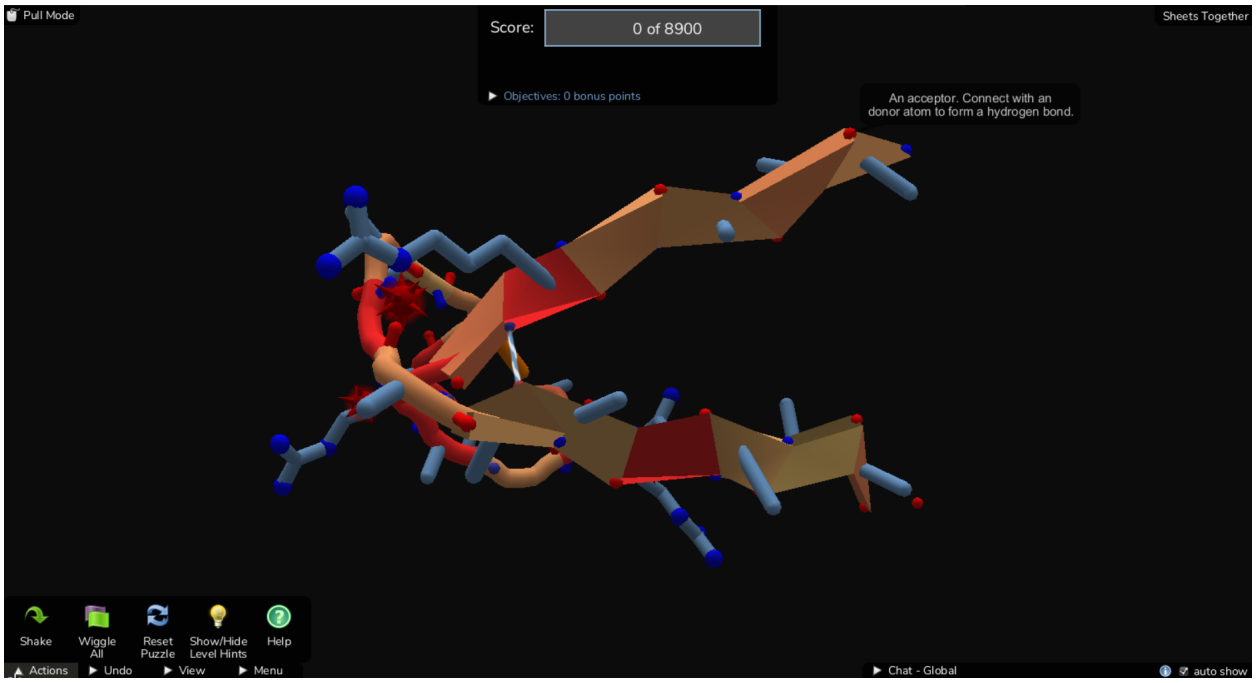
Citizen Science Featurettes

Foldit



Schedule

Thurs., Sept. 30, 2-3:30 pm ET.

See event on LibCal. Open this file in GitHub Pages.

Description

“Foldit is a revolutionary crowdsourcing computer game enabling you to contribute to important scientific research. This page describes the science behind Foldit and how your playing can help.”

Links

1. Homepage: https://fold.it/
2. Instructional resources: https://fold.it/educator

Agenda

00:00	Intro to Citizen Science	Introduction to Citizen Science
00:15	Setup	Create FoldIt account
00:20	Orientation to Project	Overview of features Training in methodology
00:30	Folding Exercise	Join Classroom Individual folding Group folding
1:30	Results and Discussion	Review group analytics Discussion
01:30	Finish	

Introduction

About Citizen Science

Opening video (optional, 9 ½ min.)

“The Awesome Power of Citizen Science.” SciShow. https://www.youtube.com/watch?v=SZwJzB-yMrU.

About FoldIt

Proteins are involved in almost all of the processes going on inside your body: they break down food to power your muscles, send signals through your brain that control the body, and transport nutrients through your blood. Proteins are present in all living things, even plants, bacteria, and viruses. Some organisms have proteins that give them their special characteristics.

