Visualizing Microbial Proteins

Ethel D. Stanley and Keith D. Stanley

Video VII: Microbial Diversity

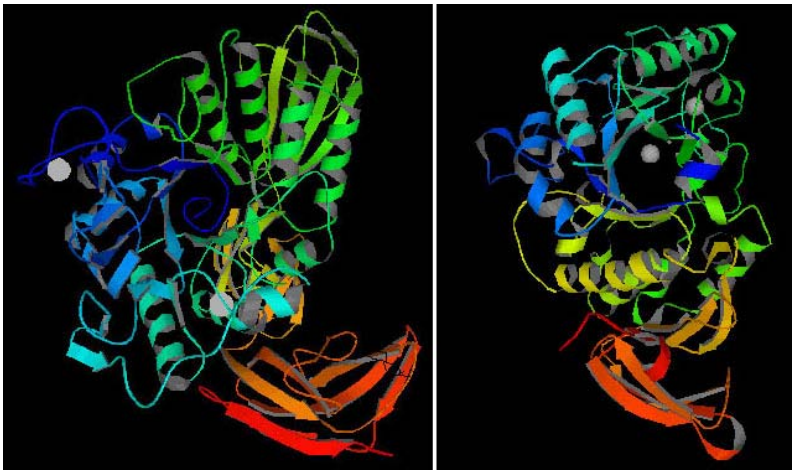

Figure 1. The CGTases share similarities with the alpha-amylases. The ribbon images of the proteins 1CDG and 6TAA were captured during an interactive 3D VRML session at the Protein Data Bank web site: http://www.rcsb.org/pdb/.

Molecular data for microbial proteins are readily accessible to anyone with a computer, but how do we use these resources? Visualization and bioinformatics tools prove useful as we take a closer look at microbial enzymes.

Part 1: Viewing Structural Models and Sequence Data

Although the two proteins above are clearly different molecules, we can see some similarity in the structure of the cyclomaltodextrin glucanotransferase (CGTase) and the alpha-amylase in Figure 1. Each of these enzymes contains a cylindrical cavity supported by alpha-helices (ribbon spirals) connected to beta-sheets (ribbon arrows) lining the interior of the cavity. Functionally similar, both proteins are active in the breakdown of starch molecules although the CGTases produce cyclomaltodextrins instead of the maltodextrins produced by alpha-amylases.

To examine the structure of the proteins above more closely, access the Protein Data Bank site at http://www.rcsb.org/pdb/.

Enter 1CDG into the box under Search the Archive and hit Find a structure. The Query Result Browser will appear. Click on 1CDG from the list of proteins to see the Summary Information page for 1CDG. (See Figure 2.)

Choose View Structure on top of the left column.

The View Structure page provides a list of options. Explore 1CDG and then 6TAA using at least the two options below:

1. To interactively explore the ribbon diagram above, click on VRML (default options) on top of the list. A 3D structural model will appear. To change the model's size, click on the Walk and move the cursor up or down. To rotate the model freely, click on Study.

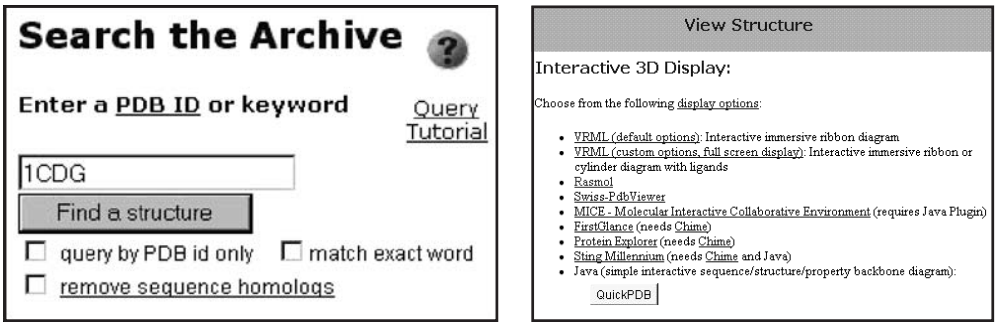

Figure 2. Screen shots from the Protein Data Bank featuring the archive search and interactive 3D display options.

