

# 5 Learning to Use Cn3D: A Bioinformatics Tool

## Introduction

Up to this point, students have seen the BRCA1 protein represented in a linear, sequential form. In this lesson, students are introduced to the high importance of a protein's three-dimensional structure. Students first engage in a short activity in which they use a pipe cleaner to perform a simple function, as an analogy for the relationship between a protein's structure and function. Students then learn to navigate between linear protein sequences and three-dimensional structures by using the bioinformatics tool **Cn3D**. Students begin by viewing and manipulating DNA—a familiar molecule—using Cn3D. When students are familiar with the program, they visualize parts of the BRCA1 protein to show how a specific mutation in the *BRCA1* gene ultimately changes or destroys the protein's function. In *Lesson Five*, students learn how **3D animators** might use bioinformatics tools in their careers.

## Learning Objectives

At the end of this lesson, students will know that:

- Bioinformatics tools like Cn3D help scientists visualize proteins.
- A protein is a physical “thing” with a three-dimensional structure that determines its function.
- All proteins are comprised of amino acids linked together by **covalent bonds** and have the same general structure: a ‘beginning’ or **N-terminus**, which contains an amino group ( $\text{NH}_3^+$ ), and an ‘end’ or **C-terminus**, which contains a carboxyl group ( $\text{COO}^-$ ).
- Each amino acid has a chemical group that is unique to it, called an R-group (also known as a side-chain). The chemistry of the amino acid R-group is important for a protein's shape and function.
- Mutations can impact the three-dimensional structure of proteins, and thus impact the protein's function, as is the case with many genetic disorders.
- Bioinformatics tools are used and created by people in many careers, including 3D animators.

At the end of this lesson, students will be able to:

- Use Cn3D to view complex biological molecules in a three-dimensional format.
- Manipulate three-dimensional images in numerous ways to deepen understanding of molecular structure.

## Key Concepts

- Genetic disorders are often caused by dysfunctional or absent proteins.
- Proteins are physical “things” with three-dimensional shapes. The shape of the protein is crucial to its function.

## Class Time

One class period of 50 minutes.

## Prior Knowledge Needed

- Basic DNA structure.
- An understanding of the Central Dogma.
- An understanding of how DNA, chromosomes, genes, and proteins relate to each other.
- The process of protein synthesis (transcription and translation).
- Basic amino acid structure.
- Basic protein structure (specifically the various forms of secondary protein structure such as alpha helix and beta sheet).

## Common Misconceptions

- Biological molecules (such as DNA or proteins) actually look like their artistic representations.

- All proteins are comprised of amino acids linked together by covalent bonds and have the same general structure: a 'beginning' or N-terminus, which contains an amino group ( $\text{NH}_3^+$ ), and an 'end' or C-terminus, which contains a carboxyl group ( $\text{COO}^-$ ).
- Each amino acid has a different R-group (also known as a side-chain). Each R-group has different chemical properties. The chemistry of the amino acid R-group is important for a protein's shape and function.
- Mutations can cause changes to the three-dimensional shape of a protein. The change in shape can alter the function, resulting in genetic disorders, including cancer.
- Programs like Cn3D allow scientists (and students!) to view macromolecules in a number of different ways to enhance their understanding of the molecule's structure and function.
- Bioinformatics tools are used and created by people in many careers, including 3D animators.

## Materials

Materials	Quantity
Copies of Student Handout— <i>Careers in the Spotlight</i> (handed out in Lesson One)	1 per student
Class set of Student Handout— <i>Instructions for Seeing DNA in 3D</i>	1 per student (class set)
Student Handout— <i>Seeing DNA in 3D Worksheet</i> [Note: This worksheet is for students' answers to lesson questions.]	1 per student
Pipe cleaners	1 per student
Teacher Answer Key— <i>Seeing DNA in 3D</i>	1
Teacher Resource— <i>Teacher Demonstration: Aligning 3D Structures with VAST</i>	1
Teacher Resource— <i>Installing Cn3D</i> [Note: If students will be asked to install Cn3D, make a class set of copies of this handout.]	1 -OR- 1 per student
Teacher Resource— <i>From Sequence to Structure</i> [Note: This is an optional lesson extension.]	1

## Computer Equipment, Files, Software, and Media

Computer with internet access and projector to display PowerPoint slides and for the Teacher Demonstration described in Teacher Resource—*Teacher Demonstration: Aligning 3D Structures with VAST*.

**Alternative:** Print PowerPoint slides onto transparencies and display with overhead projector.

Lesson Five PowerPoint Slides—*Learning to Use Cn3D: A Bioinformatics Tool*. Available for download at:  
<http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing>.

A student version of lesson materials (minus teacher answer keys) is available from NWABR's Student Resource Center at:  
<http://www.nwabr.org/students/student-resource-center/instructional-materials/introductory-bioinformatics-genetic-testing>.

**VAST** Protein Alignment for the Teacher Demonstration described on Teacher Resource—*Teacher Demonstration: Aligning 3D Structures with VAST*. Available for download from the Resources tab at:  
<http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing>.

BRCA1 Animation. Available for viewing at: [http://www.nwabr.org/education/animations/BRCA1\\_animation.html](http://www.nwabr.org/education/animations/BRCA1_animation.html).

Computer lab with internet access, a word processing program such as Microsoft Word® or Google Docs, and the program Cn3D installed on every computer.

Access to the Microsoft Word® or Google Docs document that students created in Lesson Four.

Music files (mp3 format) for Teacher Resource—*From Sequence to Structure* are available for download from the Resources tab at: <http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing>.

## Teacher Preparation

- Load the classroom computer with the *Lesson Five* PowerPoint slides.
- To maximize class time for the lesson activities, it will be useful to install the Cn3D program on all classroom computers. See Teacher Resource—*Installing Cn3D* for complete instructions on installing Cn3D. Alternatively, make a class set of Teacher Resource—*Installing Cn3D* and have students install the program.
- Download the VAST alignment file for the Teacher Demonstration on Teacher Resource—*Teacher Demonstration: Aligning 3D Structures with VAST*. The VAST alignment can be found under the Resources table at: <http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing>.
- Make copies of Student Handout—*Instructions for Seeing DNA in 3D*, one per student. This handout is designed to be reused as a class set.
- Make copies of Student Handout—*Seeing DNA in 3D Worksheet*, one per student. The worksheet is used for students to write their answers to the lesson questions.

[**Note:** You may need to contact your school administrator or IT support staff to be sure that Cn3D can be downloaded to your classroom computers. Cn3D is freely available through the National Center for Biotechnology Information, a division of the US government, US National Library of Medicine.]

## Procedure

### WARM UP

1. As students enter the classroom, display **Slide #1**. This slide highlights 3D animator Beth Anderson.

**3D ANIMATOR**  
BETH ANDERSON

**Place of Employment:**  
Arkitek Studios

**Type of Work:**  
Develops 3D animations that help scientists and students visualize biological processes.

“Bioinformatics interfaces in many ways with our work, from data that clients give us to visualize representationally, to creating actual visual data.”

Learning to Use Cn3D: **Slide #1**

2. Have students retrieve Student Handout—*Careers in the Spotlight* from *Lesson One*.
3. Students should think about, and write down, what kind of work a 3D animator might do (3D Animator Question #1). This will be revisited at the end of the lesson, including how a 3D animator might use bioinformatics in his or her job.
4. Tell students to keep their *Careers in the Spotlight* handout available for future lessons.

## PART I: Structure Meets Function: Pencil Transferase

- Explain to students the **aim of this lesson**. Some teachers may find it useful to write the aim on the board.

### Lesson Aim:

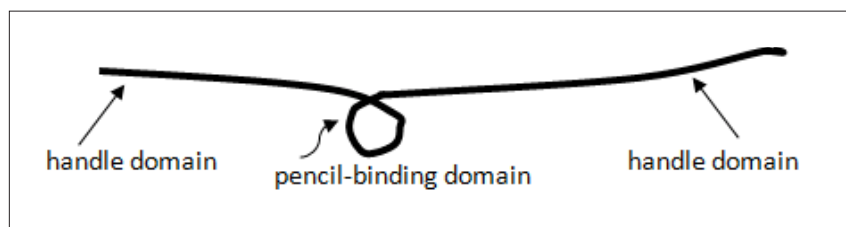
- To understand that protein structure can impact protein function, using the bioinformatics tool Cn3D to visualize molecules.

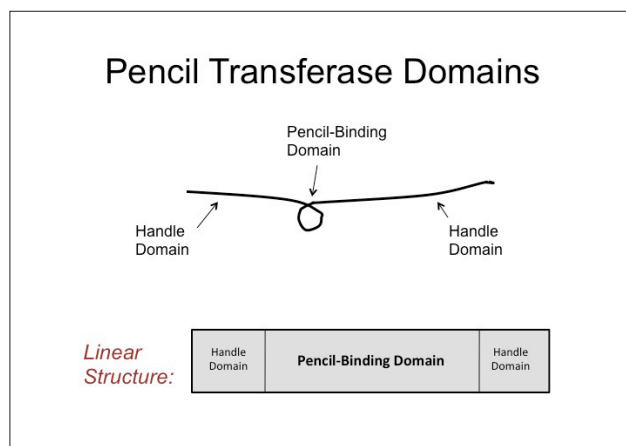
Teachers may also wish to discuss the *Learning Objectives* of the lesson, which are listed at the beginning of this lesson.

- Give each student a pipe cleaner. Tell students that their task is to **move a pencil from one desk to another desk** using the pipe cleaner.
- The rules are:
  - Students must use a pencil (not a pen or other implement with a cap).
  - Students may not touch the pencil with their hands when they are moving the pencil across desks.
  - Students may bend the pipe cleaner as needed.
  - After the pencil has been successfully moved, students should **retain the shape of the pipe cleaner** that successfully moved the pencil.
- Give students time to complete the task.
- When students have finished, have them hold up their pipe cleaners for everybody to see. The pipe cleaners will likely be very similar shapes, with a single or double loop in the middle, and a handle on each end.
- Ask students, "**What shape is your pipe cleaner? What shape is your neighbor's pipe cleaner? Are there any shapes that are predominant in the class?**"
- Draw the shape(s) on the board.
- Finally, ask, "**Why do so many of the pipe cleaners have such similar shapes?**"
- Students will recognize that the **shape of the pipe cleaner allows it to perform its function**.
- Tell students that, like the pipe cleaner, proteins are folded into specific shapes to perform their functions.
- Show **Slide #2**, which shows the structure of "Pencil Transferase." Drawn on the board, the "Pencil Transferase" protein may look like the image shown in **Figure 1**.

