

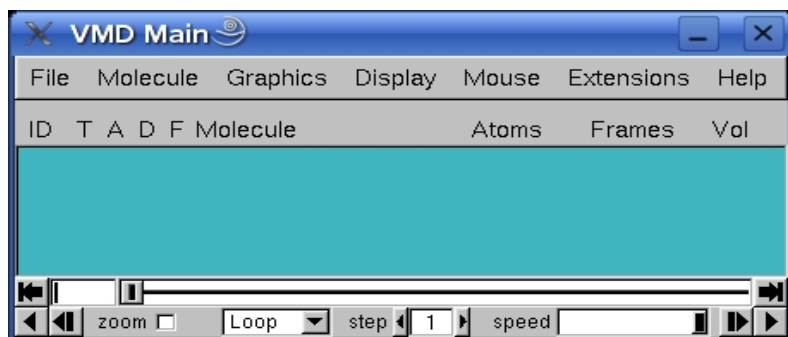
Protein Physics by Computer.

Peter Schellenberg, Institute of Physics, University of Minho,
Campus de Gualtar, 4710-057 Braga - Portugal
<http://webpe.sapia.uminho.pt/PSchellenberg/biovis.html>

Step by Step: Protein Visualization with VMD.

A step by step introduction into the visualization of biomolecules using VMD is given. In this manual, hemoglobin serves as a first example. The procedures given here correspond exactly to the course of action in the hands-on lecture. Therefore, if you get lost during the advance of the lecture you can easily regain the connection by looking at these pages. Therefore **it is highly recommended to have a hardcopy at hand !** In the course of the lecture, different structural features of hemoglobin are displayed, such as secondary, tertiary and quaternary structure. Again other presentations focus on the cofactors or the ligation of the irons by histidin.

1. Start VMD by pressing the corresponding button on the windows GUI.
2. Go to the window box VMD Main.



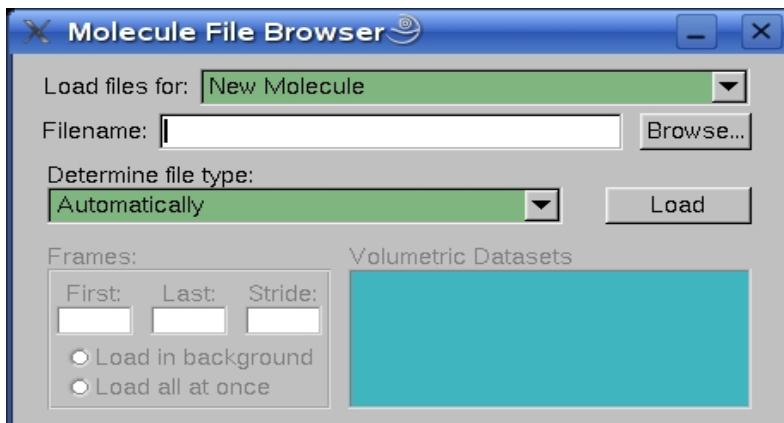
First, change the mode that defines the viewing geometry of the presentations from Perspective to Orthographic:

Display -> Orthographic

After that load the file with the molecular coordinates:

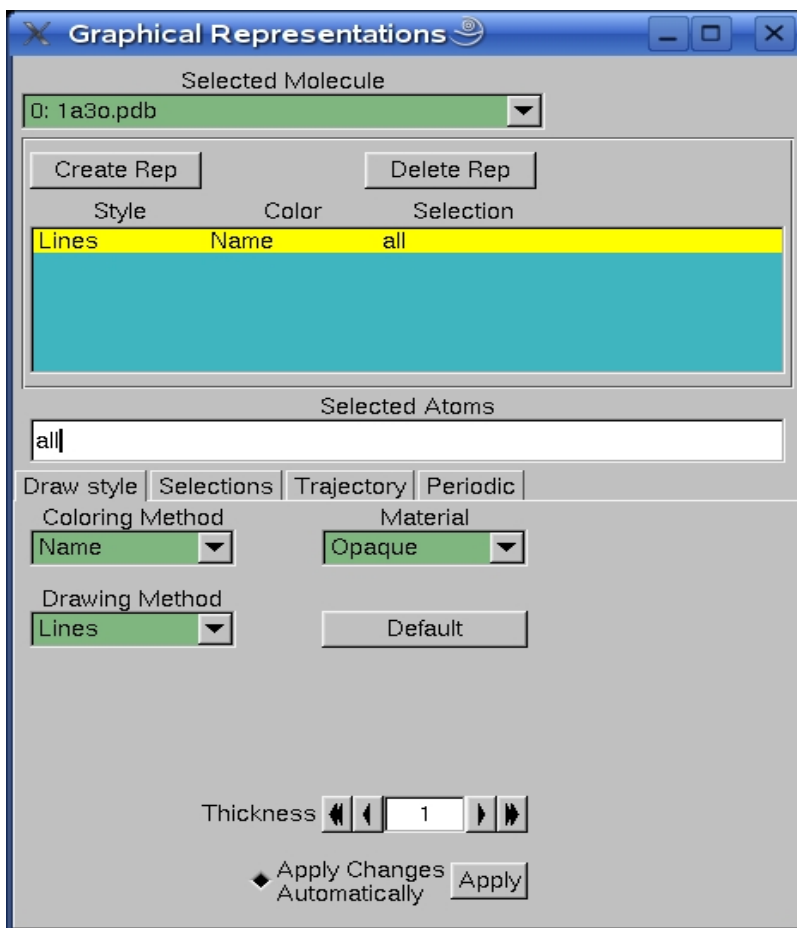
File -> New Molecule

Go to the Molecule File Browser – window box **Browse**, define the pdb-file and load it with **Load**. Then, you can close this window box again.



