JOURNAL OF CHEMICAL EDUCATION

Activity

pubs.acs.org/jchemeduc

Mechanical Modeling and Computer Simulation of Protein Folding

Maxim B. Prigozhin,*[,†] Gregory E. Scott,[‡] and Sharlene Denos*[,§]

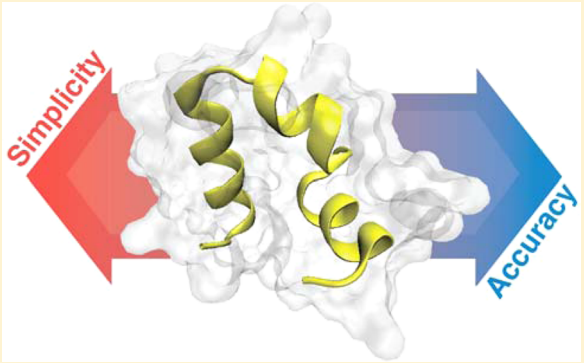
[†]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

[‡]Department of Chemistry and Biochemistry, California Polytechnic State University, San Luis Obispo, California 93407, United States

[§]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: In this activity, science education and modern technology are bridged to teach students at the high school and undergraduate levels about protein folding and to strengthen their model building skills. Students are guided from a textbook picture of a protein as a rigid crystal structure to a more realistic view: proteins are highly dynamic biological molecules in the heterogeneous environment inside the living cell. Simple mechanical models and computer simulations that evolve in their complexity as the instruction progresses are at the core of this lesson. Methods of statistics and physical chemistry (thermodynamics and kinetics) are employed to investigate a flexible version of a peptide that can sample various conformations within the folded ensemble and ultimately unfold into a random coil.



KEYWORDS: *High School/Introductory Chemistry, First-Year Undergraduate/General, Biochemistry, Physical Chemistry, Biophysical Chemistry, Proteins/Peptides, Molecular Modeling, Equilibrium, Kinetics, Molecular Biology*

INTRODUCTION

Simple models are used ubiquitously in science to describe the behavior of complex systems.[1−3] The purpose of simple models is to reduce the dimensionality of a problem so that general trends can be deduced and the behavior of similar systems can be predicted.[4] A scientific model faces the issue of the balance between simplicity and accuracy. At one extreme, the model can account for nearly every detail and degree of freedom. Although such a description will yield accurate results for specific systems, the large number of parameters can make the model inefficient and difficult to conceptualize. At the other extreme, the model is oversimplified and does not provide accurate results. Thus, the spectrum of scientific models varies in its predictive power and utility as a function of coarse-graining, with the most functional and reliable ones being somewhere in between the two extreme scenarios (Figure 1).

In introductory chemistry, many of the models to which students are introduced reside at the "simple but inaccurate" end of the spectrum. Although such models are useful for conceptual understanding of the phenomena that they describe, their use must be accompanied by a discussion of their limitations. Students often believe that models are reflections of reality, rather than having the more nuanced understanding that models are tools for explaining phenomena and have variable conceptual benefits and weaknesses. This view does not improve simply as a function of more educational experience and students need the opportunity to apply multiple models to the same system in order to develop a more sophisticated understanding of the use of scientific models.[5,6]

This problem was approached by first introducing an asymptotic case of a very simple model and then systematically increasing its complexity to make it more realistic. This interplay between accuracy and simplicity, as well as strategies for balancing the two, is at the core of this activity.

THE MODEL

The protein folding problem serves as a background concept for this activity. This problem constitutes a broad field of research in macromolecular chemistry, which contains all the aspects of thermodynamics and kinetics that a traditional "heat-engine-type" physical chemistry class would include. This area of biophysical chemistry is so rich that entire physical chemistry classes for biochemistry majors based solely on the protein folding literature were proposed[7] and stand-alone lessons on protein-folding can cater to the interests of students interested in pursuing health-related careers.[8,9]