Knowing the structure of a protein is key to understanding how it works and to targeting it with drugs. Figuring out which of the many, many possible structures is the best one is regarded as one of the hardest problems in biology today and current methods take a lot of money and time,

even for computers. Foldit attempts to predict the structure of a protein by taking advantage of humans' puzzle-solving intuitions and having people play competitively to fold the best proteins.

Since proteins are part of so many diseases, they can also be part of the cure. Players can design brand new proteins that could help prevent or treat important diseases.

Learning Objectives

By the end of this activity students will:

1. Acquire basic familiarity with biochemistry fundamentals such as protein structure and function; see how different biochemical rules govern how proteins fold and acquire their shape; and learn how proteins can be designed to combat diseases
2. Learn about the practical applications and impacts of predicting and designing proteins' structures and contribute to active, real-world research
3. Explain how crowdsourcing is a valid tool that can be used to help biochemists solve difficult problems more efficiently
4. Apply their understanding of crowdsourcing to other broader contexts where they use crowdsourced data to inform decisions in their daily lives

Setup

1. Navigate to the FoldIt website.
2. Use the links at the top right corner to sign in or register a new account.
3. Install Foldit on your computer.

Orientation to Project

Overview of Features

- Filter by difficulty and puzzle category
- Personal stats page
- AI-assisted tools
- Community Wiki
- Teaching and training materials for educators
- Contests
- Foldit Lab Reports make connections to real research

Demo of methodology

1. Welcome to FoldIt (introduction)
2. Expert Folder Protein Folder: JSnyder (intuition builder)

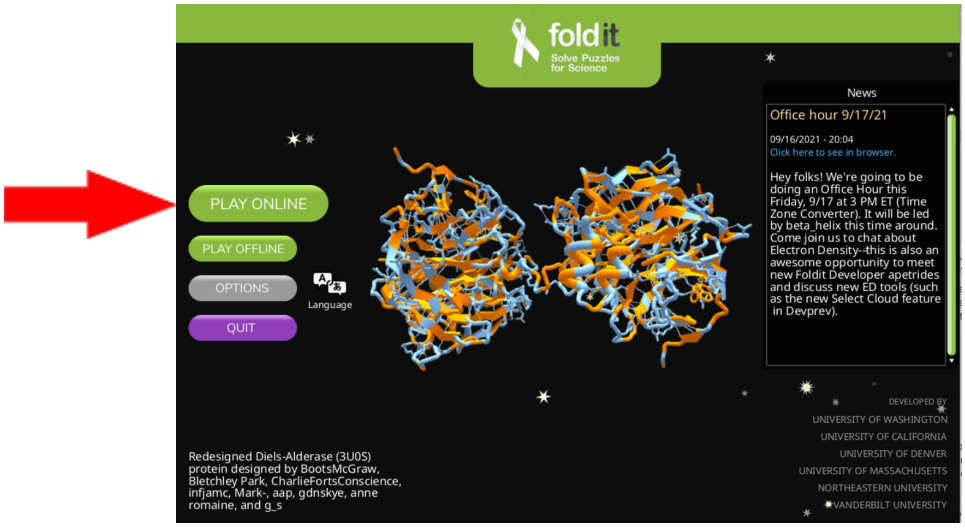
Folding Exercise

Join Classroom

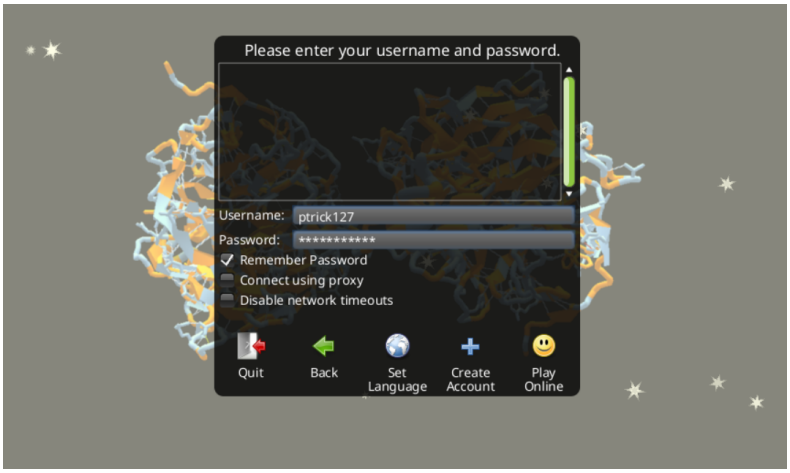
1. Navigate to this link. This is the private classroom we'll be using for this exercise.
2. Once the Zoom meeting has started,drop your Foldit username in the chat screen. (I will use this to invite you to our private classroom so we can review the group's analytics at the end of the exercise.)
3. Open your emails and click on the link to accept the invitation to the classroom. (Make sure you're using the same email account that you used to register with Foldit.)
4. Visit the classroom page and verify that your name is in the players list near the bottom.