-ase is a common suffix for enzyme names.

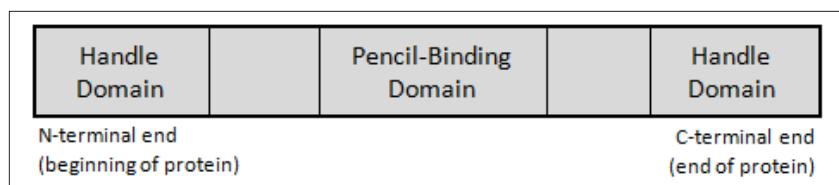
**Figure 1:** "Pencil Transferase."





16. Tell students that within “Pencil Transferase” there are areas with different structures that perform different functions. These units are called protein **domains**. Our “protein” has two handle domains and one pencil-binding domain.
17. Referring to **Slide #3**, show students the representation of the linear structure of “Pencil Transferase” on the board, as shown in **Figure 2**.

**Domain:** Specific area of a protein that performs a particular function.



**Figure 2:** “Pencil Transferase” domains.

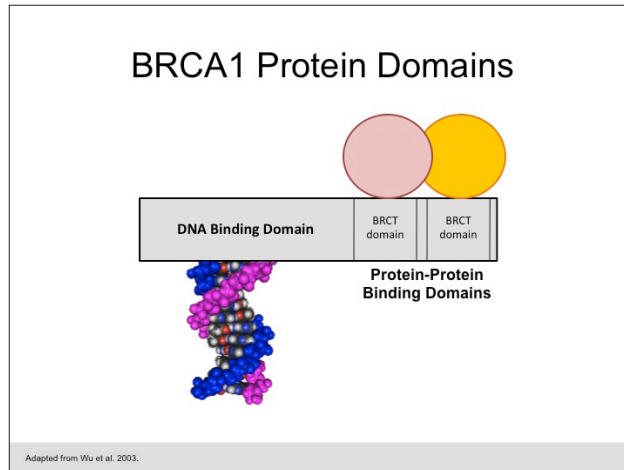
18. Point out that proteins are made up of amino acids that are held together by **covalent bonds**, and have a ‘beginning’ and an ‘end,’ labeled N-terminus and C-terminus respectively. The **N-terminus** or ‘beginning’ has an amino group ( $\text{NH}_3^+$ ), while the **C-terminus** or ‘end’ has a carboxyl group ( $\text{COO}^-$ ).
19. Show **Slide #3** or draw the representation of the linear structure of *BRCA1* on the board (see **Figure 3**). Tell students that the BRCA1 protein also has multiple domains, including a **DNA binding domain** and two **BRCT domains**. Their functions are:
  - **DNA binding domain:** To bind DNA! Remember, the function of the BRCA1 protein is to repair damaged DNA. This region binds to damaged DNA to repair it.
  - **BRCT domains:** BRCA1 needs to work with other proteins to repair DNA—it is one of many proteins that “cooperate” to do this. The BRCT domains facilitate protein-protein interactions involved in the DNA repair response. BRCT stands for Breast cancer C-Terminal domain (see **Figure 3**).

[**Note:** Need to move a piece of chalk instead of a pencil? Your “chalk-moving protein” will likely have similar domains. The amino acid sequence that folds itself into a loop to carry the pencil may be the same amino acid sequence that folds itself into a loop to carry the chalk in a different protein. The recurring domain units found in different proteins that move pencils, chalk, pens, or other similar shapes are called conserved domains.]

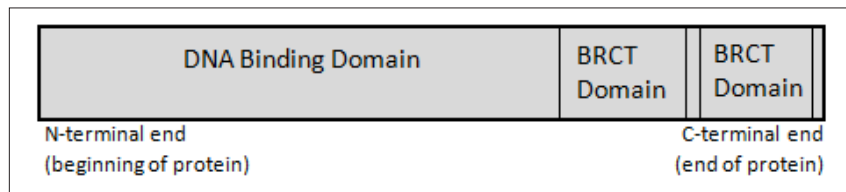
**Covalent Bonds:** A type of chemical bond that is characterized by the sharing of a pair of electrons between atoms.

**N-terminus:** The ‘beginning’ of a protein, containing an amino group ( $\text{NH}_3^+$ ).

**C-terminus:** The ‘end’ of protein, containing a carboxyl group ( $\text{COO}^-$ ).



**Figure 3:** BRCA1 domains.



[**Note:** All amino acids have the same general structure: an alpha amino group (NH<sub>3</sub><sup>+</sup>), bound covalently to an alpha carbon, with the alpha carbon also bound to a carboxyl group (COO<sup>-</sup>) and a chemical group that is unique to each amino acid called an R-group. Amino acids are held together in proteins by bonds between the alpha amino group and the carbon from the carboxyl group. This is where the terms “N-terminus” (or “amino terminus”) and “C-terminus” (or “carboxyl terminus”) come from.]

**BRCT Domain:** The Breast Cancer C-Terminal domain, a protein domain in the BRCA1 protein located at the ‘end,’ or C-terminus, of the protein that is involved in protein-protein interactions.

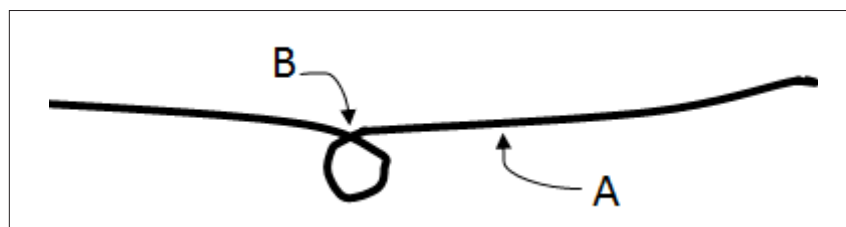
**Substitution Mutation:** When one amino acid is substituted for another as a result of mutation.

20. Inform students that the BRCA1 protein is comprised of 1,863 amino acids. In *Lesson Four*, students learned that the mutation that affects Deb (and other family members) occurs at amino acid number 1775. At position number 1775, a methionine amino acid is replaced with an arginine amino acid. This is abbreviated as a M1775R mutation.

21. Ask students, “**Knowing the position of the mutation (at amino acid number 1775 out of a protein 1,863 amino acids long), which part of the protein is most likely affected by the mutation?**”

22. Tell students that the M1775R mutation occurs near the intersection of the two BRCT domains. As such, when the class looks at the structure at NCBI, the class will only be viewing the BRCT regions, not the whole protein. Proteins as large as BRCA1 can be difficult for biochemists to work with whole, so they often crystallize one piece (i.e., one or two domains) at a time.

23. Draw students’ attention back to their pipe cleaners and tell them that different types and locations of mutations can affect the protein in different ways. For example, a **substitution mutation** (in which one amino acid is substituted for another) **may** not have a harmful effect if it is a silent mutation (does not result in a change in amino acid sequence), or if it happens inbetween functional domains, like at point (A), as shown in **Figure 4**.

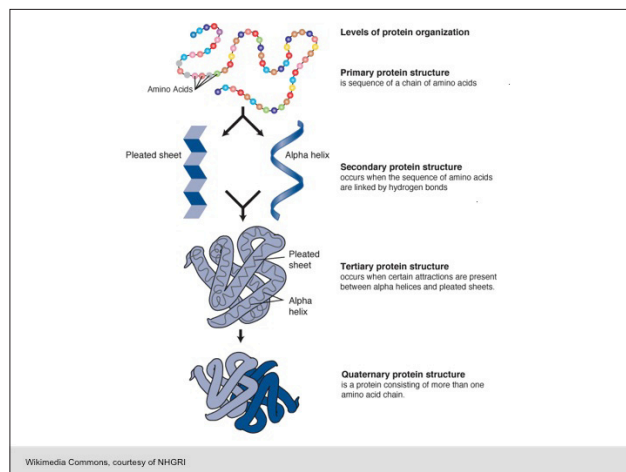


**Figure 4:** Substitution mutation on “Pencil Transferase.”

A substitution mutation at a crucial point in a functional domain, such as where the amino acids link together to make a loop (point B), may destroy the shape of the protein altogether, thus destroying its function. A **deletion** or **insertion** mutation could have similar consequences, depending on where in the protein the mutation occurs and how crucial the correct amino acid is at that point.

24. Show **Slide #4**. Briefly review levels of protein structure:

- **1° Primary:** The linear order of amino acids (the sequence of amino acids along the pipe cleaner model).
- **2° Secondary:** The alpha-helices and beta-sheets (similar to the 'loops' in the pipe cleaner model).
- **3° Tertiary:** The whole conformation or shape of the protein (including the handle domains in the pipe cleaner model) – the way the whole model folds.
- **4° Quaternary:** If the protein has more than one subunit or chain. (This can be demonstrated by putting several pipe cleaner models together.)



**[Note:** What if a point mutation in a sequence of DNA results in a premature stop codon? This causes protein translation to stop early, leading to a shorter protein.

Teachers can demonstrate a **truncation** by simply cutting off part of the pipe cleaner and asking, "How well do you think the protein would work now?"

Unfortunately, insertions and deletions may not occur in multiples of three bases, resulting in shifting of the reading frame during translation (i.e. **frameshift mutations**). This often introduces stop codons and premature truncation of the protein.]

Learning to Use Cn3D: **Slide #4**

## Part II: Seeing DNA in 3D

25. Tell students that, so far, they have been looking at the BRCA1 protein as a linear sequence of amino acids. Today they will be exploring the 3D shape of the BRCA1 protein to find out why the substitution of a single amino acid in the 1775th position has such serious consequences for the Lawler family.