3. In the graphics display you see the hemoglobin in a plain representation that is not very suitable for the understanding of biomolecules. Go to the window box VMD_Main and choose **Graphics** -> **Representations**.

The Graphical Representation window opens:



In the bluish green field, you see the presentation, which has already been automatically

created with the following parameters:

Style: **lines**, Color: **Name**, Selection: **all**.

These key words describe the present presentation in the graphical window box. You can now vary this presentation by changing the keywords. click in the list box under Coloring Method and change the keyword '**Name**' to '**Chain**'. In the graphical window box you can now easily distinguish the quaternary structure subunits of hemoglobin.

Now, go to the list box '**Drawing Method**'. Change '**Lines**' for example to '**Tube**', '**Ribbons**' or '**New Cartoon**'.

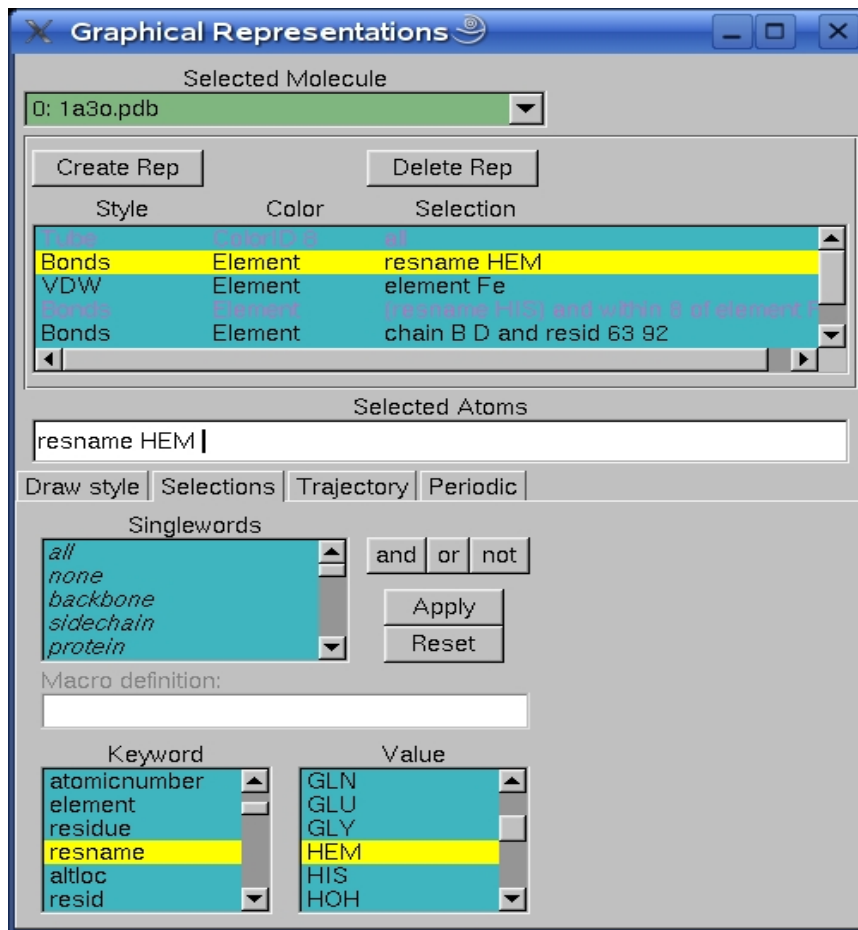
As all these presentations look nice, you may wish too somehow keep all of them. However, by changing the choice of presentation, the old selection is gone. However, VMD has a very nice solution, how to retain all those selections you wish too: click on the button '**Create Rep**'. Now you see two still identical lines, and you can alter one of these according to your needs. You can switch-off one of these presentations by double-clicks on the line: The text color changes from black to pink, and particular presentation vanishes in the graphics window box. You can reverse the event by repeating the procedure.

Make one the presentation invisible. You can now play with the others. You can certainly create more than two lines (presentations), and keep them visible or invisible as you like.

Hemoglobin contains cofactors that are mandatory for the storage of oxygen, namely HEM, and it is important to create a display emphasizing those: First, create a new presentation line by pressing **Create Rep**. Up to now, you always selected the whole protein for display. This is indicated by the keyword 'all'. Now you want to do display alterations just affecting the HEM -group of your protein. To achieve that delete the word '**all**'. Then go to the tab **Selections**.

Choose **resname** under keyword and confirm by double-clicking. The keyword '**resname**' is now written in the line 'Selected Atoms' instead of '**all**'. In the raw value you see the name of amino acids and cofactors. Scroll down. Select '**HEM**' and press return. The Selection '**resname HEM**' is taken over. Go back to the tab '**Draw style**' and change the display to make the cofactor visible, for example by choosing the Drawing Method: '**Bonds**' and Coloring Method: **Elements**. Instead of the selection '**Bonds**' you can choose one, that emphasizes the cofactors even more, for example '**Licorice**' or '**VDW**'. Play around!!!

The protein backbone can be made transparent, just change the selection in Material from '**Solid**' to '**Transparent**'. You can also change the color of the backbone to white: Go to ColorID and choose 8: white.



Next you can enhance the iron in the HEM group even further:

Button '**Create Rep**'

Delete Selected Atoms '**resname HEM**'

Tab '**Selections**'

Keyword: **element**, Value: **Fe**

press return or '**Apply**'

At this point you already did extensive work and it would be a pity if you lost it. However, there is an option to save your work. This can be done by going to the window box VMD_Main and choosing **File** → **Save State**. Select a descriptive name for your presentation.