Launch

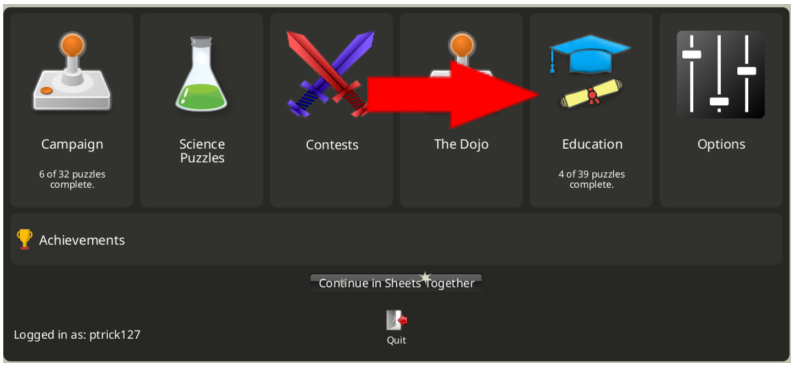
5. Locate Foldit on your computer and open it.
6. Select the Play Online option.



7. In the next screen, enter your user credentials to sign in.

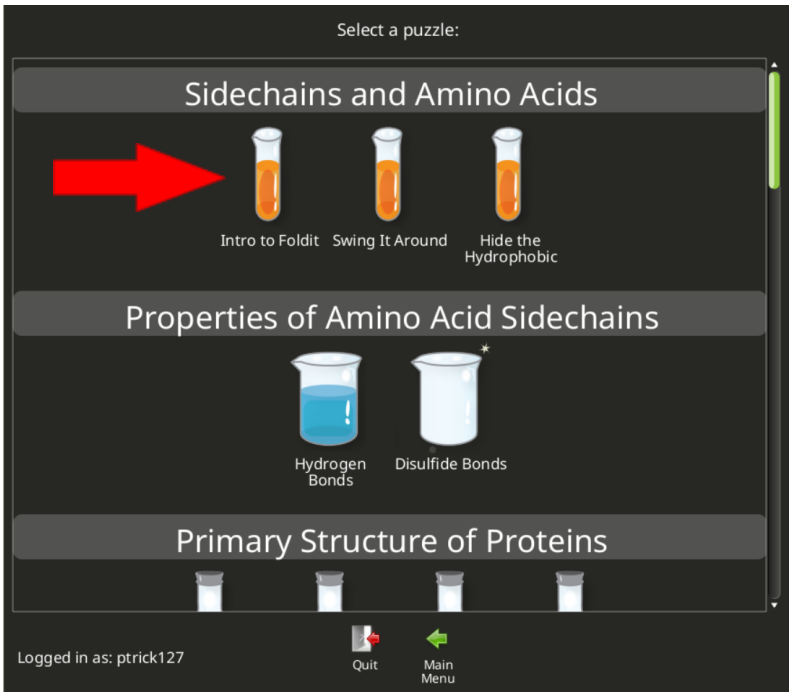


8. Select the Education Mode from the selection screen.



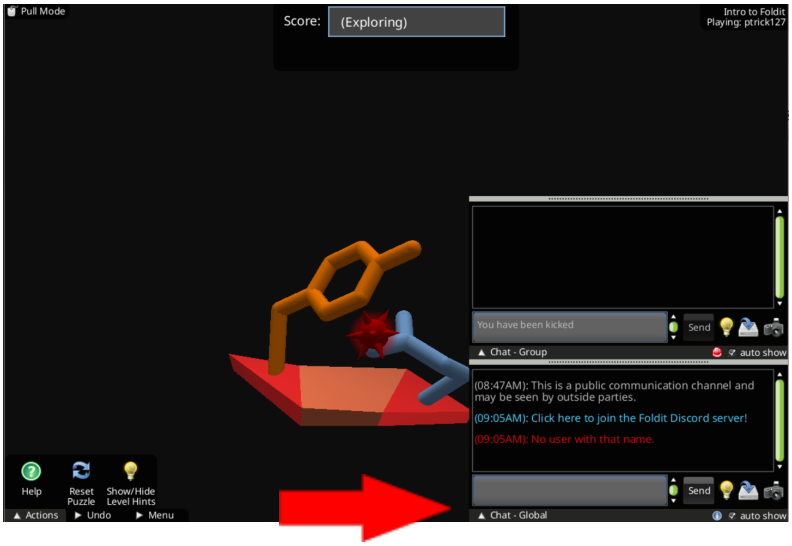
Select a puzzle