2. To explore the relationship between a wireframe structure of 1CDG and its protein sequence data, click on the QuickPDB at the bottom. You can rotate the model to view interesting structural components. Note the small Cursor window on the left. Individual amino acids and their position show up as you move over the model. Click on an amino acid in the long Sequence window across the top to highlight it in the model below.

Many proteins have been sequenced, but most lack the x-ray crystallography data necessary to generate structural models. Researchers interested in these proteins use the sequence data itself. Just as it takes time to learn to see information in a structural model, viewing sequence data meaningfully requires new analytical approaches and strategies.

Consider the sequence data for these two proteins in Tables 1 and 2. Asterisks mark the amino acid residues corresponding to an active site found in 1CDG at positions 225 - 233 and in 6TAA at positions 202 – 210. These conserved sequences confer specific properties to an active site and were retained during the evolution of these two proteins.

- Compare the conserved regions marked by asterisks in these two sequences. Which amino acids are the same?

Use the conserved sequence information in Table 3 to enter your own asterisks under the corresponding region of the alpha-amylase sequence in Table 2. (Note: The numbers beginning each row in Table 2. indicate the sequence position for the first amino acid residue in the row.)

Table 3: Well-conserved regions of alpha-amylase from *Aspergillus oryzae* with locations of specific amino acids and known structure-function relationships. (adapted from MacGregor 1996)

Starting and ending position of residues	Conserved sequence (Highly conserved in **bold**)	Some known structure - function relationships	
116 - 122	V**D**VVA**NH**	Subsite 1	H = His 122
202 - 210	GL**RID**TVKH	Catalytic site Subsite 1 Subsite 2	D = Asp 206 H = His 210 K = Lys 209
230 - 233	**E**VLD	Catalytic site	E = Glu 230
293 - 298	VEN**HD**N	Catalytic site Subsite 1	D = Asp 297 H = His 296

1	APDTSVSNKQ	NFSTDVIYQI	FTDRFSDGNP	ANNPTGAAFD	GTCTNLRLYC
51	GGDWQGIINK	INDGYLTGMG	VTAIWISQPV	ENIYSIINYS	GVNNTAYHGY
101	WARDFKKTNP	AYGTIADFQN	LIAAAHAKNI	KVIIDFAPNH	TSPASSDQPS
151	FAENGRLYDN	GTLLGGYTND	TQNLFHHNGG	TDFSTTENGI	YKNLYDLADL
201	NHNNSTVDVY	LKDAIKMWLD	LGIDGIRMDA	VKHMPFGWQK	SFMAAVNNYK
			******	***	
251	PVFTFGEWFL	GVNEVSPENH	KFANESGMSL	LDFRFAQKVR	QVFRDNTDNM
301	YGLKAMLEGS	AADYAQVDDQ	VTFIDNHDME	RFHASNANRR	KLEQALAFTL
351	TSRGVPAIYY	GTEQYMSGGT	DPDNRARIPS	FSTSTTAYQV	IQKLAPLRKC
401	NPAIAYGSTQ	ERWINNDVLI	YERKFGSNVA	VVAVNRNLNA	PASISGLVTS
451	LPQGSYNDVL	GGLLNGNTLS	VGSGGAASNF	TLAAGGTAVW	QYTAATATPT
501	IGHVGPMMAK	PGVTITIDGR	GFGSSKGTVY	FGTTAVSGAD	ITSWEDTQIK
551	VKIPAVAGGN	YNIKVANAAG	TASNVYDNFE	VLSGDQVSVR	FVVNNATTAL
601	GQNVYLTGSV	SELGNWDPAK	AIGPMYNQVV	YQYPNWYYDV	SVPAGKTIEF
651	KFLKKQGSTV	TWEGGSNHTF	TAPSSGTATI	NVNWQP	

Table 1. Sequence Information for 1CDG, a CGTase produced by a bacterium. From the Protein Data Bank: www.rcsb.org/pdb/.