Back to hemoglobin: The iron is ligated by Histidins. How can we display these Histidins?

Recommendation: Display the protein backbone as **Trace** or **Tube**.

make a copy of the line representing the HEM group

Button '**Create Rep**'

Delete Selected Atoms '**resname HEM**'

go to Tab '**Selections**'

Keyword: **resname**, Value: **HIS**

return or '**Apply**'

go to Tab '**Draw style**'

Coloring method: **ColorID: green**, Drawing method: **Bonds**

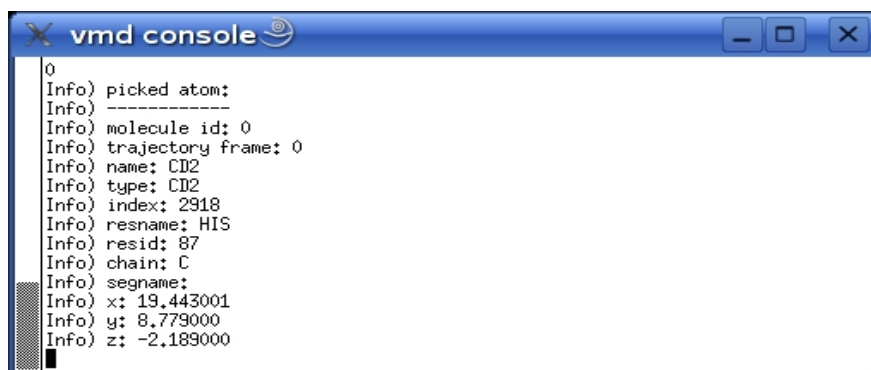
We encounter a problem: **All HIS** are displayed, not only those that complex the iron.
Solution: selection of the specific HIS. Help: Any amino acid has a unique residue ID.

How do I find out the residue ID number?

In the window box VMD_Main choose

Mouse -> Query

Go to the graphical window and click on the Histidin, that ligate the iron. You will find the required information in the console window box vmd_console:



To be exact, you get the information, which belongs to that atom on which you clicked.
This information includes the Residue ID.

As an overall result we get the following information:

chain B D and resid 63 92

chain A C and resid 58 87

Now we have a very informative representation of hemoglobin, that includes many important aspects of this protein.

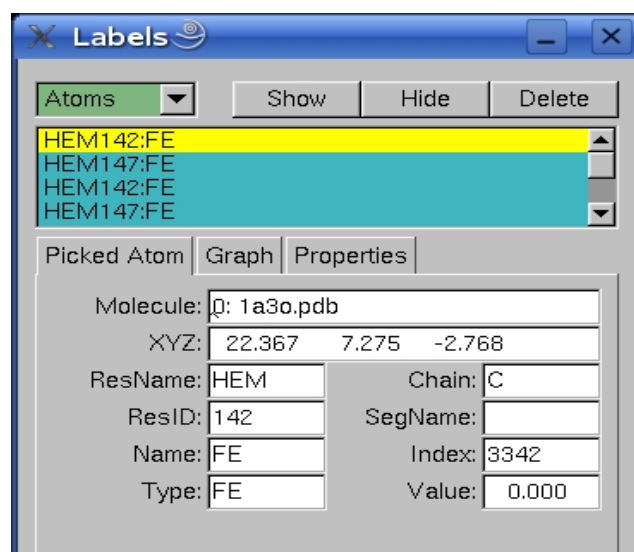
It is often meaningful to label a specific atom. This can be done by selecting

Mouse -> Label -> Atoms

and use the mouse to mark the atom. A label appears at the requested location. You can alter the label by going to the label control box:

Graphics -> Labels

The Labels window box appears:



All the previous selections can be activated, deactivated or deleted with Show, Hide or Dele^te. Furthermore, the letter size, the format and even the offset of the labels can be changed by going to the Tab Properties. For example it is a good idea just to write HEM as the label, which is easily done by modifying the format.

Correspondingly, one can also display a distance. Go to

Mouse -> Label -> Bonds

and, by using the mouse, consecutively choose the coordinates, for which you want to know the distance. Analog to the molecule labels you can modify the selection in the labels window:

Graphics -> Labels -> Bonds

Now you can alter the different presentations of hemoglobin. For example, you can alter the transparency of the protein backbone by choosing another Material. You have a wide choice of predefined materials, however, you can still define your own materials: Go to

Graphics -> Materials

In some cases it may make sense to alter the illumination of the protein. There are two light sources predefined to be switched on, whose position you can alter by choosing '**Mouse** -> **MoveLight**' and checking one of the lights.

Additional light sources can be added by activating the check boxes **Light 2** or **3** in '**Display**'.

Under '**Mouse**' you also find how to switch modes of motion for the display.

Eventually you can familiarize yourself with additional features of VMD, that can be found in the Extensions -Menu:

1. Sequence viewer:

Extensions -> Analysis -> Sequence Viewer

Upon selection of an amino acid in the **Sequence Viewer** it is depicted yellow in the graphic display and is also shown in the selection line of the **Graphical Representation Window**.

2. Ramachandran plot:

Extensions -> Analysis -> Ramachandran Plot

In the Plot window, choose Molecule 0 (1a3o). You will see the angle distribution for all amino acids. You can narrow the selection, for example by choosing chain-A. Each amino acid can be identified by clicking on the yellow mark.

3. Clipping plane tool:

Extensions -> Visualization -> Clipping Plane Tool

This is useful in volumetric presentations and in large structures, for example viruses.