Mechanical Prototype

Students use mechanical prototypes of a 10-amino-acid peptide whose folded structure approximates that of α-Synuclein in that it is half unstructured, half α-helical (Figure 2A). The prototype for the unfolded states of this model peptide consists of 9 paper clips connected together by flexible loops. The prototype for

dx.doi.org/10.1021/ed400719c | J. Chem. Educ. XXXX, XXX, XXX−XXX

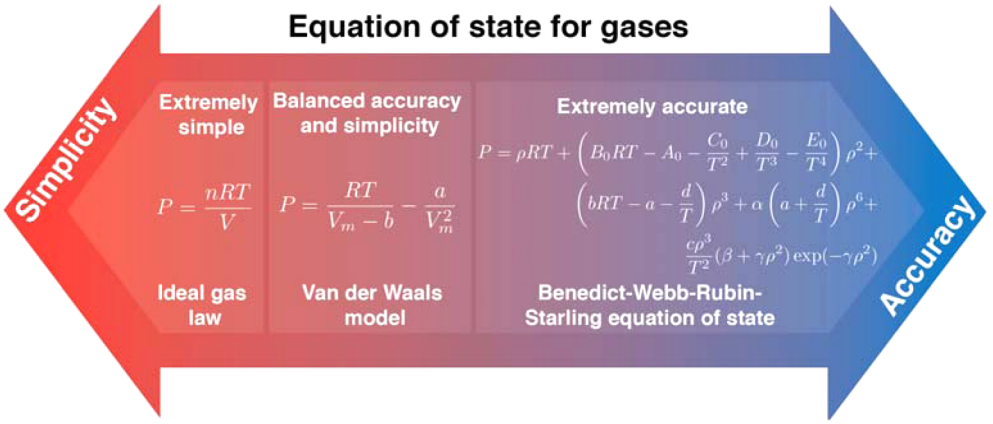


Figure 1. Equations of state for gases as a function of simplicity and accuracy of each model. The simplest model is the ideal gas law (on the left), while one of the more accurate ones is the Benedict–Webb–Rubin–Starling (BWRS) equation of state (on the right). The two extremes are balanced by introducing two (instead of 12 in the BWRS equation!) extra parameters into the ideal gas law, which results in a formula known as a van der Waals model (in the middle).

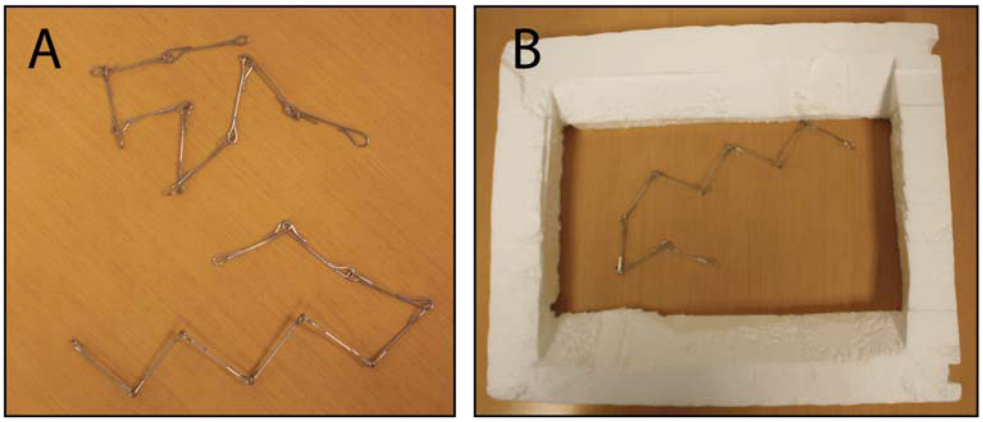


Figure 2. A two-dimensional mechanical prototype of a 10 amino-acid peptide made with paper clips. (A) The folded peptide contains a rigid zigzag created by soldering 4 joints on one side. (B) A Styrofoam box is used to confine the peptide prototype to a rectangular area to approximate cellular crowding.