26. The **BRCA1 Animation** illustrates the normal function of BRCA1, as well as the consequences of inherited mutations. Show the **BRCA1 Animation** to students. Alternatively, students can view the animation individually by visiting the Bio-ITEST website at: <http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing>.

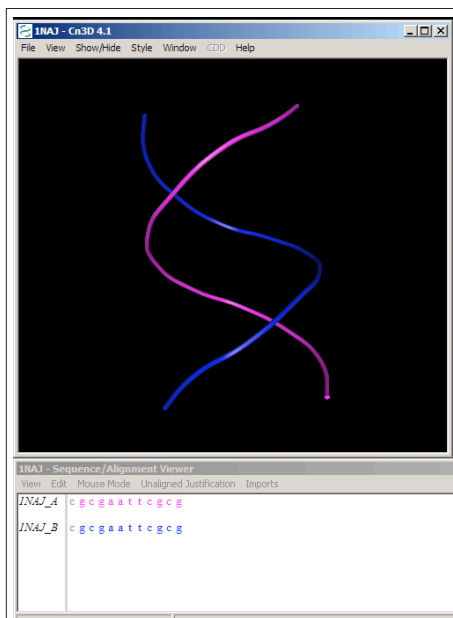
27. Tell students that they will learn how to view molecules in 3D through the use of the bioinformatics viewing tool **Cn3D**. They will learn about the program by viewing and manipulating a short DNA molecule.

**Frameshift mutation:** Also called a reading frame shift, a frameshift mutation is a genetic mutation caused by an insertion or deletion of a number of nucleotides not evenly divisible by three. Because codons are read as triplets, these insertions or deletions result in a change in the reading frame during protein translation. These reading frame shifts often result in premature stop codons and truncated proteins.

**Cn3D** (pronounced "see in 3D"): A molecular viewing software program from the National Center for Biotechnology Information. Cn3D simultaneously displays three-dimensional molecular structures along with the sequences for biological molecules such as proteins, RNA, and DNA.

28. When viewing the DNA molecule in Cn3D, students will see the anti-parallel strands, illustrated most clearly in Question #12 on Student Handout—*Seeing DNA in 3D*. When viewing the DNA structure, students may be confused by the sequence of the DNA shown in the **Sequence/Alignment Viewer** below the **Structure Window**. DNA sequences are written 5' to 3'. Therefore, both the 1NAJ\_A and 1NAJ\_B sequences are shown in the **Sequence/Alignment Viewer** as: cgcgattcgcg. (See **Figure 5**.)

**Figure 5:** Sequence/Alignment Viewer.  
Credit: Wu et al. 2003.



However, the two strands of DNA are complementary and anti-parallel in the actual DNA structure:

1NAJ\_A: 5' – cgcgattcgcg – 3'

1NAJ\_B: 3' – gcgcttaagcgc – 5'

29. After students become familiar with Cn3D, students will view the “normal” (non-mutated) version of the BRCT region of the BRCA1 protein.
30. Tell students that neither students nor scientists can look at the entire BRCA1 protein, since its structure has not yet been “solved.” They can, however, look at the BRCT domains in that protein.
31. Remind students that BRCT stands for “**breast cancer carboxy-terminal**.” Although the BRCT domain is named for the *BRCA1* gene, it is present in many proteins that function in repairing DNA, in addition to BRCA1. Each BRCT domain is about 90-100 amino acids long and has a characteristic shape.
32. Pass out Student Handout—*Instructions for Seeing DNA in 3D* and Student Handout—*Seeing DNA in 3D Worksheet*. Allow time for students to work independently on the activity.


## PART III: Teacher Demonstration—Aligning 3D Structures with VAST

33. **VAST** (Vector Alignment Search Tool) uses the locations of alpha helices and beta sheets to identify three-dimensional protein structures that might be related. Using the instructions in Teacher Resource—*Teacher Demonstration: Aligning 3D Structure with VAST*, teachers show two other protein structures associated with the BRCA1 mutations. Teachers demonstrate the structural changes between the non-mutated and mutated versions of the BRCA1 protein by aligning (overlying) the two protein structures using VAST.

**VAST:** VAST (Vector Alignment Search Tool) uses the locations of alpha helices and beta sheets to identify three dimensional protein structures that might be related. The two structures are overlaid on top of one another to compare structure similarities and differences.

**Closure: Careers in the Spotlight**

34. Today, students saw how a **single amino acid** substitution in a protein causes a very subtle shift in the protein's shape and impacts its binding to another molecule. This small change, however, has significant implications for many families, including the Lawlers. There are many mutations to BRCA1 that can cause cancer aside from the one viewed in class today. Tomorrow, students will take a closer look at the BRCA test and other genetic tests to determine how useful they may or may not be.
35. Return to the picture of the 3D animator from the *Careers in the Spotlight*, **Slide #5**.



**3D ANIMATOR**  
BETH ANDERSON

**Place of Employment:**  
Arkitek Studios

**Type of Work:**  
Develops 3D animations that help scientists and students visualize biological processes.

"Bioinformatics interfaces in many ways with our work, from data that clients give us to visualize representationally, to creating actual visual data."

Learning to Use Cn3D: **Slide #5**

36. Show **Slide #6**, which provides job information for a 3D animator. Review this information with students.



**CAREERS IN SPOTLIGHT:**  
**3D ANIMATOR**

**What do they do?**  
3D animators are visual communicators and designers. The animations and illustrations they create help communicate scientific content and ideas to scientists, students, and educators.

**What kind of training is involved?**  
Associate of Arts degree. Some also have training in biology, illustration, graphic design, and computer animation.

**What is a typical salary for a Research Scientist?**  
Ranges with experience from very little when people are first starting out to \$100,000/year (\$48/hour) or more.

Learning to Use Cn3D: **Slide #6**

37. Ask students, "**What more do we know about 3D Animators after today's lesson?**" Point out that 3D animators in the biological sciences use computer programs like Cn3D and the information in protein crystal structures like the ones seen here for BRCA1 to create animations that help us visualize biological processes. Beth Anderson and her colleagues developed the BRCA1 animation seen in this lesson.
38. Ask students to answer 3D Animator Question #2 on their *Careers in the Spotlight* handout, which has students explain how this lesson has changed their understanding of the kind of work a 3D animator does.

39. Ask students to also answer 3D Animator Question #3 on their *Careers in the Spotlight* handout, which has students explain how a 3D animator might use bioinformatics in his or her work.
40. Tell students to keep their *Careers in the Spotlight* handout available for future lessons.

## Homework

The following are suggested homework activities to follow this lesson:

- A. Students should continue to prepare for the Socratic Seminar in Lesson Six using Student Handout—*Categorizing Genetic Tests* and/or Student Handout—*Weighing the Risks and Benefits of Direct-to-Consumer Genetic Testing*. Either Student Handout can also be passed out as homework to accompany the reading, which can be used as entry tickets for that class session. These are the reading and support materials for the Socratic Seminar students will participate in during *Lesson Six*. Students may need two days to prepare fully.
- B. If students answered the Extension Questions #48-50 on Student Handout – *Aligning Sequences with BLAST Worksheet* in Lesson Four, they can expand upon their answers to these questions, describing what they have learned about how this single amino acid change can impact the structure of the BRCT domain of the BRCA1 protein. Students can also answer the questions posed on their Word® document created in Lesson Four and augmented in Lesson Five as homework (see Student Handout—*Instructions for Seeing DNA in 3D* Part 2, Questions #18 and 19).
- C. As a reflective exercise, ask students to write about the activities they did in *Lesson Four* in their lab notebooks, on another sheet of paper, or in a word processing program like Microsoft Word® or Google Docs which they then provide to the teacher as a printout or via email. This can serve as an entry ticket for the following class. Have them complete these prompts:
  - a. Today I learned that...
  - b. An important idea to think about is...
  - c. Something that I don't completely understand yet is...
  - d. Something that I'm really confident that I understand is...

[Note: Suggested scoring for reflection: +5 points if all 4 prompts are complete.]

**X-ray crystallography:** A method of determining the arrangement of atoms within a crystal, such as a crystal of a particular protein, in which a beam of X-rays strikes a crystal and deflects or diffracts into many specific directions. From the angles and intensities of these diffracted beams, a crystallographer can produce a three-dimensional picture of the density of electrons within the crystal, and thus calculate and estimate the three-dimensional shape of the molecule used to generate the crystals. This is often referred to as determining or “solving” the protein’s structure.

## Extension

- For a musical analogy with MP3 clips to describe protein structure and conserved domains, a draft extension activity for *Lesson Five* can be found at the end of this lesson plan on Teacher Resource—*From Sequence to Structure*.

## Adaptation

- Teachers can work through all or part of Student Handout—*Instructions for Seeing DNA in 3D* as a teacher-led demonstration, as desired.

## Teacher Background

**Protein Structures:** Scientists know what DNA, a protein, or other macromolecules look like by using a number of different tools, including **X-ray crystallography** and **nuclear magnetic resonance (NMR)**, to determine molecular structures. In the case of x-ray crystallography, proteins are crystallized before being bombarded with x-rays. By measuring how the crystals deflect or diffract the x-rays, scientists can determine or “solve” the protein’s structure. With NMR, molecules are subjected to a strong magnetic field that causes the natural magnets in atomic nuclei to spin in the same direction, just like iron filings line up in the presence of a magnet. Radio waves are used to disrupt this state and scientists measure how each atom responds. The characteristics of the response are used to solve the molecular structure by determining which other atoms are close by. These structures are then added to the **Protein Data Bank (PDB)**, and later, to the structure database at the NCBI for others to view using special molecular-viewing programs like Cn3D. Molecular-viewing programs like Cn3D allow users to view and manipulate three-dimensional structures on a computer screen.

**NMR spectroscopy:** Nuclear Magnetic Resonance spectroscopy, usually abbreviated as “NMR,” is a technique used to determine the three-dimensional structure of molecules, including proteins. Molecules are subjected to a strong magnetic field that causes the natural magnets in atomic nuclei to spin in the same direction, just like iron filings line up in the presence of a magnet. Radio waves are used to disrupt this state, and scientists measure how each atom responds. The characteristics of the response are used to determine or “solve” the molecular structure by determining which other atoms are close by.

