LETTER

Reply to Zimmerman et al.: The space of single domain protein structures is continuous and highly connected

At the level of structurally significant relationships between proteins, the major conclusions of our article (1) were: protein structure space is continuous [where "one can link two arbitrarily selected structures, following a path of statistically significant similar structures"(1)] and highly connected. Zimmerman et al. (2) focus on average network path lengths and question these conclusions on the basis that the average shortest path is seven. In fact, seven is not the average shortest path, but the longest minimum path length between almost all protein structures (see ref. 1 and its Figs. 2*B* and 1*B*). All arguments based on average path lengths misrepresent our results. Even with a large average path length, provided that essentially every protein structure belongs to the largest strongly connected component (LSCC), then protein structure space is continuous. Zimmerman et al. (2) also argue that our results are a feature of random networks. Indeed, this agrees with our work where we showed that, with the same high local connectivity (an intrinsic protein property) as in the protein network, "the results are very close to what happens when random digraphs with the same distribution of first neighbors are generated; Fig. 2*B*, thin line" (1).

Zimmerman et al. (2) claim that if there is extensive clustering, then the database of protein structures is either incomplete or space is not continuous. We presented strong arguments in ref. 3 that the space of compact, single domain proteins is likely complete, a widely accepted view (4). Even with extensive clustering, structure space would be continuous provided that the LSCC contains essentially all compact single domain structures, but it would be anisotropic/not highly connected, contradicting claims in ref. 1. However, Fig. 2*B* in ref. 1 strongly suggests that it is highly connected/isotropic. As shown in Fig. 1*A*, even deleting the top 10% of protein structures with the largest number of first neighbors, the LSCC has >96% of protein structures. In Fig.1*B*, for each protein structure, we plot the fraction of other protein structures in the LSCC that are no more than *k*th neighbors. By $k = 7$, protein structure space is highly isotopic; the idea of loosely connected dense clusters is incorrect. Thus, the conclusions of our article (1) remain valid.

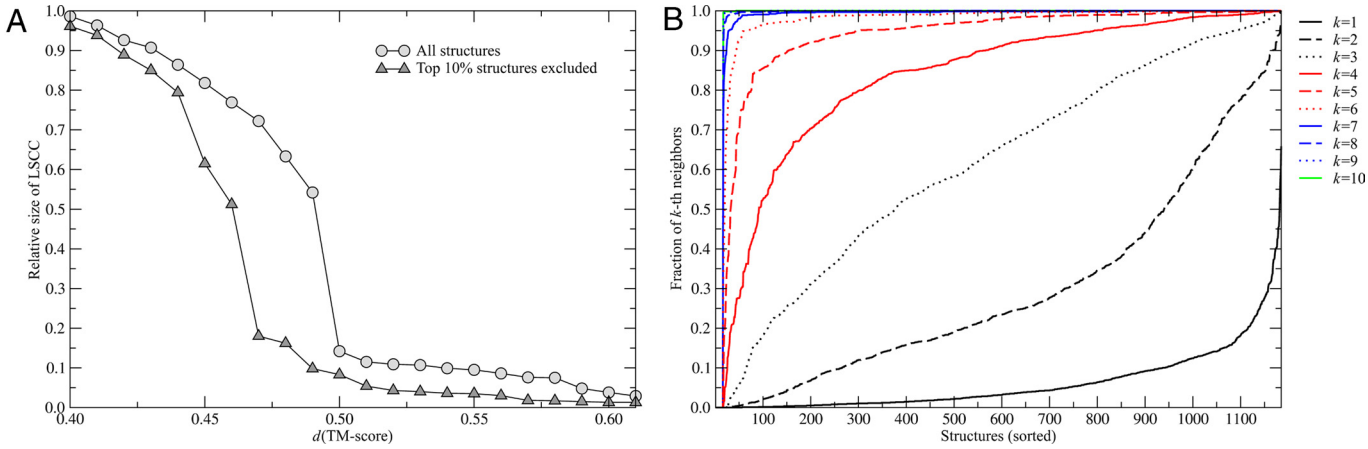

Fig. 1. High connectivity and isotrophy of protein structure space. (*A*) Relative size of the LSCC as a function of TM-score cutoff *d* for PDB200^{holo} (1), in triangles (excluding) and circles (including) the top 10% of structures with the largest number of first neighbors. (*B*) For each protein in PDB200^{holo}, the fraction of structures in the LSCC that are within *k*th neighbors; a TM-score threshold of 0.4 (indicative of significant similarity) is used.

Jeffrey Skolnick*[1], *Michal Brylinski*, *and Seung Yup Lee
Center for the Study of Systems Biology, Georgia Institute of Technology

1. Skolnick J, Arakaki AK, Lee SY, Brylinski M (2009) The continuity of protein structure space is an intrinsic property of proteins. *Proc Natl Acad Sci USA* 106:15690–15695.
2. Zimmermann M, Towfic F, Jernigan R, Kloczkowski A (2009) Short paths in protein structure space originate in graph structure. *Proc Natl Acad Sci USA* 10.1073/pnas.0912339106.
3. Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J (2006) On the origin and highly likely completeness of single-domain protein structures. *Proc Natl Acad Sci USA* 103:2605–2610.
4. Sadreyev RI, Kim BH, Grishin NV (2009) Discrete-continuous duality of protein structure space. *Curr Opin Struct Biol* 19:321–328.

Author contributions: J.S. designed research; J.S., M.B., and S.Y.L. performed research; J.S., M.B., and S.Y.L. analyzed data; and J.S. wrote the paper.
The authors declare no conflict of interest.
[1]To whom correspondence should be addressed. E-mail: skolnick@gatech.edu.