9. Select the Intro to Foldit puzzle in the Sidechains and Amino Acids section.

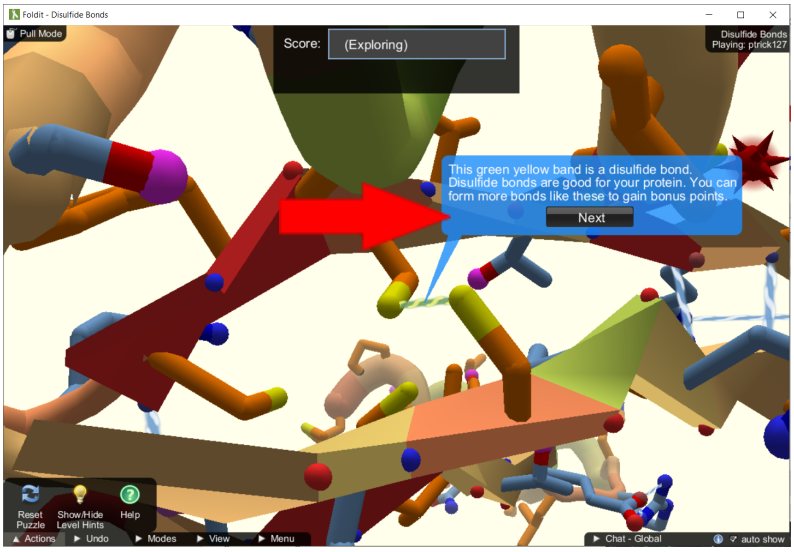


Start Folding

10. Read through the initial set of pop-ups to understand the goal of the puzzle.
11. Minimize the chat boxes in the bottom-right corner.



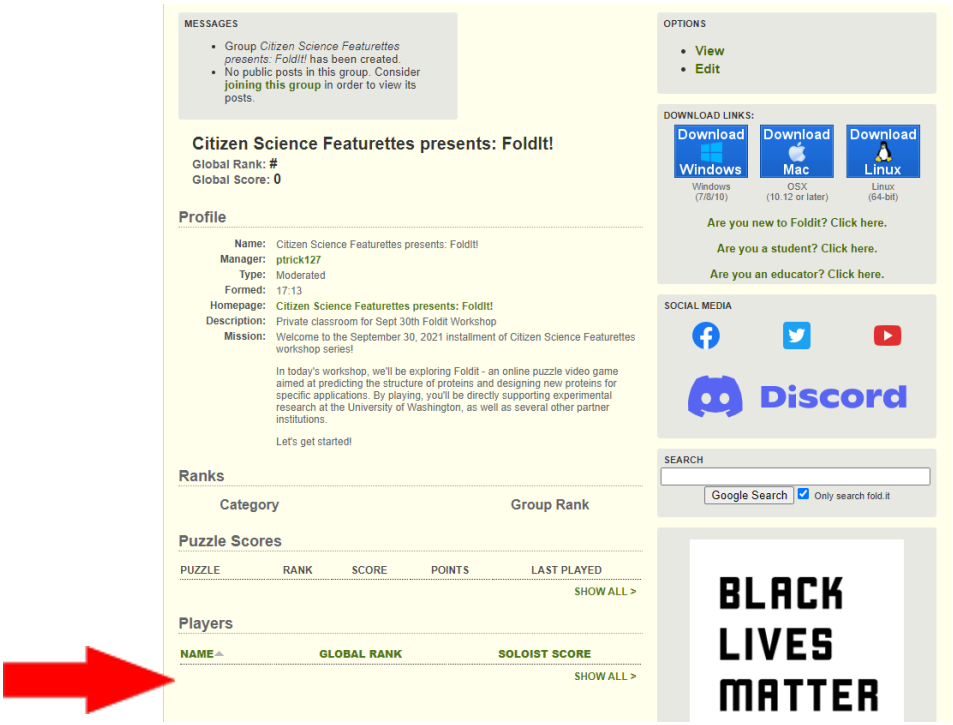
12. Follow the prompts in the pop-ups to start solving puzzles.