**Protein Data Bank (PDB):** A repository or collection of three-dimensional structures of large biological molecules, including proteins and nucleic acids, submitted by scientists from around the world. This data is typically obtained by X-ray crystallography or NMR spectroscopy.

## Assessment Suggestions

Students can be assessed on their answers to the questions posed on their Word® document created in *Lesson Four* and augmented in *Lesson Five* (see Student Handout —*Seeing DNA in 3D*).

## Glossary

**Accession number:** A unique identifier or code assigned to every entry in the National Center for Biotechnology Information (NCBI) databases. This unique code can be used to search the databases to find your gene, protein, or structure of interest.

**Alpha helix** (Plural: “alpha helices”): A common structure of proteins, characterized by a single, spiral chain of amino acids stabilized by hydrogen bonds.

**Beta sheet:** A structure that occurs in many proteins and consists of two or more parallel adjacent polypeptide chains arranged so that hydrogen bonds can form between the chains.

**BRCT domain:** The Breast Cancer C-Terminal domain, a protein domain in the BRCA1 protein located at the ‘end,’ or C-terminus, of the protein that is involved in protein-protein interactions.

**Cn3D** (pronounced “see in 3D”): A molecular viewing software program from the National Center for Biotechnology Information. Cn3D simultaneously displays three-dimensional molecular structures along with the sequences for biological molecules such as proteins, RNA, and DNA.

**Covalent bonds:** A type of chemical bond that is characterized by the sharing of a pair of electrons between atoms.

**C-terminus:** The ‘end’ of protein, containing a carboxyl group (COO<sup>-</sup>).

**Domain:** Specific area of a protein that performs a particular function.

**DNA binding domain:** Specific area of a protein (domain) that binds to DNA. The DNA binding domain in the BRCA1 protein is necessary for the BRCA1 protein to repair damaged DNA.

**Frameshift mutation:** Also called a reading frame shift, a frameshift mutation is a genetic mutation caused by an insertion or deletion of a number of nucleotides not evenly divisible by three. Because codons are read as triplets, these insertions or deletions result in a change in the reading frame during protein translation. These reading frame shifts often result in premature stop codons and truncated proteins.

**Hydrophilic:** A substance that is attracted to water. From the Greek “hydro” which means water, and “philos” meaning love. A synonym for polar. The opposite of hydrophobic or non-polar.

**Hydrophobic:** A substance which repels water. From the Greek “hydro” which means water, and “phobos” which means fear. A synonym for non-polar. The opposite of hydrophilic or polar.

**NMR spectroscopy:** Nuclear Magnetic Resonance spectroscopy, usually abbreviated as “NMR,” is a technique used to determine the three-dimensional structure of molecules, including proteins. Molecules are subjected to a strong magnetic field that causes the natural magnets in atomic nuclei to spin in the same direction, just like iron filings line up in the presence of a magnet. Radio waves are used to disrupt this state, and scientists measure how each atom responds. The characteristics of the response are used to determine or “solve” the molecular structure by determining which other atoms are close by.

**Non-polar:** A substance that repels water. A synonym for hydrophobic.

**N-terminus:** The ‘beginning’ of a protein, containing an amino group ( $\text{NH}_3^+$ ).

**Polar:** A substance that is attracted to water. A synonym for hydrophilic. The opposite of hydrophobic or non-polar.

**Protein Data Bank (PDB):** A repository or collection of three-dimensional structures of large biological molecules, including proteins and nucleic acids, submitted by scientists from around the world. This data is typically obtained by X-ray crystallography or NMR spectroscopy.

**Stop codon:** A codon (series of three nucleotides in a row) that terminates, or stops, protein translation.

**Substitution mutation:** When one amino acid is substituted for another as a result of mutation.

**Truncation:** To shorten, as if by cutting off. During translation, a growing protein chain is truncated if it encounters a premature stop codon.

**VAST:** VAST (**V**ector **A**lignment **S**earch **T**ool) uses the locations of alpha helices and beta sheets to identify three dimensional protein structures that might be related. The two structures are overlaid on top of one another to compare structure similarities and differences.

**X-ray crystallography:** A method of determining the arrangement of atoms within a crystal, such as a crystal of a particular protein, in which a beam of X-rays strikes a crystal and deflects or diffracts into many specific directions. From the angles and intensities of these diffracted beams, a crystallographer can produce a three-dimensional picture of the density of electrons within the crystal, and thus calculate and estimate the three-dimensional shape of the molecule used to generate the crystals. This is often referred to as determining or “solving” the protein’s structure.

## Credit

BRCA1 domain illustration from: <http://www.biochemsoctrans.org/bst/037/0597/bst0370597a01.gif>.

*BRCA1 Animation* developed by Beth Anderson, Arkitek Studios, and Jill DelSordi.

Anderson, Beth. Personal Interview. 2 July 2010.

Photo of Beth Anderson provided by Doug Huff.

Wu, Z., Delaglio, F., Tjandra, N., Zhurkin V.B., & Bax, A. (2003). Overall structure and sugar dynamics of a dna dodecamer from homo- and heteronuclear dipolar couplings and (31)p chemical shift anisotropy. *J. Biomol. Nmr.*, 26, 297.

Varma, A.K., Brown, R.S., Birrane, G., & Ladias, J.A.A. (2005). Structural basis for cell cycle checkpoint control by the brca1-ctip complex. *Biochemistry*, 44, 10941.

The authors wish to thank Wikimedia Commons for the definitions of some of the vocabulary terms found in the Glossary and throughout this lesson.

# 5

## Instructions for Seeing DNA in 3D

**Aim:** To understand that protein structure can impact protein function, using the bioinformatics tool Cn3D to visualize molecules.

**Instructions:** Write the answers to your questions on Student Handout—*Seeing DNA in 3D Worksheet*, in your lab notebook, or on a separate sheet of paper, as instructed by your teacher.

### PART I: Viewing DNA Structure

1. Go to the NCBI Structure database: (<http://www.ncbi.nih.gov/structure>).
2. The center top of the page has an open area for search terms. Type **1NAJ** into the search area and click the **Go** button.

We will begin our investigation of 3D structure with a look at a molecule you are already familiar with: DNA. "1NAJ" is the **accession number** for a file that contains structure information for a small piece of double-stranded DNA. An accession number is like a catalog number or bar code; it bears no resemblance to the product itself, but allows you to access information in the vast databases at the NCBI.

3. The 1NAJ record will appear as the only result from this search. **Click** the DNA image that looks like two colored strings winding loosely around each other with a black background.

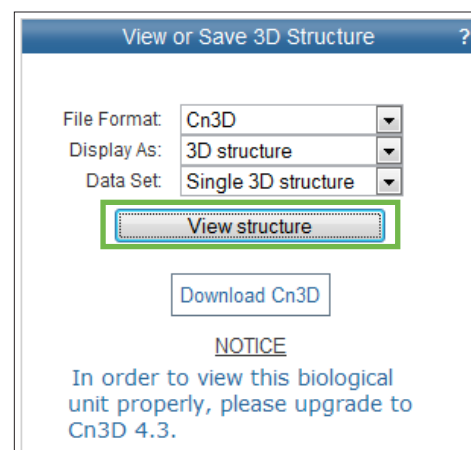
You are on the **Structure Summary** page, which includes a brief description of the source of the structure.

4. Click the "**View structure**" button to download the structure file, as shown in **Figure 1**. Save this file to your desktop, or to a folder, as instructed by your teacher. Then double-click on the file to open it.

[**Note:** The file may appear in different places depending on the type of web browser and how the computer has been set up. You may need to look for your file in your Downloads folder. This file may have a name like "mmdb.cgi."]

5. If your file opens with the Cn3D program, skip to Step #7. Otherwise, find the Cn3D program icon on your desktop or in your program menu, and open the program. Select "**Open**" from the file menu and select your file from above.

[**Note:** The Structure Window is the black square box that shows the DNA structure. Your screen may also show the Sequence/Alignment viewer, below the Structure Window.]



**Figure 1:** Click the "View structure" button to download the file.

- View the DNA in the Structure Window. Drag the structure with your mouse to turn it around. Hold the shift button down, and continue to hold it down, while dragging, to move your structure without turning it. Experiment with controlling the movement and viewing the molecule from multiple angles.

You can zoom in and out by using the **z** or **x** keys. Alternatively, you can change the size by holding down the left mouse button and dragging the mouse away from the screen.

Hold down the left mouse button and drag the mouse towards or away from the screen.



- Open the **View** menu and choose "**Animation.**" Choose the "Spin" option.

What happens when you click s and n?

Position the DNA molecule the way you want it before proceeding.



- Open the **Style** menu and select "**Rendering Shortcuts.**" There are a variety of rendering options available to view the DNA molecule. Explore each of the following:

**Worms      Tubes      Wire      Ball and Stick      Space-fill**

Imagine you are teaching a class about DNA. What rendering option(s) would you use to teach students about DNA structure? Explain the reasons for your decision.



- Open the **Style** menu and select "**Coloring Shortcuts.**" There are many options for changing the colors of molecules. Select each of the options below, and answer each of the following questions.

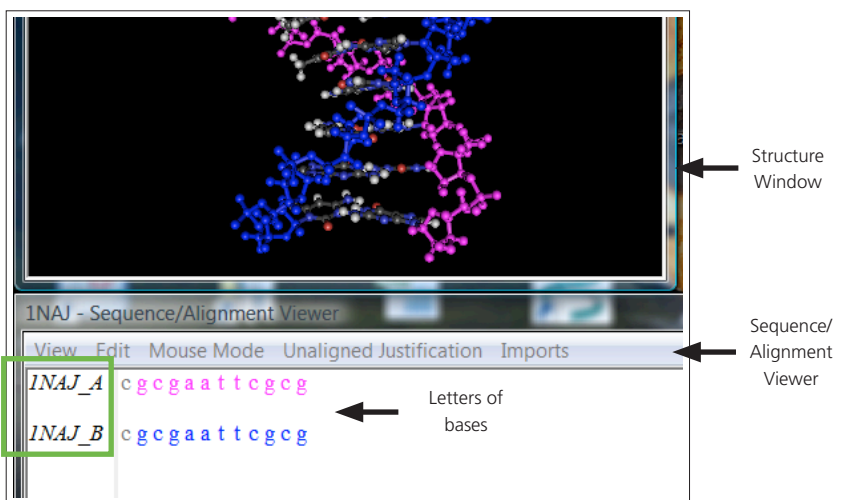
- Object:** How did this change the coloring of the DNA molecule?
- Rainbow:** 'Rainbow' uses the color red at the start (5' end) and continues through the rainbow. Why are there two red regions?
- Explore a few combinations using different rendering and coloring options. Which coloring option do you find most useful? Why?

