

## Letter to the Editor<sup>1</sup>

# The 11th Misconception?

Nicholas M. Glykos

Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Dragana, 68100 Alexandroupolis, Greece

Submitted September 9, 2010; Accepted September 30, 2010

The recent excellent essay by Robic (2010) awakened an old concern of mine regarding whether “sequences,” be they protein or nucleic acid, have escaped the limits of their useful abstraction and acquired the status of a nonexistent physical reality. The original observation that formed the basis for this notion lies with the widespread belief that protein folding involves a change of dimensionality. My sense is that the way that the central dogma is taught and the use of schematic diagrams to teach protein folding (e.g., Branden and Tooze, 1999, p. 3) combine to reinforce students’ beliefs that protein folding involves a transformation from the one-dimensional sequence space to the three-dimensional structure space. This erroneous belief is accentuated by classroom discussions about the folding problem, which are often initiated with a question about how amino acid sequence determines protein structure. I believe that students then come to visualize and conceptualize protein folding as the process through which a one-dimensional protein sequence is transformed to a three-dimensional protein structure.

When my students display this conception, I respond by asking questions such as “What is the dimensionality of what comes out of the ribosome? Does the dimensionality of an unfolded polypeptide chain differ from a folded protein? In terms of electrostatics, molecular dynamics, bond formation, and solvent interactions, which is more complex—an unfolded or a folded protein? Is protein folding a change

in dimensionality, conformation, or both?” When I confront students with these simple chemical and physical arguments, they often dismiss the whole subject as obvious. Indeed, students do know that a nascent polypeptide chain coming out of ribosome is neither a two-dimensional chemical formula nor a one-dimensional string of letters. Yet I would argue that this knowledge is obscured by a teaching approach that insists on connecting protein structure and protein folding with protein sequence. Specifically, I propose that two aspects of how sequence-folding-structure relationships are taught distract students from learning the phenomena that underpin protein folding:

1. As instructors, we insist on ignoring mounting experimental evidence that shows that proteins with no detectable sequence similarity, such as globins, can have essentially identical structures. Likewise, proteins sharing very high sequence identity can have significantly different structures (Kosloff and Kolodny, 2008). Clearly, if dissimilar sequences can lead to practically identical structures, and nearly identical sequences can lead to significantly different structures, then the mantra “sequence determines structure” becomes difficult to defend. Maybe it is time to substitute the sentence “sequence determines structure” with the sentence “unfolded structure determines folded structure.”
2. When we teach protein folding, we emphasize changes of protein conformation (unfolded–elongated–random-coil-like chain → folded–compact–stable structure) and not what really drives protein folding: interaction energies and the resulting energy landscape. For example, when “toobers” (3-D Molecular Designs, Milwaukee, WI, [www.3dmoleculardesigns.com/toobers.php](http://www.3dmoleculardesigns.com/toobers.php)) are used to teach protein folding, the emphasis should be placed not on the tube, which represents the protein backbone, but on the pushpins and their colors, which represent the physical properties of the side chains. Another strategy would be to replace Figure 1.1 of Branden and Tooze (1999; see <http://tinyurl.com/Btfig11>) with Figure 2 of Dinner *et al.* (2000; see <http://tinyurl.com/foldlandscape>) or <http://tinyurl.com/ffunnel> for a similar image).

DOI: 10.1187/cbe.10-09-0116

Address correspondence to: Nicholas M. Glykos ([glykos@mbg.duth.gr](mailto:glykos@mbg.duth.gr)).

<sup>1</sup>A short note inspired by the essay of Robic S (2010). *CBE Life Sci Educ* 9, 189–195.

© 2011 N. M. Glykos. *CBE—Life Sciences Education* © 2011 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

In my opinion, these issues are exacerbated by the massive efforts to generate sequence data and the hype surrounding “omic” projects, both of which made us believe that sequences (and not 42) are the answer to “life, the universe, and everything.” This brings me back to the opening sentence of my letter. Biological sequences are an abstraction of an abstraction. First, we substitute the complexity of a proper three-dimensional entity, such as an amino acid residue, with a two-dimensional chemical formula that describes only its molecular composition and covalent bonding. Second, we substitute these chemical formulas with single alphabet letters. *And then we forget that we are making abstractions and behave as if sequences do exist and this artificial one-dimensionality is real.* Sequence usage has become so widespread that we have started using sequences for dealing with problems, such as understanding protein folding, that by their nature defy “sequence” abstraction. Maybe, just maybe, we have had enough of “sequences.”

## ACKNOWLEDGMENTS

I thank Erin Dolan for her useful comments and suggestions on this letter, and Aaron Dinner for making Figure 2 of the Dinner *et al.* (2000) paper freely available for download.

## REFERENCES

- Branden C, Tooze J (1999). Introduction to Protein Structure, New York: Garland Publishing.
- Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 25, 331–339.
- Kosloff M, Kolodny R (2008). Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins* 71, 891–902.
- Robic S (2010). Mathematics, thermodynamics, and modeling to address ten common misconceptions about protein structure, folding, and stability. *CBE Life Sci Educ* 9, 189–195.