# FROM GENETIC CODE TO PROTEIN SHAPE USING DYNAMIC MODELING

Boris Berenfeld, Amy Pallant, Barbara Tinker, Robert Tinker, Qian Xie

The Concord Consortium
10 Concord Crossing, suite 300,
Concord Massachusetts 01742
Contact email" **apallant@concord.org**

## Abstract

Computational models engaging students in exploration of the interactions between atoms and molecules can illuminate topics in high school biology curriculum heretofore left for later years. Atomic and molecular-scale processes are typically encountered in either advanced high school biology course or in college courses, after a student has acquired a grounding of chemistry and physics. Understanding fundamental molecular concepts earlier in a science sequence is important, as many central ideas in modern biology are hard to acquire without mental models of critical molecular-level phenomena. Two such central concepts in biology are the flow of information from DNA to proteins, and the structure-function relationship in proteins.

Dynamic computational models offer the possibility of providing molecular underpinning to crucial biology concepts without requiring the mathematical formalism used in a traditional chemistry and physics course.  Such models are easily explored by students at their own pace, and can be set up to provide vital feedback to teachers about the state of student understanding.

An educational module, *From DNA to Protein Conformation*, was developed for high school biology students.  It consists of a set of four scaffolded learning activities, each using a dynamic model of protein coding and shaping. An underlying molecular engine was built to generate the models.  Finally, research was performed on student learning of key concepts in this content domain.

The materials were tested in six classrooms drawn from three suburban high schools in Massachusetts. In all classes analyzed, students scored higher on a post-test than they did on the pre-test. A paired T-test revealed that the post-test scores for content knowledge were significantly greater than their pre-test score in six classes, four at the $p. =< .001$ level.  While the suburban nature of the classrooms limits the full applicability of the findings, it appears that atomic-scale dynamic modeling can help high school students understand critical concepts in biology.

**Introduction**

The central concept of modern molecular biology (sometimes called the Central Dogma) is that genetic information flows from DNA to RNA to proteins, and that, in a polypeptide chain, the sequence of amino acids determines a protein's structure that in turn defines its biological function[1]. While there is currently an abundance of high-level pictures, diagrams and even colorful cartoon animations available on the Web and on CDs accompanying textbooks to show the transcription and translation processes, and how the polysome assembly line makes polypeptide chains, there has been less effort to explicate, let alone model interactively, the relationship between the genetic code in DNA and *the shape* of the proteins for which it codes.

Good high school biology students are typically exposed to the concept that "the order of A, T, C & G bases within the coding sequence (gene) determines the order of amino acids in a protein and that protein structure determines its function." But the most pedagogically challenging aspect of this content domain is to explain how the primary structure of a protein—its amino acid sequence—determines the shape of a protein molecule and its function.

Connecting the primary structure of a protein to its structure and function is a central endeavor in frontier biology: the new field of proteomics. The computers now available to introductory students have some of the capacity of software used in this research. The educational question is whether this power can be harnessed to give students a conceptual understanding of the associated concepts.

There are impressive visualization tools (viewers) available that allow students to explore colorful three-dimensional biological molecules. Yet, visualization tools do reveal how complex 3D forms can be made "by itself." Rendering 3D images cannot communicate the dynamic processes underlying protein folding, which includes complex interactions among amino acids in the polypeptide chain and between them and their surrounding medium, either water or lipids.

That a linear polypeptide chain may contain all information necessary to create a specific 3D conformation may be a counterintuitive idea. The challenge of the model-based curriculum described in this paper, one of a set of units in the NSF-funded Molecular Workbench project, has been to provide students with a manipulative, dynamic tool that could develop their intuition about the most basic forces determining protein folding, and could make this unusual molecular world of dynamically shaping polypeptide chains both engaging and relevant to high school biology.

**The Case for Dynamic Modeling in the Biology Classroom.**

Modeling software that is sufficiently flexible and thoughtfully applied has been shown to result in deep learning through guided inquiry (Feurtzeig & Roberts, 1999; Tinker &

---

[1]  Of course, there are exceptions such as RNA viruses and, possibly, prions, but beginning students need the rule before being exposed to the exceptions.