3. Contact map:

Extensions -> Analysis -> Contact Map

This plot is a map with the contacts between amino acids. The diagonal line present the protein backbone, of course each amino acid is next to the preceeding and following one. But you also see contacts over space due to the 3-dimensional structure of the protein.

4. Viewmaster:

Extensions -> Visualization -> Viewmaster

This is an extremely useful tool, when presenting a previously created presentation during a seminar, a talk or alike. One stores each meaningful presentation in one frame, and by clicking on this frame the presentation is shown. For example, you can store a presentation featuring the three dimensional structure in one frame and a presentation featuring the ligation of the HEM in another. During a seminar talk, you can easily switch between these without bothering about any settings. Please note that there is an alternative to that system, namely the creation of a button controlled menu within a script. This possibility is described in the article: **Improving presentations with the program Visual Molecular Dynamics by controlling the display with button navigation windows** on the lecture page, Examples exploiting that method are also given there.

An Excursion through the world of proteins.

In this second part of the workshop, we will undertake an excursion through the world of proteins and thereby study selected proteins, that are important and specific in nature and we will learn about biophysical processes taking part in these proteins.. The structures are provided as pre-created scripts in VMD containing button controls to switch between different presentations easily. For questions concerning VMD, you may consult the operating manuals [Caddigan et al.(2005), Humphrey et al.(1996), Stone et al.(2001)], which can be downloaded from the VMD -homepage. As far as the programming of the scripts and their installation is concerned, there are brief manuals available from my homepage [Schellenberg(2010)].

The particular protein presentations described here are part of the whole package `vmdscriptger.tar.gz` resp. `vmdscriptger.zip`, which contain many more scripts apart from the ones used here. For installation of the scripts, please read the installation text.

To load a particular script, to the window box VMD_Main and choosing **File** -> **Load State**. Go to the directory

```
C:/own_files/vmdscripteng
```

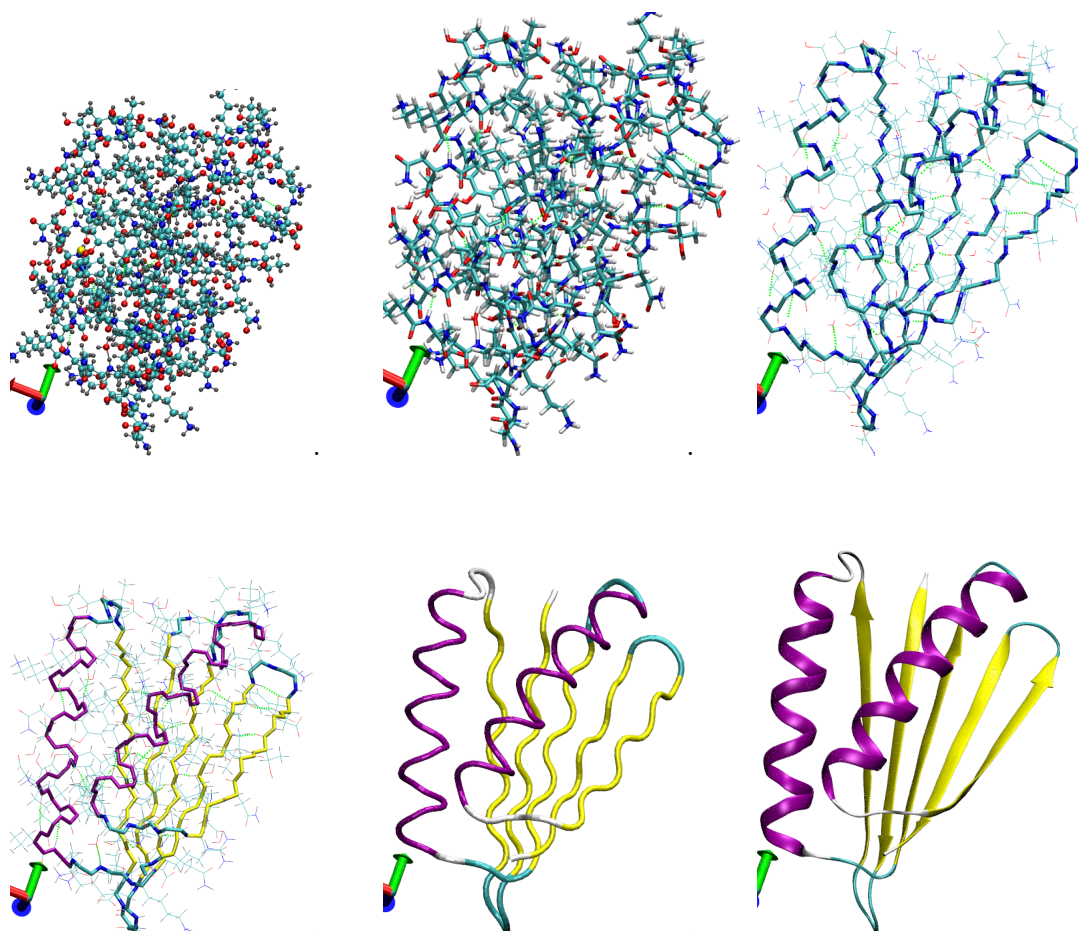
and load the vmd file

Protein structure

Scriptfile: `designerproteinhypertextengl.vmd`

The structure of this protein had been calculated in advance, prior to producing the primary structure genetically. So far for the name designer protein. The theoretical predictions turned out to be surprisingly accurate. Taking this simple and straightforward structured protein, it is demonstrated, how to get from the atomic structure model via the bonding model to the abstract protein presentations regularly used in biochemistry, such as tubes, ribbons and cartoons.