- Go to the **Sequence/Alignment Viewer** window on your screen.

[**Note:** The Sequence/Alignment Viewer is the window below the Structure Window that shows the sequence being viewed, as shown in **Figure 2**. If there is no sequence visible, open the **Window** menu and choose "**Show Sequence Viewer.**" ]

Click on the letter of a base in the Sequence/Alignment Viewer and see what happens.

Note that there are two sequences: *1NAJ\_A* and *1NAJ\_B*. Each sequence corresponds to a different strand of the DNA molecule.



**Figure 2:** The Sequence/Alignment Viewer shows the sequences being viewed.  
Credit: Wu et al. 2003.



11. In the sequence *1NAJ\_A*, click the first “g.” Next, click the first “g” of the sequence *1NAJ\_B*.

Where are these two guanines located on the DNA molecule? Explain why.

12. Choose a single base you would like to view, and click on it in the Sequence/Alignment viewer.



13. Change the **Rendering style** to “**Wire**” and the **Coloring style** to “**Object**.”

a. What does that do? Explain.

b. Move the DNA so that the ring structure(s) of the selected nucleotide’s base is clearly seen.

What do you see?



14. Open the **Select** menu and choose “**Show Selected Residues**.” Answer the following questions:

a. What do you see?

Zoom in to view the base more easily (refer to Step #7 if you need help).

b. **Challenge question:** Is the base you selected a purine or a pyrimidine, and how do you know?

From the Style menu choose “**Show Everything**” to see the complete strand.

## PART II: Viewing the BRCA1 Protein

Now that you are familiar with the Cn3D program, we will view part of the BRCA1 protein and part of a second protein that it interacts with.

15. Go to the NCBI **Structure database**: (<http://www.ncbi.nih.gov/structure>) or click the “**Structure Home**” button from the DNA structure summary page.

16. At the top of the page, find the search option. Type “**1Y98**” into the blank and click the “**Go**” button.

17. The **1Y98** record will appear as the only result from this search. Click on the **link or title** of the structure above the black box with the picture of the protein in it. You are on the **Structure Summary** page, which includes a brief description of the source of the structure.

18. Click the “**View structure**” button to download and open the structure file.

19. Take a moment to explore different **Style** options.

20. Open the **Style** menu and choose “**Coloring Shortcuts → Secondary Structure**” to highlight each type of secondary structure in a different color.

21. Look at the **Sequence/Alignment Viewer** below the structure. When viewing the DNA structure, the Alignment/Sequence Viewer showed the letters of the bases. When viewing a protein structure, the viewer shows the one-letter abbreviation for the amino acids.

Note that colors of the amino acid sequence correspond to the colors of the structure.

22. The structure contains two BRCT domains. To see each domain individually, open the **Style** menu and choose “**Coloring Shortcuts → Domains**.”

This will show the two BRCT domains in different colors. **BRCT** stands for **B**reast cancer **C**-Terminal domain, and is involved in protein-protein interactions. The alpha helices are depicted as cylinders (like crayons) and the beta sheets are depicted as flat arrows. Both the helix and strand objects point in an amino to carboxyl direction (i.e., they point toward the end of the protein).

Note that there is a third domain – this is part of another, different protein to which BRCA1 normally binds to do its job of repairing DNA. This protein is called CtIP.

23. Find the portion of the second protein, CtIP, that BRCA1 interacts with during DNA damage repair.

It looks like a little brown squiggle off to the side. In the Sequence Window, the sequence of this portion of the second protein is labeled 1Y98B. If you can't find CtIP, highlight the 1Y98B sequence and the CtIP will light up yellow in the Structure Viewer. To remove the yellow highlights, click somewhere in the Sequence Window to activate the sequence menu. Then, open the **View** menu and choose "**Clear Highlights.**"



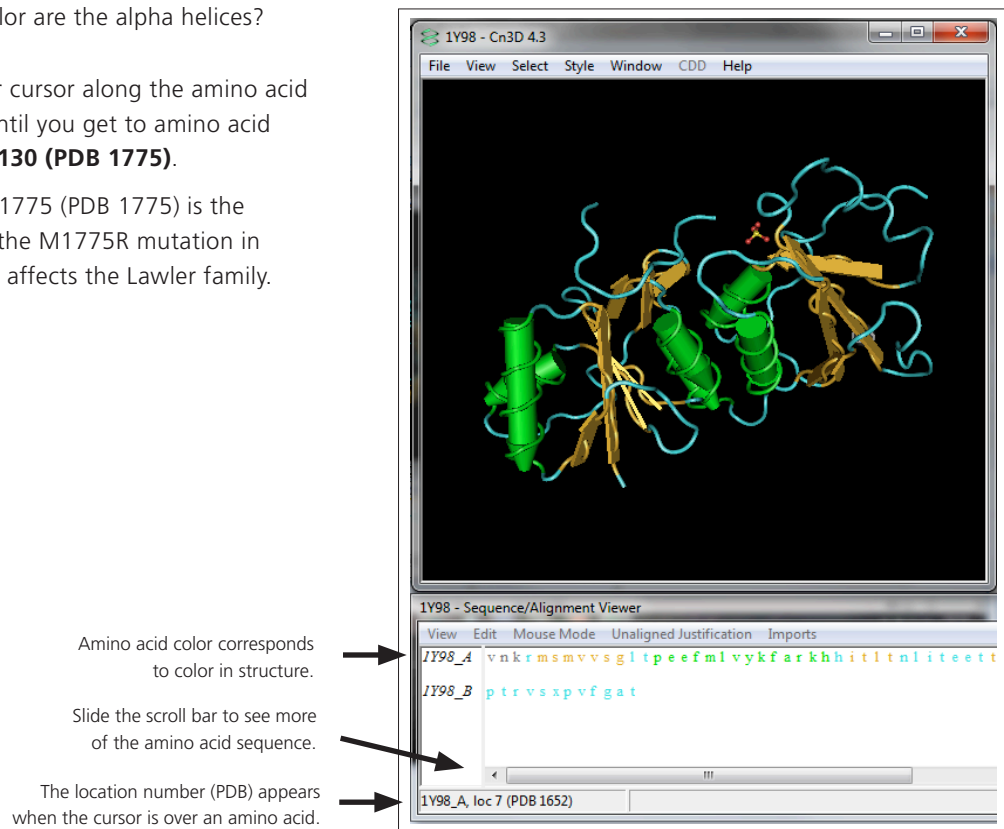
24. Now that you know what the portion of the CtIP protein looks like, go back to the **Style** menu and select "**Coloring Shortcuts → Secondary Structure.**" Move your cursor along the 1Y98A amino acid sequence in the Sequence/Alignment Viewer, as seen in **Figure 3.**

Note that the colors of the amino acids correspond to the colors in the protein structure (beta sheets, alpha helices, and random coils). The box on the lower left of the screen shows the location of the amino acid your cursor is over.

- What color are the beta sheets?
- What color are the alpha helices?

25. Move your cursor along the amino acid sequence until you get to amino acid **1Y98A\_loc130 (PDB 1775).**

Amino acid 1775 (PDB 1775) is the location of the M1775R mutation in BRCA1 that affects the Lawler family.



**Figure 3:** When you open the BRCA1 structure, you will see the BRCA1 protein that does not contain any mutations.  
Credit: Varma et al. 2005.

26. Click on amino acid number “**1775**” (M, the amino acid methionine) in the Sequence Viewer to highlight it in the structure.

It should light up yellow in the structure. You are looking at where the mutation would be located, but this structure you are looking at does not have the mutation. (Recall that the mutation converts a M to a R.)

27. Open the **Select menu → Select by Distance**. Open the **Select menu → Select by Distance → Residues only**. Leave the distance cutoff at five Angstroms to select all molecules within a radius of five Angstroms. Make sure the options “Select protein residues” and “Select other molecules only” are checked. Click “OK.”

[**Note:** An Angstrom, Å, is  $1 \times 10^{-10}$  meters.]

28. Note that part of the CtIP protein now also lights up yellow, as do other BRCA1 amino acids (residues). It is very close to the site on BRCA1 where the mutation would be. You can imagine how a change in BRCA1 might impact its ability to interact with this second protein, CtIP, which is also required for DNA repair.

29. Record your results from Cn3D by capturing the image using the following instructions:

- Open up the Word® document you created in *Lesson Four*. It should be labeled with your LASTNAME\_BRCA1\_NCBI.
- Return to the 1Y98A Cn3D page with the protein structure.
- FOR PC USERS:** Hit the “Prnt Scrn” button on your computer. This is often in the top right of the keyboard, to the right of the F12 button.

**FOR MAC USERS:** Press the keys: “**Command + Shift + 4**” at the same time.

- Return to your Word® document.

- FOR PC USERS:** Use the paste function to transfer the screen shot.

**FOR MAC USERS:** Open the **Insert** menu, choose “**Picture from file,**” and choose the image that you captured from the computer screen.

[**Note:** The image may have saved to your desktop.]

- Type these instructions and questions at the bottom of your Word® document:

- Circle the location of the mutation in the picture of the BLAST alignment.
- Circle the location of the mutation in the Cn3D picture of the protein structure.
- Explain what these pictures represent.
- Explain how the pictures are connected to each other.

- Follow the instructions you typed above and answer the questions on your Word® document. You may choose to print out the document later and circle the points of mutation, or you can make the circles directly on your Word® document by doing the following:

Use your drawing tools in Word® to draw red circles around the mutation in the BLAST sequence window and the Cn3D view. (**Insert Shapes → Oval → Shape Fill: no fill, Shape outline: red line**). You can also use the “Crop” feature in Word® to crop your image.