Horwitz, 2000). Models and modeling are considered essential components of scientific literacy (Gilbert et al, 1998a; 1998b; Gilbert, 1991; Linn & Muilenberg, 1996; Perkins, 1995). Interactivity creates the rich, highly associated mental network of concepts that we know contributes to lasting understanding (Jackson, Stratford, Krajcik, & Soloway, 1996; Snir & Smith, 1995). Interactive models can address core ideas in a visually engaging way that makes them accessible to students with vastly different learning styles. As students work with the models, changing key variables, they are more likely to remember and transfer their learning to new situations. Model-based inquiry provides an alternative way to acquire enough theoretical understanding to be effective in new situations without having to wade through the usual mathematical treatments.

A molecular-modeling approach to biology would develop in students an understanding that biological processes are based on specific physical properties of molecules, their interactions and the consequences of these physical interactions. Students need to understand that atoms are discrete and that interactions between them lead to bonding, that energy can be exchanged through interactions and that atoms and molecules move in random motion. These ideas are fundamental for understanding how the molecular world determines biological phenomena such as diffusion, osmosis, and three-dimensional structures of biomolecules.

Complex behavior like protein folding, furthermore, happens because of physical forces. If you take the time to understand these forces and how they play out, the world seems more rational and the learning can be applied in other areas as well, such as solvation, etc. As a result, there is less to learn and more of a sense of mastery.

A rule-based model differs from a visualization, such as a animation or  molecular viewer. Both may support interactions, largely simple manipulations, but the quality of those interactions with a model is different. (Horwitz & Barowy, 1994)  You can't for example, find out how an animation works and experiment with it.


Dynamic computation models, in particular, are ideal for observing interactions between atoms and molecules. Studies (Pallant and Tinker 2004, Nakeleh 1993) have shown that students who have the ability to visualize chemical and physical phenomena at the molecular dynamic level develop enduring mental models of critical macroscopic phenomena. Well-designed learning opportunities that use molecular dynamic models could help students develop these understandings.

Atomic and molecular structures, together with the interactions between molecules, are often found briefly in high school biology textbooks in a chapter named something like "Chemistry of Life."  Since ninth- and tenth-grade biology students have typically had neither physics nor chemistry, topics related to molecular structures and molecular interactions have been considered too complex for extensive use. Dynamic processes underlying protein folding or self-assembly are visually represented, perforce, by static diagrams. Yet learning by interacting with molecular dynamic tools can help students develop more scientifically accurate mental models of atomic scale phenomena  (Chi et

al. 1988), which should, in turn, help them to reason more effectively across the disciplines, including in biology.

Little research has been done on how understanding molecular dynamics modeling can enhance student understanding of biology. Canning and Cox (2001) reports that college students involved in 3-D molecular visualization of critical biological molecules have made a tremendous impact on student understanding of biochemistry and biology. Their emphasis, however, was on the viewers and not on dynamic computational models.

There have been, to be sure, significant challenges (Rieber, 1989; Tversky, Morrison, & Bétrancourt, 2002) to the use of molecular visualizations in science teaching.  Rieber has argued that in order for an animation to be a significant learning tool, it needs to support closely the theme of the model involving motion, e.g. trajectories. Tversky, Morrison, and Betrancourt maintain that studies supporting animation have not been sufficiently controlled, as animations afford more information to be delivered in a shorter amount of time.  This line of reasoning raises the question of whether any experimentally controlled condition assessing dynamic models could be designed. Tversky et al. also find that students have perceptual problems with animations that they cannot control, and are consequently easily overloaded.

Dynamic molecular models meet Rieber's criteria, as they deal with a world in continuous motion, in which the force, energy and motion of particles is the central thematic underpinning of science. The ability of such models to convey considerable information, including data on transitional states, is quite possibly a great advantage.


## RESEARCH DESIGN AND METHODS

A single educational module, *From DNA to Protein Conformation*, was developed in order to deliver to high school biology students in a scaffolded way several computational models allowing users to manipulate with the genetic code for a protein and explore in real time how the changes in the code affect the shape of a protein. An underlying molecular dynamics software application controls the motion of all objects in the system, based on approximations to the applicable physical laws. Finally, research was performed on student learning of key concepts in this content domain.

The research goal of the module was to enhance our understanding of how students reason about the molecular interactions underpinning protein folding and about the molecular processes allowing genetic control of the tertiary structure and function of proteins. The module requires four to seven class periods and combines dynamic atomic-scale computational model-based activities with kinesthetic activities and a discussion of a case study.

In the model, students can control (i) the amino acid sequence in a protein, (ii) the surrounding media (oil, water or vacuum), and (iii) the sequence of the nucleotides in the corresponding string of DNA. The string of the DNA is linked to the string of amino

acids with the genetic code. It allows students to explore interactively not only the correspondence between the sequence of nucleotides in a strand of DNA and sequence of the amino acids in a protein, to observe in real time how changes they make in the genetic code affect both the sequence of amino acids and the shape of the model protein (Figure 1 and Figure 2). These real time experimentations and observations go far beyond drill and practice with the genetic code.