First, the atomic structure model and developing into the bonding lines presentations, on the right with the protein backbone bondings drawn bold The increasingly abstract models are based on the backbone structure, with alpha -helix (violet), beta -sheet (yellow) and turns (cyan). The non-structured parts are shown in white.



Photosynthesis

Scriptfile: rclh1button.vmd

The presentation shows the bacterial reaction center from *Rhodopseudomonas Palustris* and the surrounding light harvesting complex LC I. The second protein is the LH-II complex from (*Rhodopseudomonas Acidophila*), which surrounds the RC -LH I complex. The chromophores are usually named according to their absorption wavelength as Bacteriochlorophyll B 800, B850 and B875. The first two are present in LHII and arranged in two rings, the latter, B875 is present as a one ring arrangement in LHI. The redshift of wavelength from outside to inside is like an energy funnel, in which energy is collected and followingly transferred to the reaction center by Förster Resonance Energy Transfer. For years there was a discussion going on about the nature of a strong coupling of the B850 chromophores in LHII (and accordingly in LHI). Due to recent single molecule studies, this could be evaluated.

A.W.Roszak, T.D.Howard, J.Southall, A.T.Gardiner, C.J.Law, N.W.Isaacs, R.J.Cogdell, *Crystal structure of the RC-LH1 core complex from rhodopseudomonas palustris.*, Science

vol. 302 1969 (2003)

Scriptfile: psIIhypertextengl.vmd

Photosystem II of blue algae (cyanobacteria) and of green plants is responsible for water splitting and as a consequence for oxygen evolution on earth. By far most of the atmospheric oxygen originates from this source and is probably the only source to replenish oxygen on a large scale. The appearance of oxygen on a large scale did not only trigger evolution of more sophisticated life forms, but is also responsible for a reshaping of the earth's chemical composition. Since amino acids do not absorb in the visible range of the spectrum, chromophores like chlorophylls (green) carotenoids (orange) and pheophytins (blue) are bound to the protein. The protein complex consists of light harvesting proteins (CP-43, CP-47 etc.) and as a central part of the reaction center. In plants, the protein is embedded in the thylacoid membrane which can be clearly seen from the many α -Helices spanning the membrane.

The light harvesting pigments absorb the light and transfer the excitation energy to the heart of the protein complex, the reaction center, whose pigments are drawn bold. The reaction center absorbs the light directly or gets the excitation energy from the antenna pigment. Following the excitation of the chlorophyll dimer (Special Pair, violet) an electron is transferred via a chlorophyll and a pheophytin chromophore to two quinones on the other side of the thylacoid membrane. This electron flow produces an electrical potential across the membrane, which can be used to drive biochemical reactions. However, to reduce water to oxygen, four redox equivalents are required, but only one redox equivalent is produced per excitation of the special pair. The manganese-oxygen-cluster serves as a redox depot by successive oxidation of the cluster from oxidation state zero to +IV. The four successive excitations of the special pair. The distance between the Mn-O-cluster and the special pair is in principle too large for an electron transfer, but a tyrosine from the protein backbone serves as transient electron carrier.

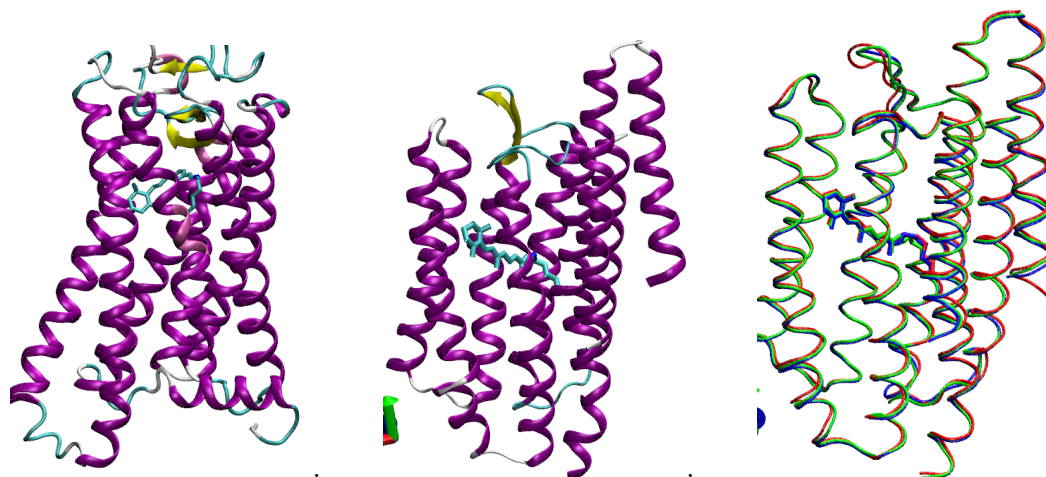
R. M. Hazen, *Evolution of Minerals*, Scientific American March 2010, *Como Evoluíram os Minerais*, Scientific American do Brasil April 2010

Vision proteins

Scriptfile: sensoryrhodopsinhypertextanimation.vmd

These sensory proteins are responsible for the primary processes of vision. Since the light reception has to happen with visible light, the protein contains a carotenoid chromophore, the opsin, which is connected covalently to the protein backbone via a nitrogen (Schiff's base). The chromophore contains conjugated double bonds to shift the absorption to the visible and reacts to light absorption with a photophysical process, a cis-trans isomerization around one of the double bonds. As in photosynthetic systems, the protein is embedded in a membrane, and consists dominantly of α -helices. The protein introduced here is not the vision protein of mammals, but instead a similarly built light detection protein from

an archaebacterium, the sensory rhodopsin. It is responsible for phototaxis (orientation of the bacterium to or away from light) This system is used here, since X-ray structures exist of transition states formed following excitation. By aligning the ground state structure and the transition state structures an animation is created that illustrates the principle of the functionality In a further presentation, the differently colored structures are superimposed. Although the protein is from an organism that even belongs to a different kingdom, the similarity to the vision pigments of mammals is surprising. In the VMD script collection you also find a structure for the vision pigment from bovine: bovinerhodopsin.vmd.