- Save and close this document. Transfer it to a thumb drive or email it to yourself if you will not have access to this computer in the future.



30. **Challenge:** To get a better sense of the interaction between BRCA1 and CtIP, Cn3D can show the amino acids that are interacting with each other in ball and stick form. Be sure that the amino acids in the previous step are highlighted before you begin.

- Open the **Style** menu and choose "**Annotate →**." Click on the box that says "**New**" to the right of "**Selection: (new)**." Another small Edit Annotation window will open.
- Click on "**Edit Style**" in the Edit Annotation window. There is no need to enter a description. A big Style Options window will open and will default to the "**Settings**" tab.
- Find the group **Protein Sidechains**, and click on the box next to it to check it.
- Immediately to the right, pull down the "**Rendering Options for Protein Sidechains**" and change them to "**Ball and Stick**."
- Click "**Apply**" and "**Done**" (and "**OK**" and "**Done**" on the remaining windows) to close all the windows and return to the structure.
- You may want to capture this image for your Word® document as well.



- Challenge Question:** What can you see now that you could not see before you annotated this structure? Does this help you understand the consequences of the Lawlers' M1775R mutation?

31. **Optional:** If time permits, you may wish to experiment with protein structures 1JNX and 1N5O. Make sure that you enter *capital O and not a zero* for "**1N5O**."

**1JNX:** This is the accession number for the BRCA domains of the **non-mutated** version of the BRCA1 protein.

**1N5O:** This is the accession number for the BRCA domains of the M1775R **mutation** of the BRCA1 protein.

Name \_\_\_\_\_ Date \_\_\_\_\_ Period \_\_\_\_\_

# 5

## Seeing DNA in 3D Worksheet

**Aim:** To understand that protein structure can impact protein function, using the bioinformatics tool Cn3D to visualize molecules.



**Instructions:** Use Student Handout—*Instructions for Seeing DNA in 3D* to complete this worksheet.

### PART I: Viewing DNA Structure

7. What happens when you click s and n?

s:

n:

8. There are many ways to view the DNA. Imagine you are teaching a class about DNA. What rendering option(s) would you use to teach students about DNA structure?

**(circle one):**    **Worms**    **Tubes**    **Wire**    **Ball and Stick**    **Space-fill**

Explain the reasons for your decision:

9a. Look at **Coloring Shortcuts**, and click "**Object**." How did this change the coloring of the DNA molecules?

9b. Now look at **Rainbow**: "Rainbow" uses the color red at the start (5' end) and continues through the rainbow. Why are there two red regions of the DNA?

9c. Which coloring option do you find most useful? Why?

# LESSON 5

## HANDOUT

11. When you click on the two guanines that are in the same place for both DNA strands, where are they located on the DNA molecule? Explain why.

13a. After changing the **Rendering style** to “Wire” and the **Coloring style** to “Object,” what does that do? Explain.

13b. Move the DNA so that the ring structure(s) of the selected nucleotide's base is clearly seen. What do you see?

14a. After selecting “**Show Selected Residues**,” what do you see?

14b. **Challenge question:** Is the base you selected a purine or a pyrimidine, and how do you know?

### PART II: Viewing the BRCA1 Protein

24a. What color are the beta sheets?

24b. What color are the alpha helices?

30. **Challenge Question:** What can you see now that you could not see before you annotated this structure? Does this help you understand the consequences of the Lawlers' M1775R mutation?

# 5

## Seeing DNA in 3D

### Teacher Answer Key

[**Note:** Suggested point values are included after each question, and are intended to provide general guidelines for the weight each question could be given. Using these suggested point values, the total value for this worksheet is **20 points, plus 4 points for the optional Challenge questions**, (Questions #14B and #31).]

### PART I: Viewing DNA Structure

7. What happens when you click s and n?

s: The image stops spinning.  
n: The image begins to spin again.  
(+1 for each correct response.)

8. There are many ways to view the DNA. Imagine you are teaching a class about DNA. What rendering option(s) would you use to teach students about DNA structure?

(circle one):   **Worms**   **Tubes**   **Wire**   **Ball and Stick**   **Space-fill**

Explain the reasons for your decision:

Student answers will vary. Some students will prefer the simplified representation of the sticks or the worms, while others will prefer to see the structure details of the wires or ball and stick (which are both good for visualizing the rings of the DNA bases), or the space-filled, which makes it possible to view the full size of the atoms, including a representation of the electron clouds.  
(+1 for selecting option; +1 for providing reasonable explanation.)

9a. Look at **Coloring Shortcuts**, and click "**Object**." How did this change the coloring of the DNA molecules?

The entire DNA molecule (the 'object') turns pink. (+1.)

9b. Now look at **Rainbow**: "Rainbow" uses the color red at the start (5' end) and continues through the rainbow. Why are there two red regions of the DNA?

Each of the two regions is a different strand of the DNA molecule.  
(+1 for 'different strands' or 'two strands' or 'DNA is double stranded'.)

9c. Which coloring option do you find most useful? Why?

Student answers will vary. Some students will prefer the simplicity of the Object coloring, while others will find the Rainbow coloring useful to identify the 5' and 3' ends of each DNA strand.  
(+1 for choosing option; +1 for providing reasonable explanation.)

11. When you click on the two guanines that are in the same place for both DNA strands, where are they located on the DNA molecule? Explain why.

The two guanines are located on opposite strands of DNA, so in the three-dimensional DNA structure, one guanine is located on the top (near the 5' end of one strand), and one is located on the bottom (near the 5' end of the second strand).

(+1 for providing location; +1 for explaining that they are on opposite strands.)

13a. After changing the **Rendering style** to “**Wire**” and the **Coloring style** to “**Object**,” what does that do? Explain?

Most of the DNA molecule is shown in pink, while the DNA base the student selected is highlighted in yellow, making this base easier to identify in the DNA structure.

(+1 for color change; +1 for explaining that the base is easier to identify.)

13b. Move the DNA so that the ring structure(s) of the selected nucleotide's base is clearly seen. What do you see?

You can visualize the structure of the base: the backbone, the ring, and the phosphodiester bond.

(+1 for structure or shape of the base.)

14a. After selecting “**Show Selected Residues**,” what do you see?

Most of the DNA structure has been removed, and only the selected base is visible. (+1.)

14b. **Challenge question:** Is the base you selected a purine or a pyrimidine, and how do you know?

The selected base is either a purine (if it has two rings) or a pyrimidine (if it has one ring).

(+1 bonus point for correct response; +1 bonus point for # of rings.)

## PART II: Viewing the BRCA1 Protein

24a. What color are the beta sheets?

The beta sheets are gold/yellow (flat arrows). (+1.)

24b. What color are the alpha helices?

The alpha helices are green (round or tube-like arrows). (+1.)

29. You will need to open the document that you saved your BLAST Alignment to (the DNA and Protein sequence). Follow the instructions on Student Handout—*Instructions for Seeing DNA in 3D*.

Students should include a screen capture image of the BRCA1 protein structure. On this screen capture image, they should circle the location of the M1775R mutation on the BLAST alignment, as well as the location of the mutation on the Cn3D picture of the protein structure (circled in red in the sample image). They should explain that the BLAST picture represents an alignment of the Lawler family protein sequences compared to a reference sequence, which shows the location of the mutation (a methionine to arginine mutation at position 1775). The Cn3D image is a picture of part of the structure of the BRCA1 protein, which also has a circle around the location of the M1775R mutation found in the Lawler family.



Credit: Varma et al. 2005.

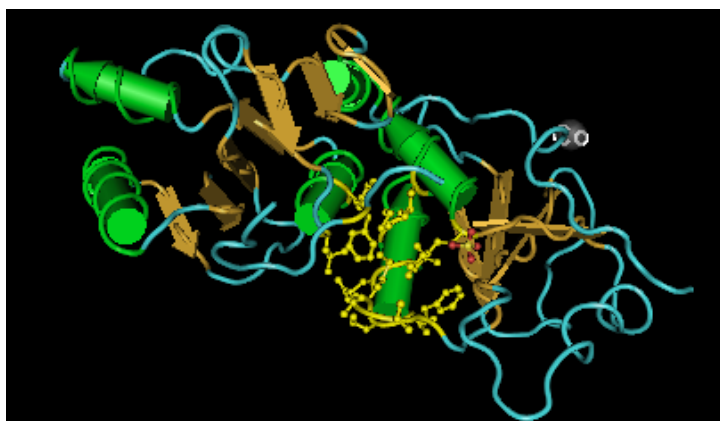
**[Note:** Some protein structures contain additional molecules, such as  $\text{SO}_4$  or metal ions, which assist in protein crystallization but are not part of the protein structure.]

(+1 for screen capture; +1 for circling mutation; +1 for explaining the BLAST picture is the protein sequence with mutation; +1 for explaining the Cn3D picture represents the protein structure with mutation. 4 points total.)

31. **Challenge question:** What can you see now that you could not see before you annotated this structure? Does this help you understand the consequences of the Lawlers' M1775R mutation?

Students can now see the methionine impacted by the Lawler family mutation, as well as which amino acids methionine 1775 interacts with, including the CtIP protein. If methionine is mutated to an arginine at position 1775, interactions with multiple amino acids in the BRCA1 protein will be impacted, as seen in the screen capture image to the right.

(+1 bonus for methionine; +1 bonus for explaining multiple interactions.)



Credit: Varma et al. 2005.

# LESSON 5

## KEY

# 5 Teachers Demonstration:

## Aligning 3D Structures with VAST

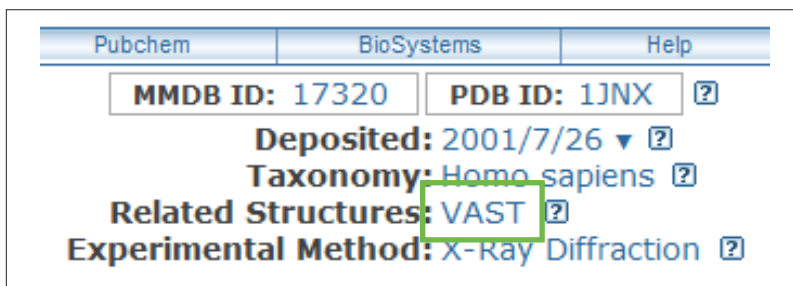
**VAST** is a **V**ector **A**lignment **S**earch **T**ool. It uses the locations of alpha helices and beta sheets to identify three dimensional protein structures that might be related.