## The Technology: Modeling from DNA to Protein

An innovative part of our model is to show a correlation between genetic codes and protein structures. To demonstrate this correlation, we have built a simplified two-dimensional molecular dynamics program that approximately models protein folding in different environments. Building a computer model capable of predicting the structure of a protein based on its genetic code is an important topic of current computational molecular biology research (Jiang, Xu, & Zhang 2002). Research on protein structure prediction focuses primarily on the mechanisms of protein folding. Because of the fundamental importance of the protein-folding problem, researchers have built numerous mathematical and computational models. We have examined many of these models. Some of them can be called simple exact models (Chan, Bornberg-Bauer, 2002), in which proteins are represented by a string of beads that occupy the vertices of a lattice.  Folding is described as a process of seeking minimum energy for the string on the lattice. The most popular lattice model is perhaps the hydrophobic-polar (HP) model (Lau and Dill, 1989), which is a two-state model in two dimensions.

While simple exact models may be useful for investigating folding mechanisms from basic physical principles, they usually have a form too abstract to be integrated into high school curriculum. For example, in a two-state HP model, one cannot talk about specific amino acids and therefore cannot introduce genetic codes. We needed a 20-state model so that we could map each state to an amino acid, and then map the DNA codons to the 20 amino acids. Furthermore, in most lattice models, confining the protein on a lattice is just an approximation to reduce the volume of the phase space of protein folding. As enumerating all possible states is not important in our educational modeling, there is no need to introduce the lattice approximation. An amino acid should have the freedom permitted by the force laws and boundary conditions to move anywhere in the space.

In the HP model, the evolution of the lattice protein over time is typically produced by using a Monte Carlo method in conjunction with a scoring function that is, for example, characterized by the number of hydrophobic contacts. A simple rule of thumb is that the more hydrophobic contacts are formed, the more stable the conformation is. Such an approach, however, does not address the contributions of other interatomic and intermolecular forces, such as van der Waals interactions, chemical bonds, electrostatic interactions, and hydrogen bonds. As a result, it does not connect to our existing curriculum about molecular interactions.

To address these problems, we have built a simple protein folding model based on our molecular dynamics engine. Our model is based on the following considerations:

Size, charge and polarity of amino acids

For simplicity, an amino acid is represented by a particle. A particle is modeled by a Lennard-Jones potential, which is attractive, but prevents other particles from getting too close to it. This reflects the fact that amino acids are stable chemical compounds that do not easily react to form a new compound. The sigma parameter of the Lennard-Jones potential also represents the size of the amino acid.

Peptide bonds

Two adjacent amino acids are linked by a spring, which represents a peptide bond. A sequence of amino acids connected by springs represents a polypeptide.

Van der Waals and electrostatic interactions between amino acids

Amino acids weakly attract each other with van der Waals forces (i.e. the negative part of the Lennard-Jones potential). Charged amino acids interact with each other with long-range electrostatic forces, which are greatly reduced when a protein is put in aqueous environment because of the electrostatic screening effect caused by the surrounding water molecules (this is modeled by decreasing the dielectric constant used in computing the electrostatic forces).

Interactions between amino acids and solvent molecules

Following common practice, solvent molecules are not explicitly taken into account in our model, because to do so would be computationally prohibitive. An effective-field approach is used to approximate the interactions of amino acids with solvent molecules.

First, we introduce a parameter h to describe hydrophobicity. A positive h value means the amino acid is hydrophobic, whereas a negative one means it is hydrophilic. The more positive the h value is, the more hydrophobic the amino acid is. Conversely, the more negative the h value is, the more hydrophilic the amino acid is. Second, we introduce a parameter E to describe the chemical property of solvent. A positive E means that the environment is oily; a negative one means that the environment is aqueous; and a zero value means that the protein is put in a vacuum. The value of E is affected by such factors as temperature. For an aqueous solution, a greater absolute value of E represents a better solubility for the hydrophilic residues.