In the figures, the vision pigment of bovine is depicted on the left, followed by a comparable picture of the sensory rhodopsin and the superimposed structures of the ground state (green) and two transition states of sensory rhodopsin.

A major structural reorientation such as a cis-trans Isomerization of the chromophores involved and accordingly a strong reorganization of the protein lattice is important for light detection, since that creates a strong signal in the environment. In nature, there are multiple classes of chromophores that act correspondingly besides carotenoids, for example open chain tetrapyrrols in the plant and cyanobacterial light sensor phytochrome. Due to its simplicity the Photoactive Yellow Protein (PYP), that contains a cinnemetic acid derivative, is particularly well researched and gave useful information for light detection in nature (see pypanimation.vmd). Another mechanism, that is correlated to a large change of the chromophore bonds, is the plant light sensing protein phototropin, that contains a flavine molecule, see phototropinanimation.vmd.

In contrast to the strong rearrangements of the chromophore environment in vision, that causes a large 'signal' amplitude, it is crucial in photosynthesis to put as little energy as possible in rearrangements resp. vibrations. Therefore the favored chromophores of photosynthesis are rigid ring systems such as chlorophyll and pheophytin, whose structure is barely altered upon excitation.

Green Fluorescent Protein (GFP)

Scriptfile: gfpswitchanimation.vmd

Green Fluorescent Protein originates from the jellyfish *Aequorea victoria*. Unlike to other chromoproteins, that use separate cofactors, the chromophore in GFP is produced from three adjacent amino acids by an cyclization and oxidation step, and the only necessity to do so is the presence of oxygen. GFP and its variants are easily expressed in *E. coli* and many other types of cells and often used as a tandem protein marker. The protein has a barrel like structure consisting of 13 beta sheets and the chromophore fixed in the inside. Today, many mutations exist, that tune the wavelength of the chromophore, and recently there have been similarly structured proteins discovered in sea corals, that extend the color range even to the red of the spectrum. In these proteins, there is a 4th amino acid and an additional reaction step forming an additional double bond involved. The chromophore is part of an extensive hydrogen bonding network, that is partly responsible for the optical properties of GFP. For example the wt-GFP undergoes an excited state proton transfer (ESPT) that shifts the emission with respect to absorption by more than 100 nm from blue to green.

One of the novel fluorescent proteins discovered in corals could be structurally optimized to reversibly photoswitch between a non-fluorescent and a fluorescent state. The protein usually rests in the non-fluorescent trans state, but upon illumination with green light, it photoisomerizes into a fluorescent cis state. The back reaction is induced by illumination with blue light. Such a photoswitchable protein can for example be used for protein tracking or has been used in Ground State Depletion Microscopy. The animation included shows the underlying cis-trans isomerization of the GFP variant.

K.Brejc, T.K.Sixma, P.A.Kitts, S.R.Kain, R.Y.Tsien, M.Ormo, S.J.Remington, *Structural basis for dual excitation and photoisomerization of the aequorea victoria green fluorescent protein*, Proc.Nat.Acad.Sci.USA, vol. 94, p. 2306, 1997

M.Ormo, A.B.Cubitt, K.Kallio, L.A.Gross, R.Y.Tsien, S.J.Remington, *Crystal structure of the aequorea victoria green fluorescent protein*, Science, vol. 273, p. 1392, (1996)

M.Andresen, M.C.Wahl, A.C.Stiel, F.Graeter, L.Schaefer, S.Trowitzsch, G.Weber, C.Eggeling, H.Grubmueller, S.W.Hell, S.Jakobs *Structure and mechanism of the reversible photoswitch of a fluorescent protein*, Proc.Natl.Acad.Sci.USA, vol. 102, p. 13070 (2005)

Molecular Motors

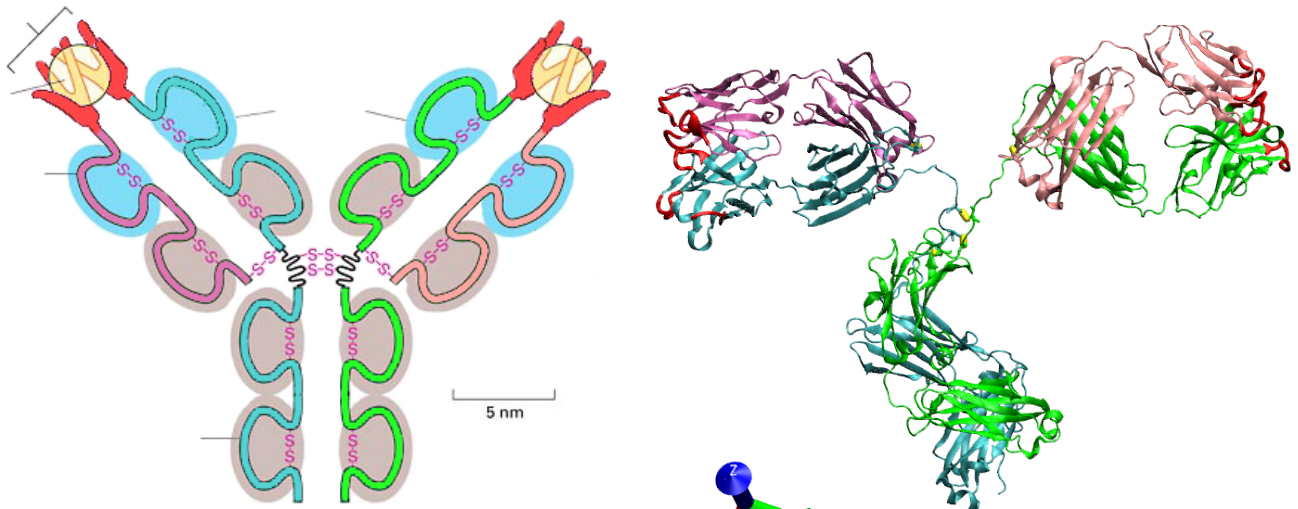
Scriptfile:atpsynthaseanimation.vmd (contributed by Christopher Bruhn)

