

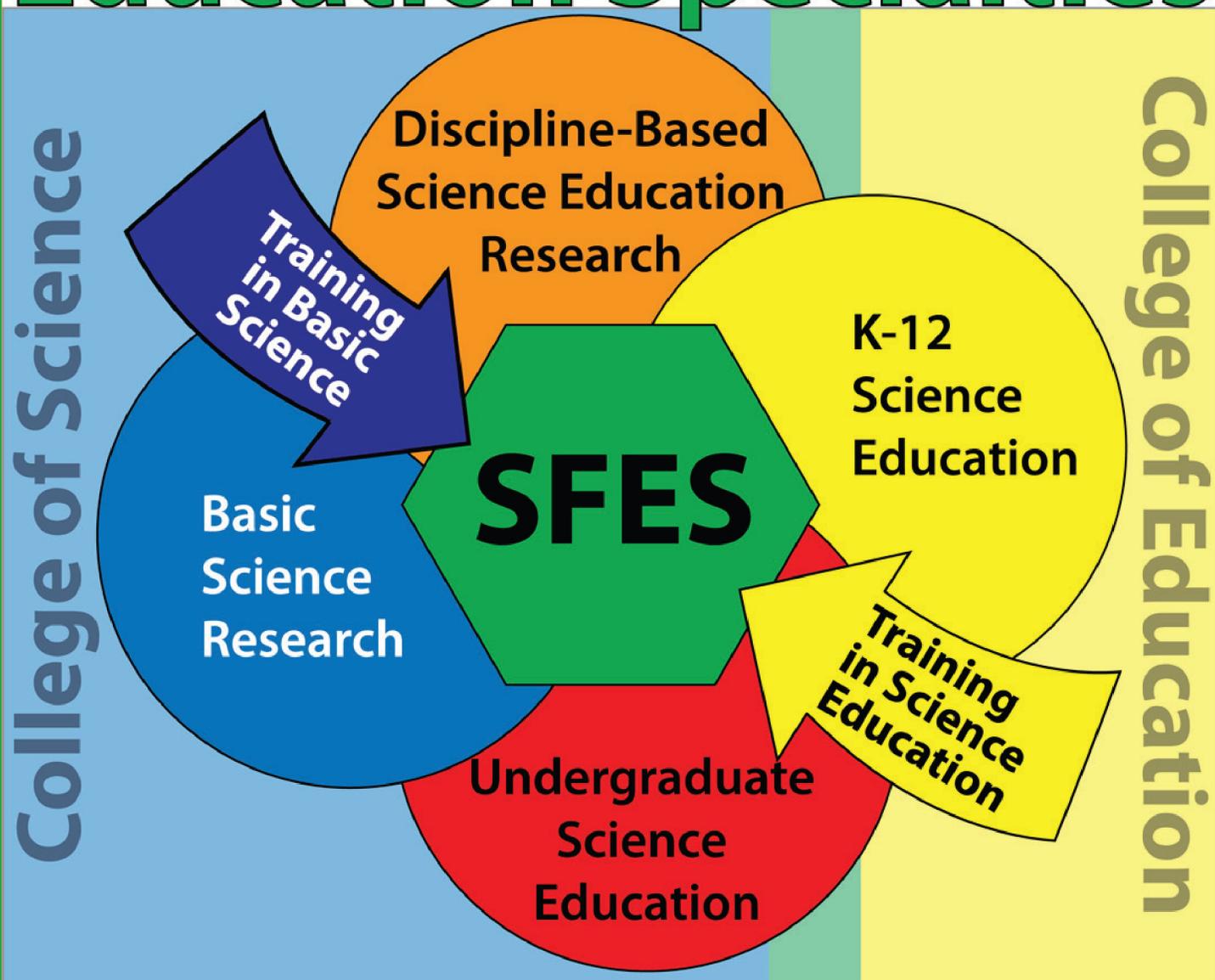
Published by

THE AMERICAN
SOCIETY FOR
CELL
BIOLOGY



Life Sciences Education

Science Faculty with Education Specialties



Letter to the Editor [§]

A short note inspired by the essay of Robic, S. (2010), *CBE Life Sci Educ*, **9**, 189-195.

The 11th Misconception ?

Nicholas M. Glykos, Department of Molecular Biology and Genetics,
Democritus University of Thrace, University campus (Dragana), 68100 Alexandroupolis, Greece,
Web: <http://utopia.duth.gr/~glykos/> e-mail: glykos@mbg.duth.gr

Dear Sir/Madam,

The recent excellent essay by Robic (2010) awakened a rather old notion of mine concerning the extent to which what we call 'sequences' (be they protein or nucleic acid), have escaped the limits of their useful abstraction and acquired the status of (a non-existent) physical reality.