Having defined these properties, we assume that the force an amino acid feels from the solvent molecules is given by:

$$h_i E \frac{\mathbf{R}_i - \mathbf{R}_c}{|\mathbf{R}_i - \mathbf{R}_c|}$$

where $\mathbf{R}_i$ is the position vector of amino acid i, and $\mathbf{R}_c$ is that of the center of mass of the protein. This simple assumption is based on the fact that a hydrophobic amino acid in water tends to shy away as much as possible from water molecules. If water molecules surround the protein from all directions, the interactions between them result in a net effect as if the hydrophobic amino acids were pulled by an effective force that helps them hide inside the protein. As a zeroth-order approximation, we assume that this effective force comes from the center of mass of the protein (Fig. 3a). This approximation works well for proteins that will fold into globular structures, though it needs to be improved to handle other situations.

For the hydrophilic amino acids that "love" water, the effect of the environment is such that they are pushed away from the center of mass, so that they can have a better chance to contact with the water molecules (Fig. 3b).
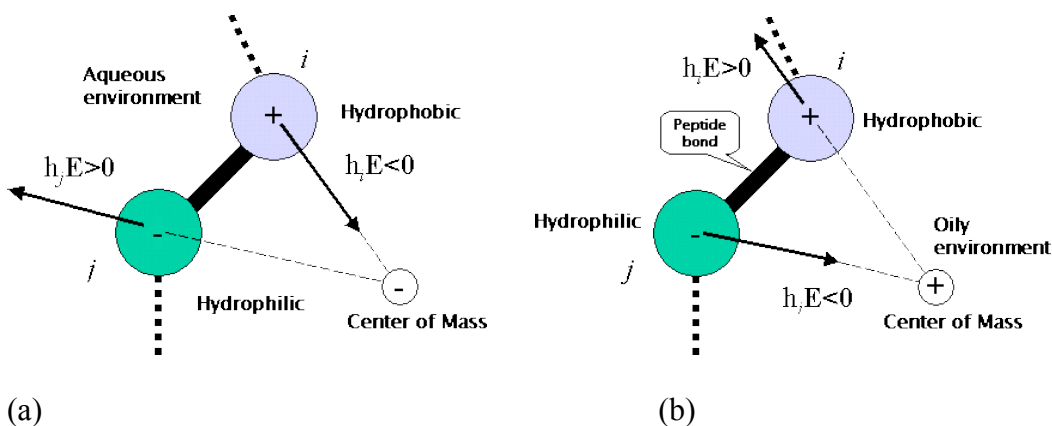


(a)                                                                 (b)

*Fig. 3. A schematic illustration of hydrophobic and hydrophilic interactions.*

The dynamic model based on these force rules and the molecular dynamics simulation engine of the Molecular Workbench software is able to produce qualitatively correct results. Fig. 4 shows a folded conformation of a protein with 94 residues in aqueous solution, started from an initial conformation generated by a self-avoided random walk algorithm. You can tell from the figure that the hydrophilic residues (cyan beads) tend to stay at the periphery, whereas the hydrophobic residues (pink beads) clump in the middle. But not all residues follow this tendency. Some hydrophilic residues are buried; some hydrophobic residues are exposed to water.
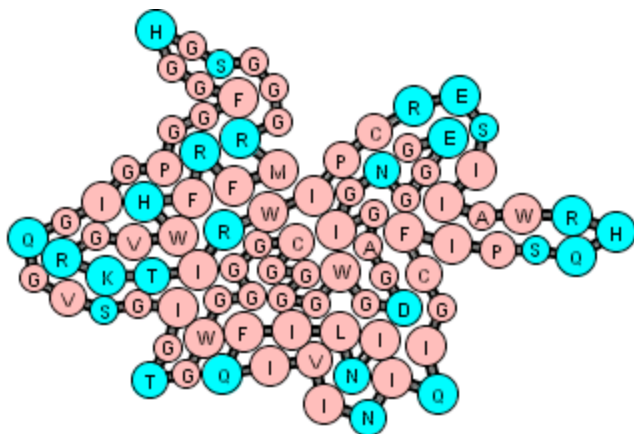
*Fig. 4. A folded conformation of a 94-residue protein in aqueous solution, started from a self-avoided random walk conformation. The pink beads represent hydrophobic amino acids. The cyan beads represent the hydrophilic amino acids. The black lines between two adjacent beads stand for the peptide bonds.*

Beginning students need to start with simple models with just a few options as described above. The research reported here used these *Molecular Workbench* models, which allow the user to specify a DNA sequence controlling the sequence of amino acids. The acids have two properties that determine the shape of the protein: charge and repulsion from water. The resulting model does not predict actual protein conformations—this is still an open research question-- but it does illustrate how temperature, charge, and water interactions can shape proteins, depending on the sequence of the amino acids.

