

Five questions for Rosetta@home:

How Rosetta@home helps cure cancer, AIDS, Alzheimer's, and more

This page attempts to explain the basic science behind Rosetta@home, and in doing so, explain how Rosetta@home can help lead to cures for human diseases.

1. [What are proteins?](#)
2. [What do proteins do?](#)
3. [How are proteins related to disease?](#)
4. [What role does protein structure play in treating disease?](#)
5. [What does Rosetta@home do on my computer?](#)

What are proteins?

Proteins are the workhorses in every cell of every living thing. Your body is made up of trillions of cells, of all different kinds: muscle cells, brain cells, blood cells, and more. Inside those cells, proteins are allowing your body to do what it does: break down food to power your muscles, send signals through your brain that control the body, and transport nutrients through your blood.

Proteins come in thousands of different varieties, but they all have a lot in common. For instance, they're made of the same stuff: every protein consists of a long chain of joined-together amino acids. Amino acids are small molecules made up of atoms of carbon, oxygen, nitrogen, sulfur, and hydrogen. To make a protein, the amino acids are joined in an unbranched chain, like a line of people holding hands. Just as the line of people has their legs and feet "hanging" off the chain, each amino acid has a small group of atoms (called a sidechain) sticking off the chain that connects them to each other (the mainchain or backbone). Every amino acid contributes the same "arms" to the mainchain, but unlike a line of people, the sidechains ("legs") of amino acids are quite different from each other. In fact, there are 20 different kinds of amino acids, which differ from one another based on what atoms are in their sidechains. The 20 amino acids have names like alanine, tryptophan, glutamine, and leucine.

Another thing proteins all have in common is that they don't like to stay stretched out in a straight line. The protein folds up to make a compact blob, but as it does, it keeps some amino acids near the center of the blob, and others outside; and it keeps some pairs of amino acids close together and others far apart. Every kind of protein folds up into a very specific shape -- the same shape every time. Most proteins do this all by themselves, although some need extra help to fold into the right shape.

It turns out that the identity of a protein is simply determined by which amino acids are in it, and what order they're in. Amazingly, a particular chain of amino acids always folds in exactly the same way -- whatever way results in the lowest overall energy. That means that to make two identical proteins with exactly the same shape and properties, all the body has to do is make two chains with the same amino acids in the same order. That's important, because there are thousands of identical copies of some proteins in every cell in your body!

Cells are always making new copies of proteins and breaking down old ones to be recycled. The blueprints for making proteins are genes, which are encoded in your DNA. There are small differences in DNA that make one person different from another. This means that one person's proteins may be slightly different than another's. It also means that half of your proteins come from the genes you inherited from your mother (and are like her proteins), and the rest come from your father (and are like his). Sometimes this is called the Central Dogma of molecular biology: every gene in your DNA gets translated into a protein in your body.

What do proteins do?

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. Many proteins act as enzymes, meaning they catalyze (speed up) chemical reactions that wouldn't take place otherwise. But other proteins power muscle contractions, or act as chemical messages inside the body, or hundreds of other things. Here's a small sample of what proteins do:

- [Amylase](#) starts the process of breaking down starch from food into forms the body can use.
- [Alcohol dehydrogenase](#) transforms alcohol from beer/wine/liquor into a non-toxic form that the body uses for food.
- [Hemoglobin](#) carries oxygen in our blood.
- [Fibrin](#) forms a scab to protect cuts as they heal.

- [Collagen](#) gives structure and support to our skin, tendons, and even bones.
- [Actin](#) is one of the major proteins in our muscles.
- [Growth hormone](#) helps regulate the growth of children into adults.
- [Potassium channels](#) help send signals through the brain and other nerve cells.
- [Insulin](#) regulates the amount of sugar in the blood and is used to treat diabetes.

Proteins are present in all living things, even plants, bacteria, and viruses. Some organisms have proteins that give them their special characteristics:

- [Photosystem I](#) is a collection of proteins in plants that captures sunlight for photosynthesis.
- [Luciferase](#) catalyzes the chemical reaction that makes fireflies glow.
- [Hemagglutinin](#) helps the influenza virus invade our cells.

How are proteins related to disease?

With all the things proteins do to keep our bodies functioning and healthy, they can be involved in disease in many different ways. Below, we list three diseases that represent different ways that proteins can be involved in disease.

HIV / AIDS: The HIV virus is made up largely of proteins, and once inside a cell it creates other proteins to help itself reproduce. [HIV-1 protease](#) and [reverse transcriptase](#) are two proteins made by the HIV virus that help it infect the body and replicate itself. HIV-1 protease cuts the "polyprotein" made by the replicating virus into the functional pieces it needs. Reverse transcriptase converts HIV's genes from RNA into a form its host understands, DNA. Both proteins are critical for the virus to replicate inside the body, and both are targeted by anti-HIV drugs. This is an example of a disease producing proteins that do not occur naturally in the body to help it attack our cells.

Cancer: Cancer is very different from HIV in that it's usually our own proteins to blame, instead of proteins from an outside invader. Cancer arises from the uncontrolled growth of cells in some part of our bodies, such as the lung, breast, or skin. Ordinarily, there are systems of proteins that limit cell growth, but they may be damaged by things like UV rays from the sun or chemicals from cigarette smoke. But other proteins, like [p53 tumor suppressor](#), normally recognize the damage and stop the cell from becoming cancerous -- unless they too are damaged. In fact, damage to the gene for p53 occurs in about half of human cancers (together with damage to various other genes).