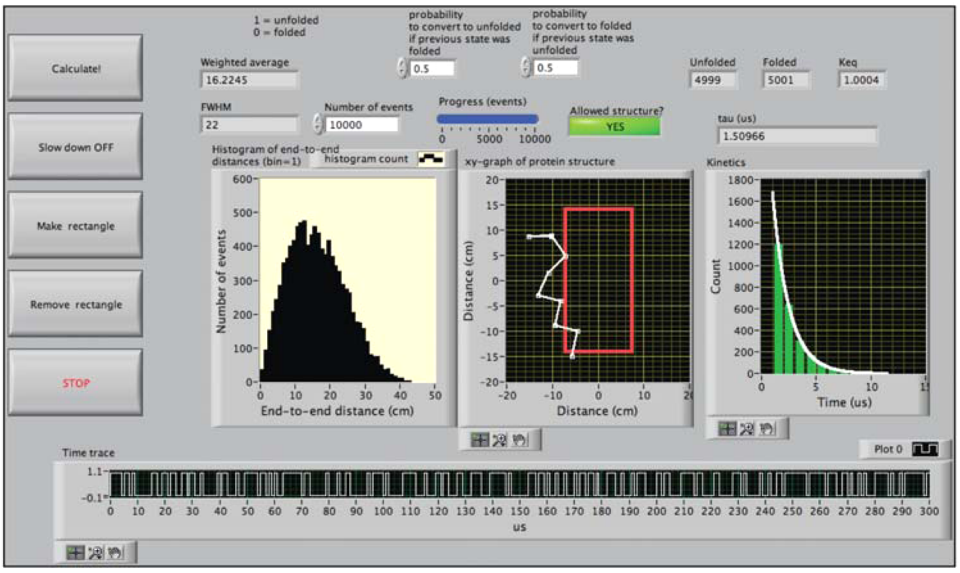


Figure 3. User interface of the computer program used to model the behavior of the prototypical peptide. As inputs, the program requires the number of samplings, the probability to switch to the unfolded state if the current state is folded, and the conjugate probability going in the other direction. The outputs include a series of folded and/or unfolded peptide conformations, a histogram of their end-to-end distances, the center of mass of the histogram, and a plot of dwell times in each state, as well as the equilibrium constant and the folding rate.

dx.doi.org/10.1021/ed400719c | J. Chem. Educ. XXXX, XXX, XXX–XXX

the folded peptide is also made of 9 paper clips, but with 4 of the joints soldered so that half the peptide is fixed in a rigid zigzag, a proxy for a two-dimensional α-helix. Each joint between the 5 cm-long connectors represents an amino acid. To investigate the two-dimensional conformational space that these two prototypical peptides would explore, the end-to-end distance is used as a reaction coordinate. End-to-end distance is a convenient variable to use in this case because it can be quickly measured with a ruler and it is also a common reaction coordinate in real biophysical studies of proteins where dyes are attached to each end of the protein and the distance-dependent energy transfer between them provides the signal.[10] Students throw the peptide prototypes on the desk several times and measure the end-to-end distances for each resulting conformation. The results expected for this peptide are contrary to those expected for the globular proteins most commonly used as instructional examples, as this peptide's average end-to-end distance will increase when the peptide is folded.

Students tabulate the resulting end-to-end distances and build a histogram of the number of occurrences vs end-to-end distance. This exercise sets the stage to discuss the physical concepts behind protein folding, including entropy, enthalpy, and the Gibbs free energy, but also to relate macroscopic measurements to the statistics of biophysical systems.

Students also consider protein folding inside a cell. The concentration of macromolecules confined to the volume of a typical *Escherichia coli* cell can be on the order of 300−400 g/L.[11] These molecules are sterically hindered from sampling the conformations that occupy large extended areas of space, i.e., the conformations that have large end-to-end distance. Students use a simple model for this crowding: a rectangular boundary made out of Styrofoam (Figure 2B). The Styrofoam frame serves as a spatial constraint preventing the peptide from extending, decreasing its conformational entropy.

Computer Simulations

There is ample evidence that computer simulations can have a positive effect on enhancing instruction, suggesting that technology is rapidly becoming a vital part of science education.[12,13] Simulations have been used previously to teach macromolecular structure[14] and kinetics;[15] however, to our knowledge, this is the first report where a computer simulation is specifically designed to reflect both structural and kinetic aspects of the mechanical model in question.

Students use a simple computer program (Figure 3) to simulate various structures that the paper clip peptide model could form, to analyze these structures systematically, and to compare experimental results with simulation (see Supporting Information). The computer model mimics the mechanical one and computes end-to-end distances in the same way. Because the program can run through thousands of iterations in a matter of seconds, students can generate histograms that are much smoother than the ones obtained manually, which provides an opportunity to elaborate on the importance of statistical sampling. In addition to the end-to-end distance statistics, the program also computes an equilibrium constant and kinetic time constant based on student input on the probability of the peptide to switch between the two states. The simulation provides an opportunity to reinforce the idea that equilibria and kinetics are governed by statistical principles. It is important to distinguish that, per Levinthal's paradox,[16] protein folding cannot occur through a random sampling of conformations. This provides an opportunity to discuss the limitations of the model and can lead to follow-up discussions about many thermodynamic and kinetic ideas appropriate to the course level, which may include the relationships between ΔG, activation barriers, and equilibrium distributions based on free energy surfaces.