**The Instructional Material**

To draw students into the study of DNA and its connection to protein structure and function and to increase the relevancy to real life, the module *From DNA to Protein Conformation* engages students in an exploration of the relationship between the macroscopic-level observable symptoms of the disease Sickle Cell Anemia to the microscopic-level phenomenon of affecting the protein (hemoglobin) shape. To do this, we expected students to reason about the forces and interactions that underpin protein folding and reason about mutation that leads to the disease and the overall genetic control of the tertiary structure and function of proteins.

The activities in the unit are organized into the following sequence:
- *Activity One: How a Protein Gets its Shape: Charged Amino Acids.* Students learn about amino acids, their structure and their charge, discovering how differently charged amino acids affect the shaping of a dynamic model of a protein fragment. They are introduced to the case of Sickle Cell Anemia, and begin their exploration of its molecular mechanism.
- *Activity two: How a Protein Gets its Shape: Interacting with Water or Lipids.* Students expand their mental model of protein folding from that of  simple attraction or repulsions between  cationic and anionic amino acids, to the more complex interactions involving all amino acids (polar and non-polar) with each

other and with the surrounding medium, water or lipids. Students work with the dynamic model making changes in its primary structure or replacing water with lipids to construct differently shaped proteins, and learn to reason about their shapes, considering the property of specific amino acids and their position in the chain and properties of the surrounding medium.

- *Activity Three: How a Protein Gets its Shape: The Role of DNA*. Students manipulate a sequence of nucleotides in DNA and explore connections between the genetic code, written as a sequence of nucleotides in DNA, the sequence of amino acids and the shape of a protein generated by the dynamic model.
- *Activity Four: Protein Malfunction and Disease*. Students compare the effect on the protein shape of two different types of mutations in DNA, one caused by a nucleotide substitution and another caused by a nucleotide deletion. Students bring what they have learned in their investigation of the coding of proteins and the relationship between protein's structure and function to the case of Sickle Cell Anemia. They reproduce a critical piece of hemoglobin structure to explore the mutation responsible for the changes in the tertiary structure of hemoglobin in a patient with Sickle Cell disease.

In summary, throughout the entire module, students work with models and are expected to apply their understanding to the cause of the Sickle Cell Disease. Students can modify the amino acid sequence directly, or indirectly by substituting or deleting nucleotides in the DNA and then observe how substitutions or deletions affect the sequence of amino acids and ultimately might affect the spatial organization of the protein.


**Classroom Studies**

Our research study looked at the ways that the Molecular Workbench modeling environment, with its interactivity, can increase cognitive competencies. Can students reason effectively about the microscopic world of dynamically interacting amino acids? Can they reason about the molecualr processes allowing DNA to affect the shape and the function of a protein, and how a mutation in DNA could have implications  for both, protein function and human health?

The module *From DNA to Protein Conformation* had its first iteration as a section of another module that was implemented in the spring of 2003. In this module, students explored the world of very large macromolecules, whose size and shape are critical for their ability to perform specific biological functions. Students constructed polymers from monomers and then, using dynamic models, explored how polymeric molecules fold into specific shapes. They learned what causes a protein to lose its ability to function, and about molecular diseases associated with changes in structure.  This pilot test, which we will refer to as *From DNA to Protein Conformation* Version A, enabled us to refine the software design, curriculum materials and the student assesment protocols. The resulting module, which does not include the materials on the macromolecules (these became a separate module on Macromolecules)  was then used in a field test during the winter of 2004.

Six preliminary classroom studies were conducted—two using the pilot version and the remaining classes using the field test module. The materials were tested in three suburban high schools in Massachusetts. These research studies were done in the following 10th grade biology courses: AP biology, College Prep biology, Biology1, Standard level and Skill building biology with an average size of 20. Class periods ranged from 45 minutes in length to 90 minutes. The regular science teacher administered the teaching intervention. Eight laptop computers with the software included were used in two classrooms, a computer lab of 12 computers was used in another classroom and the remaining classes were in a computer lab where there were enough computers to have one student per computer.

Learning goals

In each of the *From DNA to Protein Conformation* activities described above, students explored how a protein shape is related to (1) the properties of the amino acids, (2) the sequence of the amino acids in its chain, and (3) the interactions between all the amino acids (charged, polar and non-polar) with each other and with the surrounding water molecules or lipids. By directly experimenting with sequence, properties of amino acids and medium and observing the resulting conformations students are given the opportunity to develop their own conception of the forces that influence protein shape.