In this demonstration, teachers show the structural changes between the non-mutated and mutated versions of the BRCA1 protein by aligning two structures. The accession numbers for the two proteins are:

**1JNX**—This shows the BRCA domains of the **non-mutated** version of the BRCA1 protein.

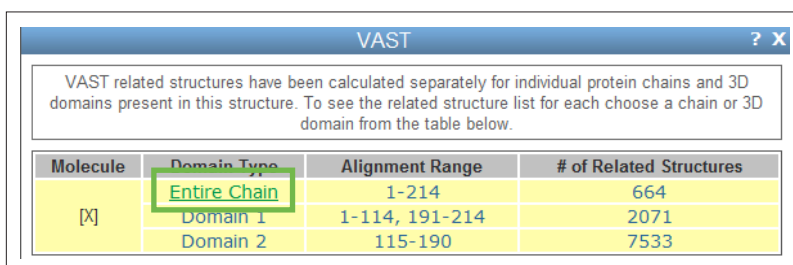
**1N50**—This shows the BRCA domains of the M1775R **mutation** of the BRCA1 protein.

1. Go to the NCBI Structure database: (<http://www.ncbi.nih.gov/structure>).
2. The center top of the page has a search box next to the drop down menu that says **Structure**. Type "**1JNX**" into the blank box and click the **Search** button.
3. The 1JNX record will appear as the only result from this search. **Click the image** to get to the **Structure Summary** page.
4. Near the top right corner of the screen, click the VAST link to search for related structures identified by VAST, as shown in **Figure 1**.



**Figure 1:** Click the VAST link near the top right corner of the screen. Credit: NCBI Structure, Molecular Modeling Database (MMDB).

5. When the VAST window opens, click on the "**Entire Chain**" link under **Domain Type** to find structures related to the full-length BRCT protein structure, as shown in **Figure 2**.



**Figure 2:** Select the "Entire Chain" option to view the entire BRCT domain. Credit: NCBI Structure, Molecular Modeling Database (MMDB).

6. To focus your search options on the 1N5O structure, which we know to be related to 1JNX, click on the + next to the phrase **Advanced related structure search** at the bottom of the yellow box in the top portion of the page (see **Figure 3**).

**Figure 3:** To expand your search options, click on the “+.” Credit: NCBI Structure, Molecular Modeling Database (MMDB).

7. Enter “1N5O” in the box labeled **related structures with MMDB or PDB ids**: Make sure the last character is a *capital O* and not a zero. Click the **Find** button (see **Figure 4**).

**Figure 4:** Enter your search term “1N5O” in the search box. Credit: NCBI Structure, Molecular Modeling Database (MMDB).

8. Check the box beside the first structure, **1N5O X**.
9. Click the gray **View Sequence Alignment** button near the top of the screen to see where the mutation is located (see **Figure 5**).
- The format is similar to results of the BLAST in *Lesson Four*: the two sequences are shown in one color (red) in areas where they are the same, and a different color (blue) where they differ. Lori’s protein (1JNX) contains a methionine, while Deborah’s protein (1N5O) contains an arginine at the same position (the M1775R mutation).

**Figure 5:** Viewing the Sequence Alignment will provide similar information as a BLAST alignment. Credit: NCBI Structure, Molecular Modeling Database (MMDB).

10. Click the **back** button on your browser to return to the previous window.

11. Click the **View 3D alignment** button (above the “**View Sequence Alignment**” button) to download a file containing the superimposed structures to your computer. Click “**Open with Cn3D**” to view the file directly with Cn3D, or save the file to your desktop and open it from there. The file may have a name like “vast.cgi” or “vast.cn3.”

The two structures are now superimposed. The mutation site appears in blue in the Sequence Viewer window while the regions of the three dimensional protein structures with identical amino acid sequences are shown in red. Amino acids shown in gray lowercase letters were not solved (or crystallized) as part of the protein structures.

12. **Click** in the structure window to activate the “**Structure**” menu.

13. Use the right and left arrow keys to switch between the views of the two protein structures.

14. From the **Style** menu, go to **Rendering Shortcuts** and select “**Space Filled.**”

Move the molecule around with the cursor, occasionally using the arrow keys to move from one structure to the next one. Find a view that you think best shows the change in structure between molecules.

15. From the **Style** menu, select “**Edit Global Style,**” and then **uncheck** the box in the “**Show**” column for the “**Heterogens**” row (see **Figure 6**). Click “**Apply**” and then “**Done.**” This will hide the heterogens. Heterogens are molecules that are neither RNA, DNA, nor proteins.

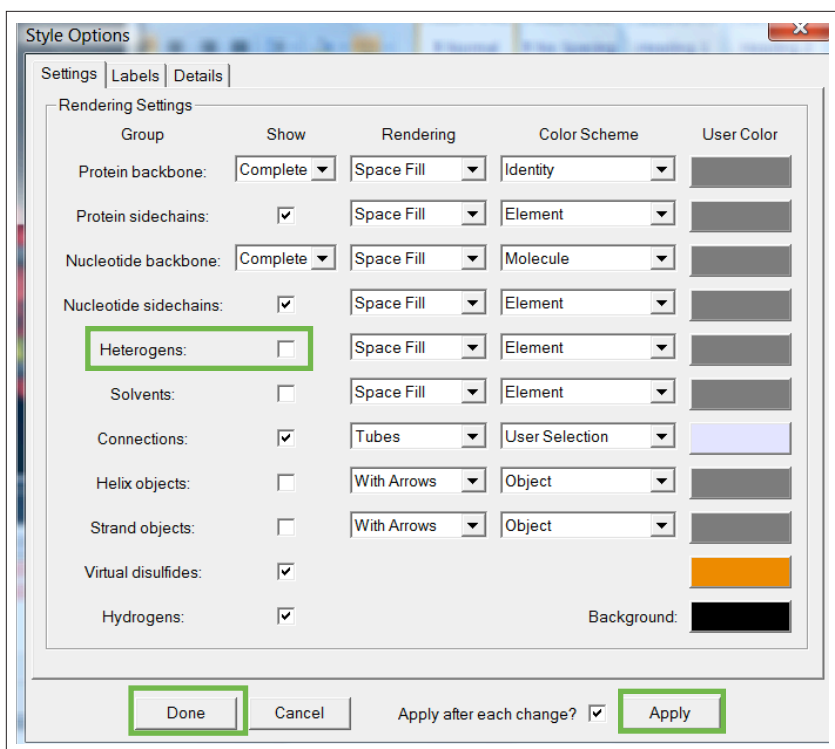
16. Look down to the Sequence/Alignment viewer. Move the cursor along the amino acid chain until you see a place where the amino acids differ.

When the cursor is directly over an amino acid, the bottom left corner of the screen shows the location of the cursor. Find “**PDB 1775**” to locate the mutation.

# LESSON 5

## RESOURCE

**Figure 6:** Hide the heterogens, which have no effect on the function of the proteins. These molecules are often artifacts of the protein crystallization process. Credit: Cn3D, Wang et al. 2000.



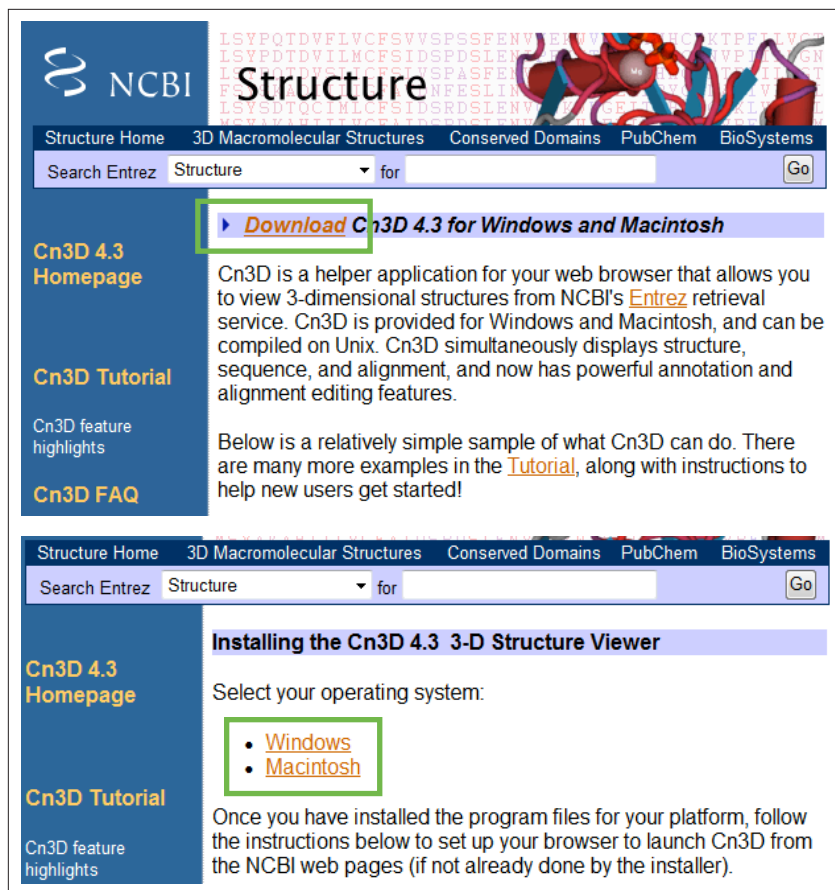
17. **Highlight the M** at 1775 in the 1JNX sequence by clicking on it.  
**Highlight the R** in the 1N5O sequence by clicking on it and pressing “Ctrl” at the same time. This also highlights the amino acids in the structure viewer so they are more visible.
18. Experiment with using the different coloring styles to observe the impact of the mutation. Every time styles are changed, however, the heterogens associated with the 1N5O sequence will need to be hidden (repeat Step #15 to do this).
19. While toggling back and forth using the arrow keys, point out to students how the change in amino acids (from M to R) changes the protein structure.  
 Methionine (M), found in the non-mutated version of the protein, is a non-polar amino acid. Non-polar (hydrophobic) amino acids fold to the inside of the protein structure, away from the aqueous environment. Arginine (R), found in the mutated version of the protein, is a polar amino acid. Polar (hydrophilic) amino acids tend to rotate to the outside of the protein structure. Students can see arginine “pop” to the outside of the mutated structure, which affects the shape of the entire structure in a subtle—yet important—way.

