

Packing proteins for pretty structures.

Bikram Chatterji

Introduction & Background

Protein packing density is a critical feature for scientists to characterize; the efficiency of packing, in terms of sheer volume of atoms per volume of space, is thought to be determined primarily by hydrophobicity concerns, and has implications for protein stability, structure, and thus function.¹ The traditional approach to studying packing density has been to (1) identify the locations (ie: x , y and z coordinates) of atoms in a 3D structure of the protein using, for instance, crystallization, (2) estimate the volume of the structure, generally using Voronoi polyhedra or some variant (perhaps using a probe-sphere to determine surface features), (3) divide a theoretical minimum volume for the structure (represented by the van der Waals volumes of its atoms) by the calculated actual volume to determine the packing efficiency.² Studies of packing density using this protocol have suggested that the interior core of a globular protein is generally very dense, but is not uniformly packed.² Other studies examining the packing of membrane protein α -helices have shown that they are of varying densities in the transmembrane (TM) region, but that the possibility for very tight packing exists and is often realized.³ Furthermore, a coarse-grain modeling study, which simplifies the task of representing a protein by operating at a large scale (ie: having the “grains” represent residues rather than atoms) also shows coordinating residues packing into the most efficient possible configuration.⁴ It is to be expected that Nature would

place a premium on developing compact proteins; such structures are stable and regular in their behavior, and furthermore are necessary to maintain the integrity of the hydrophobic core against the invasion of water molecules. On the other hand, low-density regions, characterized by cavities and voids, can potentially allow for movement of the protein (ie: it can change its shape)³ or, if close to the protein surface, can serve as docking regions for ligands.¹ In any case, it seems that the packing density is closely related to the structural features of the protein and functionality of the local region under consideration.

Proposal

My proposal stems from the realization that this relationship does not observe a strict causality. The current protocol of determining first the structure, then the volume, and finally the packing density fits in with our understanding that the density can be completely determined by the other two components, and so we should calculate them first. I suggest that we can combine packing density with volume information to conjecture possible three-dimensional structures for the protein, the reverse of the process listed above. While working in this direction will not yield an absolute answer – there can be a variety of structure which have the same packing density – it may yield interesting results and help to steer structural models in the right direction.

The advances I would draw upon in arguing for the feasibility of this “reverse-determination” technique include a recent proof⁵ of Kepler’s 400-year-old packing

conjecture and a systematic characterization of the radii and volumes of atomic groups in proteins.⁶

Evidence leading to the proposal.

Kepler's conjecture posits that the closest packing one can achieve for an infinitely large collection of same-sized spheres is that of a face-centered cubic (fcc) lattice, essentially the arrangement one sees oranges stacked in at a supermarket. In order to prove the conjecture Thomas Hales had to rule out all alternative configurations for a local area by showing that they were less efficiently packed than the fcc (in the process he developed a new variant of the Voronoi methodology called "star decomposition"). Essentially, Hales starts with a number of objects, generates configurations in which they could fit together, and then thresholds to pick out configurations (in this case a configuration) which have a particular packing density.⁵ He begins with a target density and emerges with a structure. Elimination of unsatisfactory configurations is largely automated; most of the work here could be done using linear programming and employing general arguments. In fifty cases out of 5094 to start, Hales had to look in detail at the structure to determine its packing density, but this was only necessary because he had a very stiff requirement, and a single exception to threshold would have been vital. Presumably, for our purposes, we would be happy to have a reasonably-sized but varied set of structure returned.

Interestingly, researchers recently tackled a problem similar to Kepler's, relating to the optimal packing of strings instead of spheres, and emerged with a helical arrangement bearing both qualitative and quantitative similarities to the α -helices found in proteins.⁷ In this case, the computational study was done initially and it was only retrospectively found to have biological significance. I propose using the technique as a prospective tool.

In order to achieve this, it will be necessary to have a good handle on the "objects" to be fed into Hales's algorithm. Recent efforts to definitively quantify the sizes of protein components are vital in this respect. An initial study made in 1999 characterized thirteen major atomic groups which form the building blocks for amino acids; where these thirteen groupings are based purely on chemical characteristics (number of covalent bonds, hydrophobicity, etc.) a subsequent study clustered the major groups based on numerical volume data, and made the case for expanding the number of groupings to eighteen.⁸ These groups, or 'types' are said to be capable of representing all of the various kinds of packing seen in proteins.

Potential pitfalls.

A first slightly confusing point is whether we would want to use the types, which are already biased to be in a particular configuration, or rather simply use van der Waals data to translate information from the primary sequence into atoms which can be arranged in an optimal configuration by Hales's system. The types have the advantage of representing larger-scale building blocks of amino acids, whereas van der Waals radii refer to single atoms. Another stumbling block is the necessity for specification of some of the structure (namely the amino-acid sequence) prior to execution of the algorithm. We would certainly not want to emerge with configurations which disregard the primary structure of the protein, and if this constraint could be implemented, it would hopefully

keep the number of emergent structures manageable. We should hope, therefore, that we are allowed to enter sequence data to begin with. Of course, the most important question of all is, where do we set the packing threshold? We have seen that Nature seems to favor close packing, but there is ostensibly no reason not to target a lower-density configuration. In addition, since Hales's algorithm was explicitly designed to solve a local packing problem, we could conceive of setting different thresholds for different parts of the protein, thereby emulating experimental results.

-
- ¹ Chakravarty S, Bhinge A & Varadarajan R. A Procedure for Detection and Quantitation of Cavity Volumes in Proteins. *J. Biological Chemistry*. 277(35):31345-31353 (2002).
- ² Gerstein M & Richards FM. Protein Geometry: Volumes, Areas, and Distances. Chapter 22 of *The International Tables for Crystallography*. Rosemum MG & Arnold E., eds., (2000).
- ³ Adamian L & Liang J. Helix-Helix Packing and Interfacial Pairwise Interactions of Residues in Membrane Proteins. *J Mol Biol*. (311) 891-907 (2001).
- ⁴ Bagci Z, Jernigan RL & Bahar I. Residue coordination in proteins conforms to the closest packing of spheres. *Polymer* 43: 451-459 (2002).
- ⁵ discussed in:
Cipra B. Packing Challenge Mastered At Last. *Science*. 281: 1267 (1998).
- ⁶ Tsai J, Taylor R, Chothia C & Gerstein M. The Packing Density in Proteins: Standard Radii and Volumes.
- ⁷ Maritan A, Micheletti C, Trovato A & Banavar J. Optimal shapes of compact strings. *Nature*. 406:287-290.
- ⁸ Tsai J, Voss N & Gerstein M. Determining the minimum number of types necessary to represent the sizes of protein atoms. *Bioinformatics*. 17(10):949-956 (2001).