Students were expected to be able to:
- Explain how charged amino acids within a protein chain relate to protein's three-dimensional (tertiary) structure;
- Characterize amino acids according to their response to water or lipids;
- Explain how interactions with water molecules or lipids affect the shape of a protein;
- Explore the way the sequence of DNA dictates the shape of a protein;
- Reason about the molecular origin of some diseases;
- Relate the change in the structure of proteins to changes in their function and possible implications for human health.

**RESULTS**

Interviews

Interviews with a subset of students a month after their encounters with the materials indicated that the models remained salient.  They were able to identify and discuss model representation and underlying content.

Pre-Post Test Comparisons

An examination of the effectiveness of the *From DNA to Protein Conformation* module was conducted using pre- and post-tests. The pre- and post-test questions were designed to cover content in which students were asked to reason about the interactions of molecules that underpin the biological phenomena of protein folding and reason about the genetic control of the protein structure. The pre- and post-tests consisted of three questions, in which each question had several parts. Students were asked to give brief

answers and draw diagrams. The tests did not count toward students' grades. Student answers were coded for correct content as well as for the degree to which reasoning related to the phenomena was due to the intermolecular interactions. The statistics were done using StaView® (1999).

Paired group t-tests were performed on module pre- and post-test scores for each class The t-test revealed students' post-test mean performance was significantly greater than their pre-test mean performance in each of the six classes.

| Version of module pre-post test analyzed | Class | Mean Diff. | DF | t-value | P-value |
|---|---|---|---|---|---|
| Version A | 1 | 3.875 | 20 | 5.725 | <.0001 |
| Version A | 2 | 2.75 | 19 | 3.809 | <.0012 |
| Version B | 3 | 2.909 | 22 | 6.202 | <.0001 |
| Version B | 4 | 4.731 | 13 | 9.930 | <.0001 |
| Version B | 5 | 4.47 | 9 | 5.449 | <.0028 |
| Version B | 6 | 2.909 | 11 | 3.479 | <.0059 |

Student Reasoning about the Properties of Amino Acids

Within our data set, we focused more specifically on the degree to which student reasoning about the shape of a protein was based on the properties of the amino acids and their interactions with each other and the environment.

On the pre- and post-test students were asked the following questions:
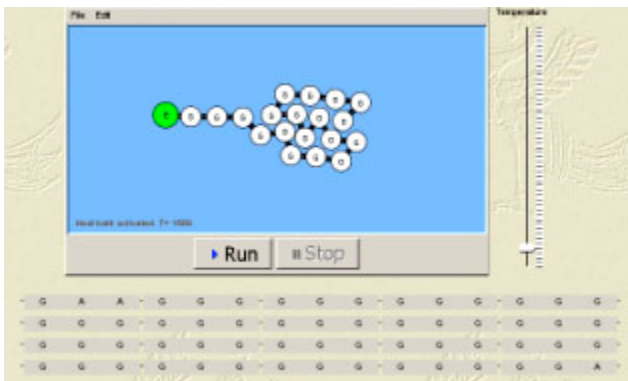1a. The model below represents a protein surrounded by water molecules. It curls into a ball like this.
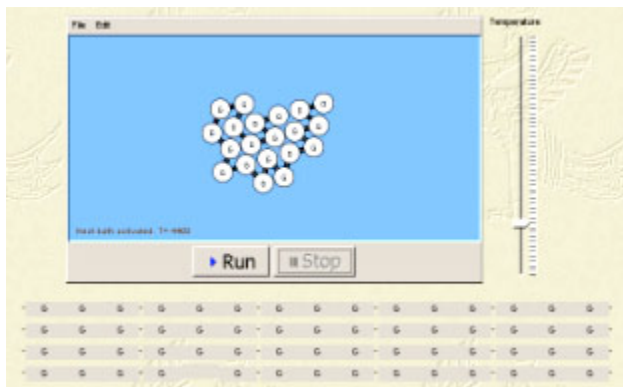


    Notice that each white circle represents the amino acid glycine. Why do you think the protein folds this way?

1b. Scientists substituted the first amino acid (glycine, G) in the chain with another one (E, glutamic acid). Look what happened to the protein (below)[2]:



Why do you think this single replacement of an amino acid affected the shape of the protein chain the way it did?

The responses were analyzed for the inclusion of references to amino acid properties such as charge or hydrophobicity in student reasoning. In coding for this, all reference to properties were acknowledged, whether or not the content was accurate. So, for example if the student said that glutamic acid had a positive charge and therefore was attracted to water students were given some credit for reasoning about the shape of proteins based on amino acid properties.