13. After 15 minutes of individual puzzle-solving, partner up with another participant (the facilitators will assign you to a group and place you in a breakout room). Continue solving puzzles for an additional 15 minutes.

Results

Check your analytics on the private classroom portal here.



Debrief and Discussion

1. What were the biggest challenges?
2. Were there certain tasks that were easier to execute than others?
3. Did you perform better (solve more puzzles, etc.) individually or as a group? What were the biggest differences you noticed?
4. What are the advantages and disadvantages of having a single individual or research team solve these puzzles on their own.
5. What are the advantages and disadvantages of having members of the public do the same task online.
6. What is the difference between prediction and design? What do you think are the applications of each puzzle type?
7. What do you think is the ideal role for humans and artificial intelligence both in predicting the shape of proteins as well as in designing new proteins?
8. Can you think of any other aspects of your life where you encounter (or participate in) crowdsourced data?

Closing video (time-permitting)

“Looking at TOP Foldit Coronavirus solutions (plus: lab testing).”
https://www.youtube.com/watch?v=e1YEyPX-Pqw.

Next Steps

Below are a few suggestions for continuing your journey into the world of gamified, crowdsourced protein folding.

Level Up as a Foldit Player

There's a lot more to discover within the Foldit community. Below, we've provided a few ways you can level up as a Foldit player and gain access to some of the other features and opportunities found there.

1. Foldit Blog
2. Watch the Foldit Lab Reports to learn more about the science that goes on behind Foldit.
3. Help beta test the new Alpha Fold AI software for protein structure prediction (learn more here)

Explore other similar projects

1. Mozak's Neuron Challenge - Mozak is a scientific discovery game about neuroscience. Help us build models of brain cells, and help scientists learn more about the brain through your efforts!

Learn more about citizen science by checking out these media links

Finally, check out the links below to learn more about the protein folding problem and how Foldit is helping scientists make progress on contemporary problems like how to fight the coronavirus pandemic.

1. FoldIt Gamers FTW (SciShow). https://www.youtube.com/watch?v=JdBcpdH_ptA.
2. The Protein Folding Problem: a major conundrum of science: Ken Dill at TEDxSBU. https://www.youtube.com/watch?v=zm-3kovWpNQ.
3. "'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures." Nature. https://www.nature.com/articles/d41586-020-03348-4.
4. Looking at TOP Foldit Coronavirus solutions (plus: lab testing). https://www.youtube.com/watch?v=e1YEyPX-Pqw.
5. TED Radio Hour. "Citizen Science." https://www.npr.org/programs/ted-radio-hour/551030943/citizen-science
6. "The Awesome Power of Citizen Science." SciShow. https://www.youtube.com/watch?v=SZwJzB-yMrU.
7. FoldIt YouTube Channel. https://www.youtube.com/channel/UCGjfDFjL-7rvhRRl2keqhcA

Appendix 1. Student Help Page

Link to source: https://fold.it/student (Updated 9/24/21)

Thank you for joining us and trying Foldit!

1. Getting Started
2. Handy Tips
3. Most Importantly

Getting Started

Download the original Foldit app on the main Foldit page.

- Please remember Foldit is used by students of all ages. Please pick an appropriate username.
- The game is much easier with a mouse than with a trackpad. We highly recommend using an external mouse.
- If you don't have access to chat channels, it's probably because your instructor has turned them off.
- These are the community rules (https://fold.it/portal/communityrules) which help keep Foldit a great place to play and learn. If your instructor didn't mention these, you probably should read these before diving into chat.