ATP-synthase is a large trans-membrane complex that exploits a proton gradient across the membrane to produce ATP from ADP, which is the cells primary energy source. In the frame of the process, parts of the protein rotates in this process. The process has been investigated down to the single protein level.

Antibodies

Scriptfile: antibodyhypertextengl.vmd

Antibodies are the central proteins of the immune system. The protein consists of two heavy and two light chains, which are connected by disulfide bonds. On the tip of the antibody wings, the hypervariable regions are located, which bind to certain structures (Epitopes) of the binding partner (Antigene) specifically. In the following pictures, a schematic drawing of the antibody structure is shown (left), and the corresponding abstract structure presentation in the same colors produced with VMD. The heavy chains are green and cyan, the light chains are violet and pink, and the loops of the hypervariable regions are depicted in red.

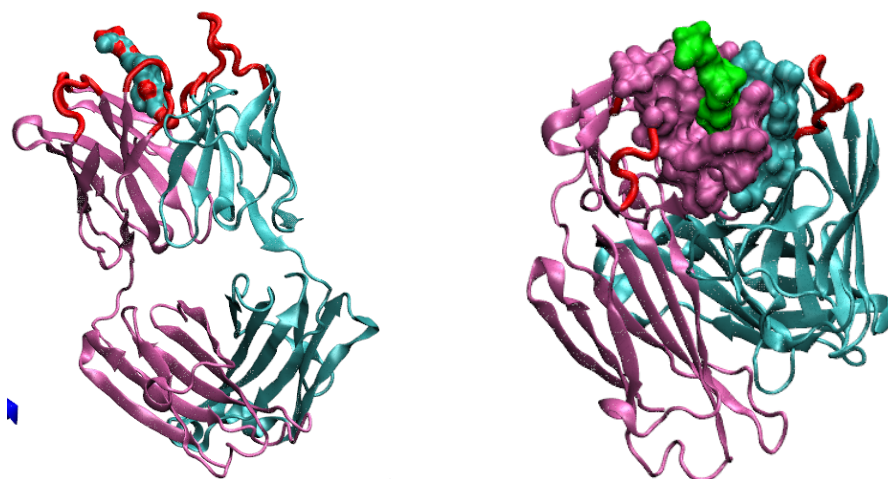


One can enzymatically cut the protein parts with the binding regions without loss of functionality. These protein parts are called FAB's. The disulfid bridges are also shown. In a further presentation, they are enhanced and the bridges in the hinge region, at the base of the FABs and in the individual fragments are distinguished by color.

The immune system produces antibodies against larger intruders like viruses and bacteria. The antibodies bind to particular locations of the intruder, for example to certain enzymes on the surface of viruses. Small molecules or proteins are usually not attacked. However, one can produce antibodies against these systems for analytical purposes or for research

applications. In this case, the molecule is bound to a large particle, that is intravenously utilized to an animal. The immunsystem identifies this particle as an intrudor and produces antibodies to structures on the surface of the particle, e.g. the bound molecules. This mechanism may also play a role in the occurance of allergic reactions. In this case the increase of microparticles in the environment due to civilisatory effects may be crucial. These particles may adsorb otherwise harmless substances from the environment.

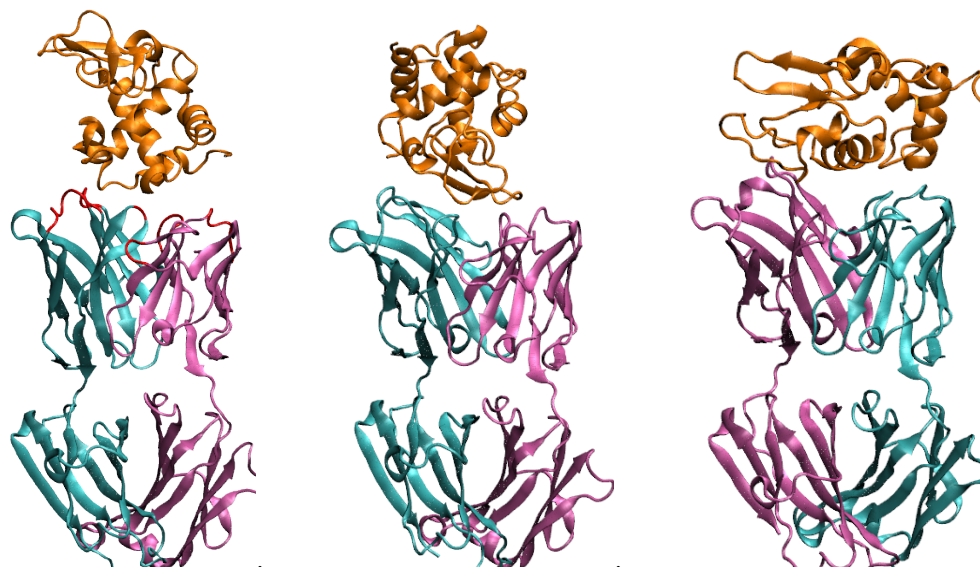
The presentations show antibody –antigen pairs produced in this manner. By far most of the analysed structures only deal with FAB –fragments, usually with the respective antigens. As an example for a small antigen, digoxigenin is chosen, which is a therapeutically active steroid. This fit into a binding pocket, that is shaped by the hypervariable regions. Different presentations illustrate the shape of the binding pocket and the position of the digoxigenin.