CONCLUSIONS

This activity provides students with an opportunity to use an increasingly complex model to sample different equilibrium states for a peptide. It was piloted in one, 3-h laboratory section of 24 students in a first-year undergraduate general chemistry course that focuses on thermodynamics and kinetics. The lesson was also done at the high school level in a class of 20 students. Students in both settings enjoyed the activity and were surprised to learn that the same kinetics and thermodynamics concepts they had been applying to small molecules could also be applied to proteins, which they had mostly learned about in descriptive terms in biology classes.

ASSOCIATED CONTENT

Supporting Information

The simulation program, additional background, and instructions for teachers and students. This material is available via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

* E-mail: m.prigozhin@gmail.com.
* E-mail: denos@illinois.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (1430124). The authors are grateful to David Bergandine at University High School in Urbana, Illinois for allowing us to pilot this activity with his students. At the time when this work was performed, M.B.P. was a Howard Hughes Medical Institute International Student Research Fellow.

REFERENCES

(1) Coll, R. K.; France, B.; Taylor, I. The Role of Models/and Analogies in Science Education: Implications from Research. *Int. J. Sci. Educ.* **2005**, *27*, 183−198.

(2) Harrison, A. G.; Treagust, D. F. A Typology of School Science Models. *Int. J. Sci. Educ.* **2000**, *22*, 1011−1026.

(3) Gobert, J. D.; Buckley, B. C. Introduction to Model-Based Teaching and Learning in Science Education. *Int. J. Sci. Educ.* **2000**, *22*, 891−894.

(4) Justi, R. S.; Gilbert, J. K. Modelling, Teachers' Views on the Nature of Modelling, and Implications for the Education of Modellers. *Int. J. Sci. Educ.* **2002**, *24*, 369−387.

(5) Grosslight, L.; Unger, C.; Jay, E.; Smith, C. L. Understanding Models and Their Use in Science: Conceptions of Middle and High School Students and Experts. *J. Res. Sci. Teach.* **1991**, *28*, 799−822.

(6) Harrison, A. G.; Treagust, D. F. Modelling in Science Lessons: Are There Better Ways To Learn with Models? *Sch. Sci. Math.* **1998**, *98*, 420−429.

(7) Anthony-Cahill, S. Using the Protein Folding Literature To Teach Biophysical Chemistry to Undergraduates. *Biochem. Mol. Biol. Educ.* **2001**, *29*, 45−49.

(8) Bowen, R.; Hartung, R.; Gindt, Y. M. A Simple Protein Purification and Folding Experiment for General Chemistry Laboratory. *J. Chem. Educ.* **2000**, *77*, 1456−1457.
(9) Smith, C. A. How Do Proteins Fold? *Biochem. Educ.* **2000**, *28*, 76−79.
(10) Forster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **1948**, *2*, 55−75.
(11) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **1991**, *222*, 599−620.
(12) Rutten, N.; van Joolingen, W. R.; van der Veen, J. T. The Learning Effects of Computer Simulations in Science Education. *Comput. Educ.* **2012**, *58*, 136−153.
(13) Linn, M. C. Technology and Science Education: Starting Points, Research Programs, and Trends. *Int. J. Sci. Educ.* **2003**, *25*, 727−758.
(14) White, B.; Kahriman, A.; Luberice, L.; Idleh, F. Evaluation of Software for Introducing Protein Structure: Visualization and Simulation. *Biochem Mol. Biol. Educ* **2010**, *38*, 284−289.
(15) Gonzalez-Cruz, J.; Rodriguez-Sotres, R.; Rodriguez-Penagos, M. On the Convenience of Using a Computer Simulation to Teach Enzyme Kinetics to Undergraduate Students with Biological Chemistry-Related Curricula. *Biochem. Mol. Biol. Educ.* **2003**, *31*, 93−101.
(16) Levinthal, C. How to Fold Graciously. In *Mossbauer Spectroscopy in Biological Systems*; Proceedings of a Meeting held at Allerton House, Monticello, Illinois; Debrunner, P.; Tsibris, J. C. M.; Münck, E., Eds.; University of Illinois Press: Urbana, IL, 1969.