# 5

## Installing Cn3D

The Cn3D software allows you to view the three-dimensional structures of macromolecules like proteins and DNA. If Cn3D has not already been downloaded to your computer, follow the instructions below.

1. To download the Cn3D program onto your computer, visit the Cn3D homepage at: <http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>.
2. Click **“Download.”**
3. Select your Operating System (Windows, Macintosh).
4. To install Cn3D on a Mac, click the link to download the file. When the file has been downloaded, double-click the icon to install Cn3D. To install Cn3D on a computer with Windows, click the link to **“Download the Cn3D 4.3 installer here.”**
5. Close all internet browsers and then double-click on the .msi file or select **“Run.”** This will launch the Windows Installer application if you have it already. If so, enter the information at the prompts, and you are done. If not, or if Windows tells you that you need a newer version of the installer software, you can download the latest Windows Installer from Microsoft® by following the appropriate links listed on the download page for your operating system.
6. Once Cn3D is installed, you can click the **“View Structure in Cn3D”** link from the Structure summary pages to download and view structures from the NCBI.



Credit: NCBI Structure, Molecular Modeling Database (MMDB).



# 5 From Sequence to Structure

## Introduction

This lesson extension uses music as an analogy for protein structure and function. *Part I* compares primary, secondary, tertiary, and quaternary protein structures to a musical composition. *Part II* relates the idea of conserved domains in proteins to the song “Happy Birthday.” All mp3 music files for this optional lesson extension can be found under the Resources tab on the Introductory curriculum webpage at: <http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing>.

## PART I: Primary, Secondary, Tertiary, and Quaternary Protein Structure

### A. Primary Structure

Music is made up of a sequence of notes, and proteins are made up of a sequence of amino acids. Knowing only the **order** of the amino acids in a protein tells us the **primary structure** for that protein. This does not, however, tell the whole story of the protein. In a musical composition, knowing the **order** of the notes is of course important but also does not tell the whole story. For example, the musical notes for the beginning of a piece of music are:

DABADABAAABAAABAABAGF

If we could search a large database containing music for individual notes (iTunes® is a large database containing music, but it cannot search for sequences of notes), we might be able to find out more about our piece of music. Are there many other songs that contain the same sequence? A musician might be able to make an educated guess that this piece is in the key of D. Does that musical key link it to other musical pieces? Of course, since this is a piece of music, we can actually listen to it.



And it would sound like this:  
Play the “Primary Structure Song.”

It doesn’t sound like much, does it?

Now, let’s take a look at the beginning sequence of amino acids from the *BRCA1* reference sequence that we used for the BLAST alignment in *Lesson Four*. The beginning sequence of amino acids for the *BRCA1* gene is:

MDLSALRVEEVQNVINAMQKILECPICLE

Again, knowing the **order** of the amino acids is important but does not tell the whole story. Using the NCBI, we can search a database that might compare this sequence with other sequences, which could tell us how common the sequence is, or perhaps even whether it is linked to other proteins with similar sequences. But, again, this does not tell us very much about what the protein **does**.

### B. Secondary Structure

The basic shape of a protein begins to emerge when we look at the **secondary structure** that is formed as the **alpha helices** and **beta sheets** take shape.

Let's return to the music analogy. The order of the notes gives us an indication of what the music should sound like, but we need more information to make the music take shape. Music is given shape by rhythm, dynamics (how loudly or softly sections are played), accents, the key in which it is written, and many other factors. If we add rhythm, some tone, and accents to the string of notes we heard earlier, it begins to take shape and becomes much more recognizable:



Play the "Secondary Structure Song."

How does the BRCA1 protein take shape? It has to do with the interactions between the different amino acids and their functional groups that constitute the protein—interactions such as hydrogen bonding, **non-polar** interactions, ionic bonding, and even **covalent bonding**. If you take a closer look at the string of amino acids from our primary structure—look at the chemical structure and composition of each amino acid—you will see that some of the amino acids have **polar (hydrophilic, or water loving)** side groups and some have non-polar (**hydrophobic, or water fearing**) side groups. The non-polar side groups shy away from the watery solution the molecule is in, and bend the protein so that they can be gathered together on the inside of the structure. Hydrogen bonds also form between the amino acids themselves, which twists the protein into an **alpha helix**. Sometimes the interactions between the amino acids cause the string to fold back and forth in a flattened zigzag, which is called a **beta sheet**.

Just as rhythm, tone, and accent give shape to music, these chemical interactions twist and fold our amino acid chain into a basic three-dimensional structure.

### C. Tertiary Structure

**Tertiary structure** is completed when the attractions between alpha helices and beta sheets cause the protein to fold back on itself and make "bridges" between sections of the protein. This gives the protein its final, folded, and compacted shape.

Using the music analogy, we have seen the music take shape from a string of single notes to, perhaps, a recognizable tune. At the level of tertiary structure, additional connections and bridges are made within the tune itself. The original sequence showed (and you have only heard) the first 21 notes of a much longer composition. If the sequence were written out in its entirety, there may be places that bridge to and interact with the beginning sequence. It might sound like this:



Play the "Tertiary Structure Song."

**Non-polar:** A substance that repels water. A synonym for hydrophobic.

**Covalent bonds:** A type of chemical bond that is characterized by the sharing of a pair of electrons between atoms.

**Polar:** A substance that is attracted to water. A synonym for hydrophilic. The opposite of hydrophobic or non-polar.

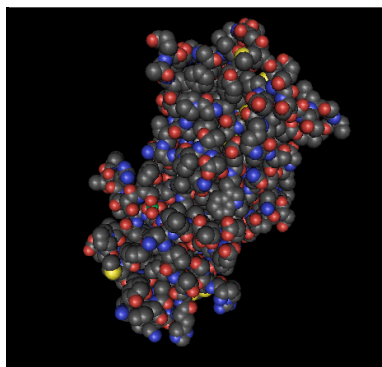
**Hydrophilic:** A substance that is attracted to water. From the Greek "hydro" which means water, and "philos" meaning love. A synonym for polar. The opposite of hydrophobic or non-polar.

**Hydrophobic:** A substance which repels water. From the Greek "hydro" which means water, and "phobos" which means fear. A synonym for non-polar. The opposite of hydrophilic or polar.

**Alpha helix** (Plural: "alpha helices"): A common structure of proteins, characterized by a single, spiral chain of amino acids stabilized by hydrogen bonds.

**Beta sheet:** A structure that occurs in many proteins and consists of two or more parallel adjacent polypeptide chains arranged so that hydrogen bonds can form between the chains.

The BRCA1 protein, likewise, has its basic shape but requires additional infrastructure to be fully functional. For many proteins, their structures are dictated by additional folding that occurs when important covalent bonds form between the sulfur-containing amino acid cysteine. The final structure looks something like this:



Credit: Varma et al. 2005.

This final shape, or **structure**, allows the BRCA1 protein to perform its function. It is bent, folded, and twisted to make it just the right size and shape so that it fits along the DNA double helix and does its job of repairing broken strands of DNA.

### D. Quaternary Structure

Some proteins have an additional level of structure, the **quaternary structure**. If the fully functional end-protein is composed of more than one protein chain, the assembly of this larger unit is the fourth level of structure.

In the music analogy, we need to look at the end product. The composition was not written for a single violin, but for orchestra and choir. The composer built additional musical “bridges” between the different instruments and choir to create one cohesive piece of music. When it is all put together, it sounds like this:



Play the “Quaternary Structure Song.”

**Something to think about:** There are many levels of organization in making a “functional” musical composition and in folding functional proteins. This humble protein may not look like the equivalent of the “Hallelujah Chorus,” but BRCA1 is found in many species and performs a crucial function in DNA repair. In fact, George Frideric Handel would not have existed to write the “Hallelujah Chorus” without functional BRCA1 proteins.

### Sources:

#### 3-D Molecular Designs

<http://www.3dmoleculardesigns.com/news2.php>

#### Survey and Summary: Structural Classification of Zinc Fingers

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC140525/?tool=pubmed>

#### Wikipedia: Zinc Finger

[http://en.wikipedia.org/wiki/Zinc\\_finger](http://en.wikipedia.org/wiki/Zinc_finger)

### PART II: Conserved Domains

**Domain:** Specific area of a protein that performs a particular function.

Nature has versions of its own “greatest hits” collection. These are the sequences of amino acids that code for proteins or regions of proteins that have crucial functions for many organisms. If many organisms require the same function (i.e., DNA replication and repair, energy metabolism, cell division) they will need similar proteins, made from similar sequences of amino acids. These related sequences are said to be **conserved** sequences, or conserved **domains**.

Note that these conserved sequences are **similar**, not **identical**. They need to be similar enough that the resulting protein can do the required job for that organism, yet organisms have a certain amount of genetic diversity. How can living things make proteins to do a specific job, yet allow for genetic diversity?

**Back to the music analogy:** You need a song with a specific function. Let’s say your friend is having a birthday and you would like to celebrate it with a song. We have, of course, a song just for this function.



Play the “Happy Birthday” song.

This is perhaps the most well-known song in the world—if it were a protein, it would be **highly conserved**. Would your friend, however, recognize this version?



Play the “Happy Birthday Improvisation” song.

Perhaps we do not need the entire original song to perform our function of wishing your friend a happy birthday—maybe the first six notes are crucial to the song, and then some variation can occur. Perhaps, throughout our improvisations, we need to have a few notes along the way that anchor us back to the original melody. One can also imagine that too much improvisation could make the melody totally unrecognizable and your friend would not know that you are trying to wish him or her a happy birthday. A song like that would not fulfill its function, and would probably not become known by too many people.

In a similar way, amino acid sequences that code for a functional protein can vary from one organism to another, as long as the conserved portion of the gene continues to code for the crucial elements that make up that protein.