The original observation that formed the basis for this notion lies with the widespread belief –both amongst students, but occasionally also between authors of published research papers in acclaimed journals– that there is a change of dimensionality involved in protein folding. I believe that the combination of how the central 'dogma' is being taught, together with the introductory schematic diagrams concerning protein folding [see, for example, page 3 of the second edition of Branden & Tooze

(1999), reproduced on p.4 of this manuscript], lead the students to believe that protein folding involves a change of dimensionality, ie. that it involves going 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 start with a sentence in the spirit of “how does the sequence determines the structure?”. Such approaches, I believe, lead the students down the wrong path of visualizing (and, thus, conceptualizing) protein folding as the process through which a one-dimensional protein sequence is transformed to a three-dimensional protein structure.

It is worth noting that when students bearing such ideas are confronted with simple physics-based

§ Glykos, N.M. (2011), "The 11th Misconception?", A letter to the Editor, *CBE Life Sci Educ*, **10**(1), 1-2.

arguments (see footnote¹ for two examples), they are quick to dismiss the whole subject by saying something in the spirit of “Oh, well, I obviously already knew that”. The author senses that refuting this misconception, is, in a sense, the correct (or, at least, a healthy) reaction. The students do indeed know that a nascent polypeptide chain coming out of ribosome is neither a two-dimensional chemical formula, nor a one-dimensional string of letters. But this knowledge is obscured by a teaching approach that insists on connecting protein structure and protein folding with protein sequence. To make this clear : I believe that there are three major fallacies in the way the sequence-folding-structure relationships are being taught.

The first fallacy is our insistence to ignore mounting experimental evidence showing that on one hand proteins (like globins) with no detectable sequence similarity can have essentially identical structures, and on the other hand proteins sharing very high sequence identity can have significantly different structures (Kosloff & 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” does sound somewhat difficult to defend. Maybe it is time to

¹ Questions like the two quoted below have been found effective by the author :

(a) “Would you agree that an unfolded polypeptide chain is as three-dimensional an entity as the folded protein ? That what comes out of the ribosome is *not* a one-dimensional entity but a three-dimensional structure ? And that in terms of its physical reality (eg. electrostatics, bond and angle vibration and stretching, non-covalent bond formation and breaking, interaction with solvent, etc), an unfolded protein is as complex as its folded counterpart ?”, and,

(b) “So, would you agree that protein folding only involves a change of conformation of a three-dimensional structure, and not a change of dimensionality ? That we *do not* go from the 1D sequence space to the 3D structure space, but that we have been in three dimensions right from the beginning ?”

substitute the sentence “sequence determines structure” with the sentence “(unfolded) structure determines (folded) structure” ?

The second fallacy that I perceive has to do with our insistence to place the emphasis of teaching protein folding on the changes of protein conformation (unfolded-elongated-random-coil-like chain → folded-compact-stable structure) and not on what really drives protein folding, that is, the interaction energies (and the resulting energy landscape). To use a popular teaching aid as an example, I think that when “toobers” (<http://www.3dmoleculardesigns.com/toobers.php>) are being used to teach protein folding, the emphasis should be placed not on the tube (representing the protein backbone), but on the push pins and their colors (representing the physical properties of the side chains). To put this differently, I believe that teaching protein folding would be much easier if we substituted Figure 1.1 of Branden & Tooze (1999) with Figure 2 of Dinner *et al.* (2000) [to simplify access, these figures are reproduced without permission on the next page of this manuscript].

The last fallacy –at least in the writer's opinion– is that the sheer amount of sequence data together with their continuous everyday usage in modern Molecular Biology and the hype surrounding the massive genomic (and other related -omic) projects, made us think that sequences (and not 42) is the answer to “life, the universe, and everything”. Which brings me back to the opening sentence of this letter :

Biological sequences are an abstraction of an abstraction: In the first level, we substituted the complexity of a proper three-dimensional entity (like an amino acid residue) with a two dimensional chemical formula describing only composition and covalent bonding. In the second abstraction layer, we substituted these chemical formulas with single

alphabet letters. *And then we forgot about it, and started behaving as if sequences do exist, as if this artificial one-dimensionality is real.* Sequence usage became so widespread, that not only we forgot that these one-dimensional strings of letters do not (and never did) exist, but we have started using them for dealing with problems (like protein folding) that by their nature defy this whole 'sequence' abstraction. Maybe, just maybe, we have had more than enough of 'sequences' ?

References

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.

Branden, C. & Tooze, J. (1999). *Introduction to Protein Structure*. Garland Publishing Inc., New York.

Kosloff, M. & Kolodny, R. (2008). Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins* 71: 891-902.

Dinner A.R., Sali A., Smith L.J., Dobson C.M., Karplus M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci.* 25: 331-339.