1	ATPADWRSQS	IYFLLTDRFA	RTDGSTTATC	NTADQKYCGG	TWQGIIDKLD
51	YIQGMGFTAI	WITPVTAQLP	QTTAYGDAYH	GYWQQDIYSL	NENYGTADDL
101	KALSSALHER	GMYLMVDVVA	NHMGYDGAGS	SVDYSVFKPF	SSQDYFHPFC
151	FIQNYEDQTQ	VEDCWLGDNT	VSLPDLDTTK	DVVKNEWYDW	VGSLVSNYSI
201	DGLRIDTVKH	VQKDFWPGYN	KAAGVYCIGE	VLDGDPAYTC	PYQNVMDGVL
	*********				
251	NYPIYYPLLN	AFKSTSGSMD	DLYNMINTVK	SDCPDSTLLG	TFVENHDNPR
301	FASYTNDIAL	AKNVAAFIIL	NDGIPIIYAG	QEQHYAGGND	PANREATWLS
351	GYPTDSELYK	LIASANAIRN	YAISKDTGFV	TYKNWPIYKD	DTTIAMRKGT
401	DGSQIVTILS	NKGASGDSYT	LSLSGAGYTA	GQQLTEVIGC	TTVTVGSDGN
451	VPVPMAGGLP	RVLYPTEKLA	GSKICSSS		

Table 2. Sequence Information for 6TAA, an alpha-amylase (also known as TAKA amylase) found in *Aspergillus oryzae*. From the Protein Data Bank: www.rcsb.org/pdb/.

Now that the conserved sequence VDVVANH is marked by asterisks in Table 2, try to locate the corresponding sequence for CGTase in Table 1. Enter asterisks appropriately in Table 1. (Note: This is not easy to do. Imagine trying to manually compare hundreds of proteins for multiple sites! Part 2. introduces bioinformatics tools that will align and identify conserved sequences in two or more proteins.)

- List the starting and ending positions and the amino acid sequence you've found in 1CDG that correspond to this active site in 6TAA.
- Which amino acids are conserved?

Variations in the amino acids can change the functionality of the active site. For example, CGTase sequences have a phenylalanine residue located near the active site that corresponds to the glutamic acid residue (E230) in the alpha-amylase sequence EVLD shown in Table 3. The phenylalanine is required for cyclization of the cleaved maltodextrin fragments. (MacGregor, 1993)

Part 2: Using Sequence Data with Online Bioinformatics Tools

The sequence information for one CGTase can be used to probe for similar proteins in an online molecular database. (See "Orientation to the *Biology Workbench*" on the *Microbes Count!* CD and the "Searching for Amylase" activity in Chapter 2 for an explanation of how to get started.) In *Biology Workbench*, use Ndjinn to search for the 1CDG protein in the PDBSEQRES database.

- View the record. What organism produces this protein?

Import the sequence, and use BLASTP to find similar proteins in the SWISS-PROT database.

- Choose three of the CGTases that are displayed, but make sure each is produced by a different organism. List the three CGTases and the organism that produces them.

Use CLUSTALW to align the sequences for these three proteins with the 1CDG protein. Consider the output of aligned sequences. Are you surprised by the number of amino acid residues that are conserved?

- CLUSTALW also produces an unrooted tree from the sequence data. Usually the tree is near the bottom of the output file. Use the tree to decide which of your CGTases are the most similar to the 1CDG protein. Defend your choice.

Not all CGTases are used in industrial starch processing. A CGTase used to make alpha-cyclodextrins commercially is produced by *Paenibacillus macerans,* formerly known as *Bacillus macerans* (Guzman-Maldonado and Paredes-Lopez, 1995). The SWISS-PROT label for this CGTase is CDG1_PAEMA.