Handy Tips

- Foldit is a game about learning and scientific discovery. While it's not an easy game, hundreds of thousands of people with no experience or training in science have been trying out and playing Foldit for years. You can do it!
- If you are stuck on a tutorial please reset the tutorial and try again. All the tutorials contain everything you need. You just have to read them very closely.
- For even more information, be sure to check out the wiki (http://foldit.wikia.com/wiki/FoldIt_Wiki) there is a lot of great info there, tips, tricks, info about structures, as well as videos for an entire biochemistry class if you want to learn more!
- If you are wondering about protein folding and why is it important, look here (https://fold.it/portal/info/about). We are glad you asked, because it is great to have some context with what you are doing.

Most Importantly- have fun!

Having issues? Email mail.fold.it@gmail.com

Appendix 2. Extract from *Reinventing Discovery*

Below is an extract from Michael Nielsen's 2014 book, *Reinventing Discovery*, in which he discusses how amateur FoldIt players rival professional biochemists in predicting the shape of proteins.

When Amateurs Rival Professionals

It's not just in astronomy that citizen science is useful. One of the big open problems in biology is to understand how the genetic code gives rise to an organism's form. Of course, we've all heard many times that DNA is the "blueprint for life." But even though the slogan is familiar—it is, after all, the fate of great slogans to become cliches—that doesn't mean anyone yet understands in detail how DNA gives rise to life. Suppose biologists had never seen an elephant's trunk. Could they look into an elephant's DNA and somehow see the trunk there—that is, predict the trunk's existence based solely on the sequence of base pairs in an elephant's genetic code? Today, the answer to this question is no: how DNA determines an organism's form is one of the mysteries of biology.

To help solve this mystery, a citizen science project called Foldit is recruiting online volunteers to play a computer game that challenges them to figure out how DNA gives rise to the molecules called proteins. That challenge may sound a far cry from deducing the existence of the elephant's trunk—it is a far cry—but it's a crucial step along the way, because proteins carry out many of the most important processes in our bodies. Aside from its intrinsic scientific interest, Foldit is also interesting as a demonstration of the great complexity of work that can be done by volunteers. In Galaxy Zoo, participants mostly carry out simple tasks, such as classifying a galaxy as spiral or elliptical. In Foldit, players are asked to tackle tasks that would challenge a biochemistry PhD. And, as we'll see, the top Foldit players are doing those tasks extraordinarily well.

Before we discuss Foldit in detail, let's talk a bit about proteins in general. Biologists are obsessed by proteins, and with good reason: they're molecules that do everything from digesting our food to contracting our muscles. A good example of a protein is the hemoglobin molecule. Hemoglobin is one of the main components in our blood: it's the molecule our bodies use to move oxygen from our lungs to the rest of our body. Another important class of proteins are the antibodies in our immune system. Each antibody has its own special shape that lets it lock on to viruses and other intruders in our body, tagging them for attack by our immune system.

At present we only partially understand how DNA gives rise to proteins such as hemoglobin. What we do know is that certain sections of our DNA are protein coding, meaning that they describe a specific protein. So, for example, there's a protein-coding section for hemoglobin somewhere in your DNA. That region is a long string of DNA bases, which starts: CACTCTTCTGGT. . . . It turns out to be helpful to divide that string of bases into triplets, which are called codons: CAC TCT TCT GGT . . . . The way proteins are formed is that each codon in the protein-coding section of your DNA is transcribed into a corresponding molecule in the protein called an amino acid. So, for example, the first codon for hemoglobin, CAC, gets transcribed into an amino acid known as histidine. I won't explain exactly what histidine is, or

what it does—for us it doesn't much matter. What matters is that everywhere the CAC codon appears in the DNA sequence for hemoglobin (or any other protein), it gets transcribed to histidine. In a similar way, the second codon, TCT, gets transcribed into the amino acid serine. And so on. The resulting protein is a chain containing all those amino acids—so hemoglobin is a chain containing histidine, serine, and so on.