As an immune response of the body, multiple different structures against one antigen are produced, first with low binding constant and specificity. In the course of the immune response, antibody forming cells are selectively activated that produce more specific and better binding antibodies. (clonal selection theory). Even at that stage different antibodies exist, that bind differently to the antigen. In principle, every antibody producing cell (B – lymphocyte) forms an unique antibody. This is advantageous, since the intruder is recognized via different positions, and can not become resistant due to an alteration in one position. The heterogenous mixture is referred to as polyclonal antibody. A standardization of such a mixture is difficult, since the immune response is individually specific for each host animal. It would be advantageous for many experiments to have a homogenous single antibody. For structure determination, this is crucial, since differently formed hypervariable regions and the differently binding antigens would lead to an overlap of the different structures, that would prevent structure determination in the crucially interesting parts of the antibody. In the 1980's a method was developed, how an individual antibody forming cell could be bred, therefore producing a large amount of an individual antibody (monoclonal antibody). To this end a B –lymphocyte is fused with a myeloma cell, and this hybridoma cell is

selected and bred. In this manner, a large amount of monoclonal antibody can be produced. Since the cells can be frozen, one can produce the same type of antibody any time later.

As an example for an antibody for a large antigen, three different monoclonal antibodies to the protein lysozyme are shown. The lysozyme is colored in orange, the two chains of the antibody are cyan and violet. One can clearly distinguish the different epitopes, of lysozyme to which the antigen binds. An overlap of the three structures is also shown.



Apart from the naturally occurring principal structure of the antibodies, several other proteins were modified in a way to emulate antibodies. Particularly proteins, that already fulfill a binding function to specific molecule in nature, for example maltose binding protein or lipocalins, are genetically modified to bind to molecules of interest. As an example the script anticalin.vmd is included.

Ribosome

Scriptfile: ribosomebutton.vmd

the ribosome is a protein-RNA-complex, which is responsible for the translation of the genetic code into the amino acid sequence of the protein. The complex consists of protein parts (blue) as well as of ribosomal RNA, (r-RNA), namely the 16s r-RNA (olive), the 24s r-RNA (ocre) and the 5s-RNA. Not only the protein acts as a catalyst, but the r-RNA (ribozyme) as well. The ribosome consists of two subunits. Similar to the function of a tape recorder, the messenger -RNA (m-RNA) is read by transfer -RNA (t-RNA) while slipping through the gap between the two subunits. In the presentation the two subunits as well as three t-RNAs and a piece of m-RNA about to be translated are shown. There exists at least one t-RNA for each amino acid, which is responsible for the translation of the nucleotide codon corresponding to the amino acid into the peptid sequence. The nucleic acid triplet is located on one end of the t-RNA, while on the other end, the RNA is loaded with the respective amino acid, that is bound to the end of the emerging peptid chain.

Unlike the the usually plain structure of DNAs, r-RNA as well as t-RNA show a complex three dimensional structure, which is mostly due to the additional –OH group on the ribose sugar. The structure of ribosomes and of t-RNA as well as the genetic code is universal, which suggests, that these systems were also present in the last common ancestor of all living beings, prior the divergence into Archae, Bacteria and Eucariotes. The r-RNA is an important part of the ribosome complex, in structural as well as katalytic respect. Such an RNA is also called a Ribozyme. It is assumed that the present function of DNA (information storage) and protein (maschinery) was first both fulfilled by RNA (RNA –world).

Nitrogenase

Scriptfile: nitrogenasebutton.vmd

The Haber-Bosch process to get Ammonium and in following processes Nitrates from Nitrogen and Hydrogen is one of the most energy consuming processes in chemical industry. Although the reaction is exothermic, it has a large reaction barrier, as nitrogen is very stable due to its triple bond. To initiate the reaction, a mixture of nitrogen and hydrogen is heated to several hundred degrees and several hundred atm of pressure. With the help of the enzyme nitrogenase, which contains a very specific metal cluster (iron-molybdenium-sulfur-cluster) as a cofactor, the process takes place in nature under ambient conditions.

References and Notes

- [Caddigan et al.(2005)] E. Caddigan, J. Cohen, J. Gullingsrud, & J. Stone (2005). *VMD User's Guide*. Theoretical Biophysics Group, University of Illinois and Beckman Institute, Urbana, IL.
- [Humphrey et al.(1996)] William Humphrey, Andrew Dalke, & Klaus Schulten (1996). 'VMD – Visual Molecular Dynamics'. *Journal of Molecular Graphics* **14**:33–38.
- [Schellenberg(2010)] P. Schellenberg (2010). 'Biomolecules: Databanks, visualization and computations'. <http://webpe.sapia.uminho.pt/PSchellenberg/lecturelinksenglish.html>. also available in german: http://webpe.sapia.uminho.pt/PSchellenberg/biovis_de.html.
- [Stone et al.(2001)] John Stone, Justin Gullingsrud, Paul Grayson, & Klaus Schulten (2001). 'A system for interactive molecular dynamics simulation'. In John F. Hughes & Carlo H. Séquin (eds.), *2001 ACM Symposium on Interactive 3D Graphics*, pp. 191–194, New York. ACM SIGGRAPH.