- Devise a strategy using sequence information to determine which of the CGTases you are working with (including 1CDG) would be most likely to function similarily in the breakdown of starch. Explain your results.

Part 3. An Extended Research Problem: Making the Case for a New Microbial Enzyme

- The protein AMYR_BACS8 is produced by a microbe that can digest raw starch. Is this advantageous for the *Bacillus*? Why?
- Would this be of value to humans? Explain.
- You have been asked to present a report at the planning meeting for a starch processing plant on possible applications of this microbial enzyme. Your role is to provide information on the protein itself. Identify three questions you anticipate will be asked during the meeting.
- Prepare a short presentation that both the research scientists and the marketing specialists will understand. In addition to looking for information from publications on the protein, present the available molecular data drawing from your experience with the visualization and bioinformatics techniques in this activity.

Web Resource Used in this Activity

Biology Workbench (http://workbench.sdsc.edu)
Originally developed by the Computational Biology Group at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. Ongoing development of version 3.2 is occuring at the San Diego Supercomputer Center, at the University of California, San Diego. The development was and is directed by Professor Shankar Subramaniam.

Protein Data Bank
http://www.rcsb.org/pdb/

VRML Plug-in: Cosmo Player
You will need the Cosmo Player plug-in to use VRML. See download information on the PDB site or http://www.karmanaut.com/cosmo/player/

Additional Resources

Available on the *Microbes Count!* CD

Text

A copy of this activity, formatted for printing

"Orientation to the *Biology Workbench*"

Related *Microbes Count!* Activities

Chapter 2: Searching for Amylase

Chapter 4: Molecular Forensics

Chapter 4: Exploring HIV Evolution: An Opportunity for Research

Chapter 6: Proteins: Historians of Life on Earth

Chapter 6: Tree of Life: Introduction to Microbial Phylogeny

Chapter 6: Tracking the West Nile Virus

Chapter 6: One Cell, Three Genomes

Unseen Life on Earth Telecourse

Coordinates with Video VII: Microbial Diversity

Relevant Textbook Keywords

Active site, Amino Acid, Bioinformatics, Conserved sequence, Molecular model, Visualization, X-ray crystallography,

Related Web Sites (accessed on 4/23/03)

Examining protein structures
http://www.ornl.gov/TechResources/Human_Genome/posters/chromosome/pdb.html

Microbes Count! Website
http://bioquest.org/microbescount

Molecular Graphics Manifesto
http://www.usm.maine.edu/~rhodes/Manifesto/text/01Intro.html

Tutorials on How to Use RasMol and Chime
http://www.umass.edu/microbio/rasmol/rastut.htm

Unseen Life on Earth: A Telecourse
http://www.microbeworld.org/htm/mam/is_telecourse.htm

Bibliography

Guzman-Maldonado, H. and O. Paredes-Lopez (1995). Amylolytic enzymes and products derived from starch: A review. *Critical Reviews in Food Science and Nutrition.* 35 (5):373-403.

Janacek, S., E. Leveque, A. Belarbi, and B. Haye (1999). Close evolutionary relatedness of alpha-amylases from archaea and plants. *J. Molecular Evolution* 48:421-426.

MacGregor, E. A. (1988). a-Amylase structure and activity. *J. Protein. Chem.* 7:399-415.

MacGregor, E. A. (1993). Relationships between structure and activity in the alpha-amylase family of starch-metabolising enzymes. *Starke* 45:232-237.

MacGregor, E. A. (1996). Structure and activity of some starch metabolizing enzymes in *Enzymes for Carbohydrate Engineering*. Park, K-H., J. F. Robyt, and Y. D. Choi, Editors. Elsevier: Amsterdam.

Figure and Table References

Figure 1. Courtesy Sam Donovan (Beloit College)

Figure 2. Modified from *Biology WorkBench* (http://workbench.sdsc.edu)

Table 1. From the Protein Data Bank: www.rcsb.org/pdb/

Table 2. From the Protein Data Bank: www.rcsb.org/pdb/

Table 3. Adapted from MacGregor 1996