On the pre-test, 28% of the students from the field test wrote responses that indicated they thought the properties of the amino acids were responsible for the shape they observed whereas on the post-test 94.5% included reference to such properties in the responses. The majority of students in the post-test did include charge and hydrophobicity in their answers but did not include in their explanation precisely "the mechanism of" *how* these properties influenced the interactions with the environment and with other amino acids.

The following are examples to show the range of the student responses to the questions. Student A post-test response to question 1A (Why does the protein fragment of glycines fold in the way it did?):
*Because the amino acids all have a neutral charge.*
Student B post-test response to question 1A:
*The protein most likely folds this way because the amino acids are hydrophobic meaning they are trying to get away from water.*
Student C post-test response to question 1B:*Because it [glutamic acid] likes water and maybe its charge impacted the shape.*
Student D post-test response to question 1B (Why does a single replacement of glycine with glutamic acid affect the shape of the protein the way it did?):

---

[2] The degree of the extension into water relates to the assumptions built into the model.  We made the decision that it was more important for students to see that there was an effect than to worry about perfect model accuracy.

*All of them are hydrophobic except the charged amino , which is hydrophilic. Therefore it pulls that part of the chain away.*
Student E post-test response to question 1B:
*The glutamic acid is hydrophilic so it tries to string out and get the most contact with water.*
These initial findings suggest that we have successfully created a learning environment in which students analyze protein structures based on the properties of the amino acids in the chain. Student explorations of these models appear to lead to some understanding of the interactions of the properties with the molecules in the environment and the surrounding, however, it also appears that the materials need to be revised in order to engage students more with the ways these same properties influence the interactions of the amino acids with each other and the molecules around them.

<u>Student reasoning about the relationship of the genetic code to the shape of proteins</u>

One measure of transfer was obtained by providing students on the post-test with a case history of a genetic disease, Cystic Fibrosis, which they had not studied prior to the test. Student responses were examined to determine the degree to which students connected the genetic code with the sequence of amino acids in a protein and with its shape and, in this case, its function. The question read as follows:

> There is a dangerous human disease, called Cystic Fibrosis that affects various organs, making it hard to breathe. It often shortens peoples' lifespan. It is known that Cystic Fibrosis is an inherited disease, that is, one that can be passed on from parents to children.
>
> Trying to uncover the cause of the disease and seeking a cure, scientists noticed that those with this disease have very salty (chloride-filled) sweat. Scientists isolated a protein from the skin of healthy individuals and found that it works like a control valve, regulating how much chloride goes out with the sweat. People with Cystic Fibrosis are missing that protein valve. Scientists discovered that, because the protein was misshapen, it could not take its place in the cell membrane, where it controls the flow of salt in and out of the cell.
>
> Imagine yourself working with the team of scientists looking for the cure of Cystic Fibrosis. How would you present the connection between the genes (sequence of DNA) and the "salty sweat" symptoms to fellow scientists?
>
> A.  Fill in the two missing boxes below.

Students were prompted to fill in a flow chart that showed the relationship between the inherited gene and the misshapen protein. The data was analyzed to determine whether or

not students revealed that the DNA code was transcribed and translated into a protein, and that the resulting protein contained the wrong sequence of amino acids. The data was examined, further, for signs that students recognized that, because of the genetic fault, the protein folded into the wrong shape, resulting in protein molecule that could not be used to control the flow of salt.

On the pre–test, 70% of the students did see the correlation between the genetic code and the function of the protein, but only 45% of them included the relationship between the sequence of the amino acids and the shape of the protein. In the post-test analysis, 57% of the students illustrated they understood the connection between the DNA code and the shape of the protein in their responses. When analyzed by class, students in the College Prep and AP classes, which had a strong prior understanding of the processes of translation and transcription, were better able to transfer the acquired knowledge to the concept that the shape of the resulting protein was responsible for the cystic fibrosis disease. 80% of the students with this prior knowledge were able to clearly connect the genetic code to the form and function of the protein. A student, for example, would say, "A person inherits a gene of Cystic Fibrosis, the gene is transcribed, it is then translated into a protein that has an incorrect sequence and is misshapen as a result."

The findings indicate that the molecular workbench models can increase student understanding of the relationship between protein form and function and the genetic code. It is possible, however, that for lower level students, the question and flow chart as written on the pre and post-test may have been an obstacle to their comprehension. Other questions on the pre- and post-test that asked directly about the relationship between an insertion and the resulting protein sequence show no discernable differences in student response levels among the different classes.

