for protein structure refinement and validation, Protein Science **19**, 1991 (2010).