Okay, so far, so good: DNA can be used as a recipe for generating proteins. Proteins, however, differ from DNA in that they each have their own special shape, unlike the completely regular structure of DNA. That shape is tremendously important. For example, as I mentioned before, the antibodies in our immune system are proteins, and the shape of an antibody determines which viruses it can lock onto. What's going on is that as the information in the DNA is transcribed to form the amino acids in the protein, the protein "folds" into its shape. How this folding occurs is still only partially understood, but there are some basic rules of thumb that should give you the flavor of what's going on. Some amino acids like to be near water—they're called hydrophilic, from the Greek roots "hydro" and "philia," for water and love, respectively. Since proteins inside a cell are surrounded by water, the protein will tend to fold so the hydrophilic amino acids sit on the outside, near the water. Histidine and serine are both examples of hydrophilic amino acids. By contrast, hydrophobic amino acids—amino acids that don't like water—end up bundled up tight inside the protein. Sometimes these tendencies conflict: neighboring amino acids in the protein may be alternately hydrophobic and hydrophilic, with the result that the protein can end up folding into a very complex shape.

There's an incredibly clever trick here that nature is using. The DNA is a completely regular arrangement of information, which makes it both easy to copy and relatively straightforward to transcribe into amino acids. But then competition between hydrophilia, hydrophobia, and other forces means that the protein can fold up to form complex shapes. By changing the DNA we can change the amino acids in the protein, which in turn causes the shape of the protein to change. What's clever about this is that it takes us from the regularly arranged information in the DNA, which is easily copied, to the many possible shapes of the protein. A priori, shapes don't seem so easy to copy. It's as though you could trace over the blueprint for a house, and the traced version would then somehow spring into existence as a tiny model house. The DNAprotein connection is Nature's way of making easy the seemingly impossible task of copying complex shapes.

But there's a problem with this neat story. Just because we know the DNA sequence for a protein doesn't mean we can easily predict what shape the protein has, or what the protein will do. In fact, today we have only a very incomplete understanding of how proteins fold. Complete structures—the exact shapes—are known for only 60,000 proteins, despite the fact that we know the DNA sequences for millions of proteins. Most of those complete structures have been found using a technique called X-ray diffraction—basically, shining X-rays at a protein and figuring out its shape by looking carefully at the X-ray shadow it casts. It's slow, expensive, painstaking work, and the techniques are only gradually getting better. What we'd really like is a fast and reliable way to predict the shape from the genetic description. If we could do that, cutting out the slow and expensive X-ray diffraction step, we'd go from knowing the shape of 60,000 proteins to knowing the shape of millions. Even more significantly, such a method would be a tremendously powerful tool for helping us design proteins with desired shapes. This would, for instance, help us engineer new antibodies to fight disease.

To solve the protein folding problem, biochemists have turned to computers in an attempt to predict protein shape from the genetic description. To make their predictions they use the idea that a protein will eventually fold into its lowest energy shape, much as a ball will roll to the bottom of a valley between two hills. All that's needed is good method for finding the lowest energy shape of a protein. This sounds promising, but in practice it's hard to search through all the possible shapes, looking for the shape with the lowest energy. The difficulty is the number of different shapes a protein can potentially fold into. Proteins typically have hundreds or even thousands of amino acids. To determine the structure means knowing the exact position and orientation of every single one of those amino acids. With so many amino acids involved, the number of possible shapes is astronomical, far too many to search through even on a very powerful computer. Enormous effort has been put into finding clever algorithms that can be used to restrict the number of configurations that must be examined, and the algorithms are getting pretty good. But there's still a long way to go before we can use computers to reliably predict protein shapes.

In 2007, a biochemist named David Baker and a computer graphics researcher named Zoran Popovic, both from the University ´ of Washington, in Seattle, had an idea for a better way of solving the problem. Baker and Popović's idea was to create a computer game that shows a protein to the player, and gives them controls to change the shape, rotating the protein, moving amino acids around, and so on. Some of the controls built into the game are similar to the tools used by professional biochemists. The lower the energy of the shape the player comes up with, the higher their score, and so the highest scoring shapes are good candidates for the real shape of the protein. Baker and Popovic hoped that this might be a better approach to protein folding than the conventional approaches, combining state-of-the-art computational techniques with computer gamers' persistence and abilities at pattern matching and 3-D problem solving.