**Conclusions**
These data indicate that through guided explorations of our models, students can acquire fundamental biological concepts and transfer them to other materials appropriately. They were able to reason about the ways in which the properties of amino acids might affect the shape of a protein. Students regularly employed the ways in which charge and hydrophobicity help describe the interactions of the different amino acids with each other and with the solvent. Additionally, students could correlate how a change in the sequence of DNA could have implications for protein function (but doesn't necessarily have to) and in some cases for human health. In analyzing a case of Cystic Fibrosis, one not previously encountered in the curriculum, student responses indicated that they were able to figure out the importance of the mutation both at the microscopic level (it alters the sequence of the protein, and in turn, the shape) and at the macroscopic level (protein doesn't function, and this causes symptoms of the disease).

References:

Canning, D. R. and Cox, James R. (2001). Teaching the structural nature of biological molecules: Molecular visualization in the classroom and in the hands of students. *Chemistry Education Research and Practice in Europe* Vol. 2, No. 2., pp109-122

Chan,H. S.  and E. Bornberg-Bauer, (2002). *Applied Bioinformatics*, 1, 121-144.

Chi, M. , Glaser, R., and Farr, M. (Eds.) (1988). *The Nature of Expertise*, Erlbaum, Hillsdale, NJ.

Feurtzeig, W., & Roberts, N. (Eds.) (1999). Modeling and simulation in science and mathematics education. New York: Springer.

Jackson, S. L., Krajcik, J. S., Soloway, E. (1998). The Design of Guided Learner-Adaptable Scaffolding in Interactive Learning Environments, ACM CHI '98 Human Factors in Computer Systems Proceedings, Los Angeles, CA, Addison-Wesley, 187-194.

Jiang T. , Y. Xu, and M. Q. Zhang, (Eds.) (2002). *Current Topics in Computational Molecular Biology*, The MIT Press, Cambridge, Massachusetts,

Gilbert, J. K., Boulter, C. J., Rutherford, M. (1998a). Models in Explanations, Part 1: Horses for Courses?. *International Journal of Science Education* 20(1), 83-97.

Gilbert, J., Boulter, C. J., Rutherford, M. (1998b). Models in Explanations, Part 2: Whose voice, whose ears ?. *International Journal of Science Education*, 20(2), 187-203.

Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1), 73-79.

Horwitz, P., & Barowy, W. (1994). Designing and using open-ended software to promote conceptual change. *Journal of Science Education and Technology, 3*, 161-185.

Lau, K. F. and K. A. Dill, Macromolecules, 22, 3986-3997 (1990). Proc. Natl. Acad. Sci. USA, 87, 638-642.

Linn, M. C., & Muilenburg, L. (1996). Creating lifelong science learners: What models form a firm foundation? *Educational Researcher*, 25(5), 18-24.

Tversky, B.; Morrison, J.B. & Bétrancourt, M. (2002) Animation: can it facilitate?  International Journal of Human-Computer Studies, Vol.57: 4, 247-262.

Nakhleh, M. B. (1992). Why some students don't learn chemistry. *Journal of Chemical Education* 69:191-196.

Pallant, A and R. Tinker (2004). Reasoning with Atomic-Scale Molecular dynamics models. *Journals of Science Education and Technology* Vol. 13, No.1, April.

Perkins, D. (1995). *Smart schools: Better thinking and learning for every child*. New York: The Free Press.

Rieber, L. P. (1989). *A review of animation research in computer-based instruction*. In M. R. Simonson and D. Frey (eds.), Proceedings of Selected Research Paper Presentations at the Convention of the Association for Educational Communications and Technology and Sponsored by the Research and Theory Division. Washington, DC: Association for Educational Communications and Technology, 369-389.

Sanger, Michael J. Computer Animations in Chemistry: What We Have Learned http://faculty.cns.uni.edu/~sanger/Review

Snir, J., Smith, C. (1995). *Constructing Understanding in the Science Classroom: Integrating Laboratory Experiments, Student and Computer Models, and Class Discussion in Learning Scientific Concepts,* IN D. N. Perkins, J. L Schwartz, M.W. West, & M. S. Wiske (Eds.) Software Goes to School: Teaching for Understanding with New Technologies. Oxford University Press, NY.

Tinker, R., & Horwitz, P. (2000). Modeling Across the Curriculum: IERI Planning Grant: National Science Foundation (REC-0089198).