Alzheimer's: In some ways, Alzheimer's is the disease most directly caused by proteins. A protein called [amyloid-beta precursor protein](#) is a normal part of healthy, functioning nerve cells in the brain. But to do its job, it gets cut into two pieces, leaving behind a little scrap from the middle -- amyloid-beta peptide. Many copies of this peptide (short protein segment) can come together to form clumps of protein in the brain. Although many things about Alzheimer's are still not understood, it is thought that these clumps of protein are a major part of the disease.

What role does protein structure play in treating disease?

Proteins are very small, too small to see, even with a microscope. However, using special x-rays or very powerful magnets, scientists have been able to figure out the structures of some proteins -- what they would look like if we could see them. A complete structure defines the three-dimensional position of every atom in the protein.

In order for a protein to function normally, it must usually bind to and interact with at least one chemical compound or other protein. The site of interaction is called the protein's binding site (or active site, for enzymes that carry out chemical reactions). The interaction depends on a nearly perfect fit between the shape of the binding site and the thing it binds to, like a key fitting into a lock. Solving the structure of a protein allows us to see the exact shape and position of its binding site(s).

Most drugs do their work by targeting the binding or active site of a particular protein. For example, the anti-cancer drug tamoxifen fits into the binding site of [estrogen receptor](#). Without the drug, estrogen will bind to estrogen receptor, which can contribute to the uncontrolled growth of cancer cells. With the drug in place, estrogen can't get in, and so the cancer's growth is slowed down.

Traditionally, drugs have been discovered by what amounts to trial and error. But if a specific protein is known to be involved in a disease, and if the structure of that protein is known, then chemists can try to design a drug to bind to the protein. If it works, the new drug will bind to the target protein and keep it from doing whatever it does. For example, two proteins from the HIV virus, [HIV-1 protease](#) and [reverse transcriptase](#), have been targeted this way. Unfortunately, this is still a time-consuming process and success is not guaranteed. However, many people believe that knowing protein structures will play an increasingly important role in the future of drug discovery.

Of course, drug design isn't the only role for protein structure in treating disease: the structure of a protein helps explain what it does and how it does it. That can give insight into how particular processes work in the body, and how they go wrong in disease. This kind of basic understanding can contribute to the treatment of a disease apart from any specific drug.

What does Rosetta@home do on my computer?

The computer program Rosetta works on several kinds of calculations, but they all relate to protein structure. A large number of computers are needed because the calculations take a long time, and many different possibilities must be explored in order to discover the right answers.

For more information, [Dr. David Baker's Rosetta@home Journal](#) is a great source of timely information about what new projects are being run on Rosetta@home, and how they relate to important biomedical problems. You can also see the [Active WorkUnits Log](#), which has up-to-date information about what Rosetta@home is working on *right now*, but descriptions may be terse and/or technical.

Design of therapeutic proteins: Since proteins are part of so many diseases, they can also be part of the cure. The Baker lab is using Rosetta@home to design brand new proteins that could help prevent or treat important diseases. For example, Rosetta is being used to redesign parts of the HIV virus' coat so they can be administered as an effective vaccine. We are also working on antagonists for androgen receptor (a protein involved in prostate cancer) and on novel endonucleases (protein enzymes that cut DNA) for gene therapy approaches to treating a variety of hereditary diseases. See [Disease Related Research](#) for more details.

Protein structure prediction: As described above, knowing the structure of a protein is key to understanding how it works and to targetting it with drugs. Rosetta attempts to predict the structure of a protein computationally, as opposed to determining it experimentally. This problem, often called "the protein folding problem", is regarded as one of the hardest problems in biology today. A computational solution is desirable because experimental methods generally take months to years of time and cost hundreds of thousands of dollars per protein. (And human beings have tens of thousands of different proteins -- not to mention all the proteins in other organisms.)

Some of the structure predictions on Rosetta@home are for proteins whose structure is truly not known. The resulting models are used to address some specific biological question, such as a disease mechanism. Other predictions are tests where we already know the answer, aimed to improve Rosetta itself. A specific case of this is the CASP competition, where teams of researchers around the globe try to predict the structures of proteins where the answer was recently determined experimentally but is not yet public knowledge. Rosetta is consistently among the best performers in CASP.

Protein-protein docking: Docking problems focus on predicting how two things bind to each other -- in this case, how two proteins bind to each other. Knowing which parts of the proteins interact and their relative orientations in space helps explain the functions of those proteins. It also suggests ways to create drugs that would disrupt the interaction, if the interaction is part of a disease. (For instance, HIV proteins that bind to cell surface proteins so that the virus can infect the cell.) In some cases, the structures of the two proteins by themselves are already known, but in other cases we must first predict them (see above). To improve Rosetta's performance on protein-protein docking, the Baker lab also competes in CAPRI. As in the CASP competition, in CAPRI researchers try to predict protein-protein interactions that have recently been determined but are not yet public knowledge.

Drug docking and design (coming soon): Rosetta can also be used to dock drug-like molecules with proteins to see how they might bind to each other. By trying many potential drugs from a large library of molecules, we may be able to discover a drug that binds to a protein of interest. On the other hand, if a drug is already known to target that protein, then by predicting how it binds to the protein we can predict ways of improving the drug. (These calculations are not currently running on Rosetta at home, but we expect that they will be soon.)

Please direct questions and comments about this page to iwd@u.washington.edu.