I was skeptical when I first heard about Foldit. It sounded like the dull educational computer games I saw in school when I was growing up in the 1980s. But I downloaded the game, and spent hours playing it over several days. At that point, the excuse "I'm doing research for my book" was rapidly becoming a euphemism for "this is a great way to procrastinate on writing my book," and I forced myself to stop. So far, more than 75,000 people have signed up. People play the game because it's good. It has the compelling, addictive quality all good computer games have: a task that's challenging but not impossible, instant feedback on how well you're doing, and the sense that you're always just one step away from improvement. It's the same addictive quality we saw earlier in the MathWorks competition, and which is also felt by many participants in Galaxy Zoo. Furthermore, like Galaxy Zoo, Foldit is deeply meaningful to many of the players. Einstein once explained why he was more interested in science than politics by saying, "Equations are more important to me, because politics is for the present, but an equation is something for eternity." Each time you classify a galaxy or find a better way to fold a protein, you're making a small but real contribution to human knowledge. For many participants, Foldit and Galaxy Zoo aren't guilty pleasures, like playing World of Warcraft or other online games. Instead, they're a way of contributing something important to society. One of the top Foldit players, Aotearoa, describes it as "the most challenging, exciting, stimulating, intense, addictive game I have ever played," and comments that it provides a way for people to "offer something proactive to solving some of the worlds/societies most

complicated puzzles, rather than waste time playing a ‘game’ that does not provide the same ‘rewards’ as folding protein does, this way!”

In addition to the individual motivation to play, Foldit also encourages collective problem solving by the players. There is an online discussion forum and a wiki, where players share news and discuss their strategies for protein folding. The game incorporates a simple programming language that players can use to create scripts—short programs—that automate game tasks. A typical script might implement a strategy for improving a fold, or identify which part of the protein’s current shape is in most need of improvement. Hundreds of such scripts have been publicly shared—an open source approach to protein folding. Many of the players work in groups, sharing their insights about the best ways of folding. All this work is greatly informed by the game score, which, as in the MathWorks competition, focuses participants’ attention where it will be most useful: when one of the high-scoring players shares a strategy tip or a script, other players pay attention. The players themselves are wildly varied, ranging from a self-described “educated redneck” from Dallas, Texas, to a theater historian from South Dakota, to a grandmother of three with a high-school education.

Just how good are the Foldit players at folding proteins? Every two years since 1994, there’s been a worldwide competition of biochemists using computers to predict protein structures. The competition, called CASP—Critical Assessment of Techniques for Protein Structure Prediction—is very important to the scientists who work on protein structure prediction. Before the competition starts, the CASP organizers approach some of the facilities that determine protein structure using the traditional approach of X-ray diffraction, and ask them what protein structures they expect to complete in the next couple of months. They then use those proteins as puzzles in CASP. Starting with the sequence of amino acids making up the protein, the CASP competitors are asked to predict the structure. At the end of the competition, teams are ranked on how close they come to the actual structure.

Foldit players competed in both the CASP 2008 and 2010 competitions. They performed extremely well, finishing near or at the top on many of the CASP challenges. Foldit developer Zoran Popovic summed up the results of the 2008 competition by saying that “foldit players are on a par, but not better than protein folding experts at trying to solve the same problem with all tools available to them. It also appears that foldit outperformed all fully automated server submissions.” Thus, a team of amateurs can be competitive with some of the world’s top biochemists, equipped with state-of-the-art computers. Popovic told me that his “ultimate goal is to show that experts are unequivocally inferior to the general population with this problem . . . a biochemistry PhD does not self-select for spatial reasoning. Structure prediction is all about 3d problem solving and very little about biochemistry.” Indeed, even specialists in protein structure prediction usually spend only a small fraction of their time working directly on predicting protein structures. And while they have expertise that the amateurs don’t, much of that knowledge is incarnate in the mechanics of the game. That levels the playing field enough that the remaining disparity in expertise can be overcome by the greater time commitment of the Foldit players. It’s a symbiosis: the professionals develop the systematic understanding that underlies the mechanics of the game, and the amateurs then supply the dedicated artistry required to take best advantage of that systematic understanding.